Tensor-Based Low-Rank and Sparse Prior Information Constraints for Hyperspectral Image Denoising

GU XI WANG1,2, HONG WEI HAN3,4, EMMANUEL JOHN M. CARRANZA5, SI GUO6,7, KE GUO3, AND KEYAN XIAO8

1 College of Architecture and Environment, Sichuan University, Chengdu 610000, China
2 National Institute of Measurement and Testing Technology, Chengdu 610000, China
3 Geomathematics Key Laboratory of Sichuan Province, Chengdu University of Technology, Chengdu 610059, China
4 Engineering and Technical College, Chengdu University of Technology, Leshan 614000, China
5 Discipline of Geological Sciences, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal at Westville Campus, Durban 3629, South Africa
6 School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 610000, China
7 CCDC, Geological Exploration and Development Research Institute, Chengdu 610000, China
8 MIR Key Laboratory of Metallogeny and Mineral Resource Assessment, Institute of Mineral Resources, Chinese Academy of Geological Sciences, Beijing 100037, China

Corresponding author: Hongwei Han (hhw_666666@163.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0601505, in part by the National Natural Science Foundation of China under Grant 41672325, and in part by the Geomathematics Key Laboratory of Sichuan Province Foundation under Grant scsxdz201702 and Grant scsxdz2019yb01.

ABSTRACT Hyperspectral data have been widely used in various fields due to its rich spectral and spatial information in recent years. Yet, hyperspectral images are always tainted by a variety of mixed noises. These noises seriously limit the accuracy of subsequent applications. To remove noise, this paper, based on low-rank tensor decomposition, combined with non-local self-similar prior information, proposes a tensor-based non-local low-rank denoising model, where non-local self-similarity uses mainly spatial correlation while low-rank tensor decomposition method uses mainly spectral correlation between bands. Traditional tensor-based methods are commonly NP-hard to compute and are sensitive to sparse noise. However, the method proposed in this paper can separate efficiently the low-rank clean image from Gaussian noise and sparse noise (pulses, deadlines, stripes, speckle, etc.) by using a new tensor singular value decomposition (T-SVD) and tensor nuclear norm (TNN). The NP-hard task was also achieved well by the alternating direction multiplier method. Due to the full use of spectral and spatial information of the data, Gaussian noise and sparse noise can be effectively removed. The effectiveness of our algorithm was verified through experiments using simulated and real data.

INDEX TERMS Hyperspectral image denoising, non-local self-similar, low-rank tensor decomposition, sparse representation.

I. INTRODUCTION
The hyperspectral imaging spectrometer has important contributions to earth observation and remote sensing. Hyperspectral images obtained by such sensors have rich spectral and spatial information [1]. In recent years, hyperspectral data have been used widely in various fields, e.g., earth observation, environmental protection [2], target identification and tracking [3], food safety [4], precision farming, urban planning [5], and so on. However, in real scenarios, due to the limitations of the equipment, such as the dark current, sensor sensitivity, transmission errors and calibration during the imaging process, hyperspectral data are certainly tainted by a variety of noises. These noises include Gaussian noise, impulse noise, deadlines, stripes, speckle, among other [6]. The presence of noise in hyperspectral images not only affects the visual effect of images, but also limits the accuracy of ensuing processing, such as classification, spectral unmixing, target detection, and so on [7]. Therefore, it is necessary to remove noise before succeeding work [8]. Up to now, several...
hyperspectral denoising algorithms have been proposed, and these methods can be roughly divided into two categories: 2D extended method and tensor-based method.

The 2D extended methods use traditional 2D image denoising algorithms to process different bands of hyperspectral images [9]. Over the past 50 years, several 2D denoising methods have been proposed, including the classic K-SVD [10] and BM3D [11], which can be used directly for hyperspectral denoising. However, this extension ignores the inherent characteristics of hyperspectral images and usually fails to achieve good performance in practical applications. In recent years, many methods for hyperspectral noise reduction based on low-rank matrix decomposition [12]–[15] have been proposed. These methods consider clean hyperspectral data to be in a low-rank subspace, which can be modeled as a linear combination of a finite number of end-members. Zhang et al. [7] proposed the method of low-rank matrix recovery (LRMR), which performed excellently for removal of mixed noise. However, only spectral correlation but not spatial correlation is taken into account by the LRMR method. To overcome this limitation, the non-local self-similarity across space (NSS) priori [16]–[21] and some spatial information methods are used for noise reduction. In [22], a new low-rank spectral non-local approach (LRSNL) was proposed. This method considers both spectral and spatial information, and can effectively remove mixed noise. Similarly, to improve the low-rank model, a total variation regularization term was considered, and a total variation (TV) regularized low-rank matrix factorization (LRTV) method [9] was proposed. These improved methods are effective in hyperspectral denoising. However, these methods transform hyperspectral images into 2D images by vectorizing the data of each band, resulting in the loss of some useful structural information. To overcome this limitation, many tensor-based methods [23], [24] have been proposed for hyperspectral image denoising.

A hyperspectral image is composed of a stack of two-dimensional images, which can be considered naturally as a 3-order tensor. Tensor-based methods mainly use tensor decomposition to hyperspectral image denoising [25]. There are usually two kinds of tensor decomposition used for denoising, namely CANDECOMP/PARAFAC (CP) decomposition [26] and Tucker decomposition [27]. By using Tucker decomposition, a low-rank tensor approximation (LRTA) [28] method was proposed. Liu et al. [24] designed the PARAFAC method by using parallel factor analysis. The two methods consider the correlation between spectral bands and ignore the spatial structure information. To improve these algorithms, a new method based on tensor dictionary learning was proposed in [29], which considered both global correlation among the bands and non-local self-similarity across space. In recent years, researchers have proposed many low-rank tensor denoising methods [30]–[32]. These new methods use tensor singular value decomposition (T-SVD) [33]. In other words, a tensor nuclear norm (TNN) [34] defined based on T-SVD is used for noise reduction. A tensor robust principal component analysis (RPCA) [35] method was proposed by using T-SVD and tensor nuclear norm. This method takes advantage of the low rank and sparsity of the data, whereby it can remove effectively mixed noise in hyperspectral image data. Recently, Fan et al. [36] proposed a spatial-spectral total variation regularized low-rank tensor factorization (SSTV-LRTF) method. The total variation (TV) regularization [37] can preserve edge information effectively while obtaining low rank correlation among adjacent bands. Since the tensor-based denoising methods fully retain useful spatial structure information, they perform better than the 2D extended methods [38]–[40]. These existing algorithms have achieved good results in hyperspectral noise reduction. However, because noise is complex and diverse, no method exists that can deal with all kinds of noise simultaneously. Therefore, there is still a lot of room for improvement.

As mentioned above, non-local self-similarity (NSS) [41], [42] uses mainly spatial correlation while low-rank matrix decomposition method [41] uses mainly spectral correlation between bands. To utilize both spectral correlation and spatial information in hyperspectral data, a tensor-based non-local low-rank denoising model is proposed here. This method can eliminate both Gaussian noise and sparse noise (including impulse noise, deadlines, stripes, and speckle). The main ideas of this work can be summarized as follows:

1) Hyperspectral data are regarded as a third-order tensor. Based on the low-rank prior of clean hyperspectral data and the sparse nature of non-Gaussian noise, the low-rank model is applied to tensors composed of non-local similar full-band patch (FBP) [25] groups, and a tensor-based low-rank non-local denoising model is proposed.

2) In our model, low-rank tensors are constrained by tensor nuclear norm based on T-SVD, and non-convex term (logarithm sum) is used instead of tensor sparse measure \( \ell_0 \) term. Finally, ADMM (the alternating direction multiplier method) [9], [32] is used to design an effective algorithm to solve this problem. ADMM can turn a complex problem into a series of simple problems, and has been successfully used in [35], [36], and [40] for global convergence guarantees.

3) The experiment verifies that the algorithm we proposed can not only effectively remove Gaussian noise, but also has a good effect on the mixture of several kinds of noise.

The organization of the remainder of this paper is as follows. The next section gives some notations and explanations for tensors. In the third section, the proposed tensor-based non-local low-rank denoising model is described. Experimental results and analysis are described in the fourth section. Conclusions are then given in the last section.

II. NOTATIONS AND PRELIMINARIES

Succinct definitions and notations used in this are given in this section. Here, scalar, vector, matrix and tensor are represented as lowercase non-bold letters, lowercase boldface letters, capitalized boldface letters and uppercase calligraphic letters, respectively. A \( N \)-tensor is expressed as \( \mathbf{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N} \). Elements of \( \mathbf{A} \) are denoted as \( a_{i_1 \cdots i_N} \) where \( 1 \leq i_n \leq I_n \).
There are two norms that are often used, the Frobenius norm of a tensor \( A \) is defined as \( \| A \|_F = \left( \sum_{i_1, \ldots, i_n} |a_{i_1 \ldots i_n}|^2 \right)^{1/2} \), and the other is the matrix nuclear norm \( \| A \|_n = \sum_i \sigma_i(A) \), where \( \sigma_i(A) \) is the \( i \)th singular value of the matrix \( A \).

In the following chapters, we consider a third order tensor whose \( i \)th horizontal, lateral and frontal slices are expressed as \( A(i, :, :) \), \( A(:, i, :) \), and \( A(:, :, i) \), respectively. A frontal slice \( A(:, :, i) \) can also be expressed as \( A^{(i)} \). \( \hat{A} \) represents the fast Fourier transform (FFT) along the third dimension of tensor \( A \), namely \( \hat{A} = \text{fft}(A) \). Similarly, \( A \) can be obtained by \( \text{ifft}(\hat{A}, [1, 3]) \). For a 3-order tensor \( A \in \mathbb{R}^{I_1 \times I_2 \times I_3} \), let \( A \) be a block diagonal matrix defined as

\[
\tilde{A} = \text{bdiag}(\hat{A}) = \begin{bmatrix} \hat{A}^{(1)} \\ \hat{A}^{(2)} \\ \vdots \\ \hat{A}^{(I_3)} \end{bmatrix}
\]

where each block on its diagonal is the frontal slice \( \hat{A}^{(i)} \) of \( \hat{A} \). In order to define the multiplication between tensors, called the t-product [33], [35], [40], we first define the block circular matrix. For a 3-order tensor \( A \in \mathbb{R}^{I_1 \times I_2 \times I_3} \), its block circular matrix has a size of \( I_1 \times I_2 \times I_3 \), and is defined as

\[
\text{bcirc}(A) = \begin{bmatrix} A^{(1)} & A^{(I_3)} & \ldots & A^{(2)} \\ A^{(2)} & A^{(1)} & \ldots & A^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ A^{(I_3)} & A^{(I_3-1)} & \ldots & A^{(1)} \end{bmatrix}
\]

Likewise, the following two operators – unfold and fold – are defined as:

\[
\text{unfold}(A) = \begin{bmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(I_3)} \end{bmatrix}, \quad \text{fold(unfold}(A)) = A
\]

Below we give the product of two tensors, defined as follows:

Definition 1 (T-Product) [33]: Let \( A \in \mathbb{R}^{I_1 \times I_2 \times I_3} \) and \( B \in \mathbb{R}^{I_2 \times I_3 \times I_3} \) be two third-order tensors. Then, the t-product \( A \ast B \) is the \( I_1 \times I_2 \times I_3 \) tensor

\[
A \ast B = \text{fold(bcirc}(A) \cdot \text{unfold}(B))
\]

Definition 2 (Conjugate Transpose) [35]: The conjugate transpose of a tensor \( A \) of size \( I_1 \times I_2 \times I_3 \) is the \( I_2 \times I_1 \times I_3 \) tensor \( A^* \) obtained by conjugate transpose of each frontal slice and then reversing the order of transposed frontal slices 2 through \( I_3 \).

Definition 3 (Orthogonal Tensor) [35]: A tensor \( Q \in \mathbb{R}^{I_1 \times I_1 \times I_3} \) is orthogonal if

\[
Q^* \ast Q = Q \ast Q^* = I
\]

where \( I \in \mathbb{R}^{I_1 \times I_1 \times I_3} \) is the identity tensor, which first frontal slice in the \( I_1 \times I_1 \) identity matrix, and whose other frontal slices are all zeros.

\[
\text{Definition 4 (T-SVD) \cite{33}: Let } A \in \mathbb{R}^{I_1 \times I_2 \times I_3} \text{ be a third-order tensor. Then, } A \text{ can be factored as }
\]

\[
A = U \ast S \ast V^* \tag{6}
\]

where \( U \in \mathbb{R}^{I_1 \times I_2 \times I_3} \) and \( V \in \mathbb{R}^{I_2 \times I_3 \times I_3} \) are orthogonal, and \( S \in \mathbb{R}^{I_1 \times I_2 \times I_3} \) is a \( f \)-diagonal tensor.

\[
\text{Definition 5 (Tensor Multi-Rank and Tubal Rank) \cite{40}: The tensor multi-rank of a tensor } A \in \mathbb{R}^{I_1 \times I_2 \times I_3} \text{ is a vector } r \in \mathbb{R}^{I_1}, \text{ in which the } i \text{th entry is the rank of } \hat{A}^{(i)}. \text{ The tensor tubal rank denoted as rank}_t(A) \text{ is defined as the number of non-zero singular tubes of } S, \text{ where } S \text{ is obtained from the T-SVD of } A \text{ in definition 4. That is, }
\]

\[
\text{rank}_t(A) = \# \{ i : S(i, i, :) \neq 0 \} = \max_i r_i \tag{7}
\]

\[
\text{Definition 6 (Tensor Nuclear Norm) \cite{35, 40}: The tensor nuclear norm of a third-order tensor } A \in \mathbb{R}^{I_1 \times I_2 \times I_3} \text{, denoted as } \| A \|_n, \text{ is defined as the average of the nuclear norm of all the frontal slices of } \hat{A}, \text{ that is } \| A \|_n = (1/I_3) \sum_{i=1}^{I_3} \| \hat{A}^{(i)} \|_n. \text{ Notice that the tensor nuclear norm above is defined in the Fourier domain and is akin to the nuclear norm of the block circular matrix in the original domain. For detailed discussion, please refer to \cite{35}.}
\]

\section{III. TENSOR-BASED NON-LOCAL LOW-RANK DENOISING MODEL}

\subsection{A. Tensor-Based Low-Rank Denoise Model}

As mentioned above, a third-order tensor can be used to express a hyperspectral image. Hyperspectral images are typically contaminated by a variety of mixed noise. This paper mainly studies Gaussian noise and sparse noise, such as pulse noise, dead pixels or lines and stripes [4]. Denote \( d_H, d_W, \) and \( d_S \) as the spatial height, spatial width and number of spectral band of a hyperspectral image, the observed hyperspectral image \( Y \in \mathbb{R}^{d_H \times d_W \times d_S} \) can be expressed as the sum of the following three parts:

\[
Y = L + S + \mathcal{N} \tag{8}
\]

where \( Y \) is the noisy hyperspectral image, \( L \) is the clean hyperspectral data, \( S \) is sparse noise, and \( \mathcal{N} \) is Gaussian noise, and \( L \) is a low tubal rank tensor [32]. In fact, if a tensor is a low-rank tensor, then it is a low tubal rank tensor as well. Thus, low-rank tensor \( L \) is likewise recoverable by low tubal rank tensor constraint [40]. To restore the tensor \( L \), the following optimization model is considered:

\[
\min_{L, S, \mathcal{N}} \text{rank}_t(L) + \lambda \| S \|_0 + \| \mathcal{N} \|_F^2 \text{ s.t. } Y = L + S + \mathcal{N} \tag{9}
\]

The rank and \( \ell_0 \) terms in (9) can only take discrete values, which make the combinatorial optimization problem difficult to solve in the application. For the ease of solving, the tensor tube rank is replaced by the tensor nuclear norm, which is a convex function. In addition, \( \ell_0 \) is relaxed to log-sum form to simplify the calculation. This relaxation is proved an effective strategy for solving minimization problems [43], [44].
Then, the objective function is transformed into solving the following optimization problems:

$$\min_{L,S,N} \|L\|_s + \lambda P_{ls}(S) + \tau \|N\|_F^2$$

s.t. \(Y = L + S + N\) \(\text{(10)}\)

where \(P_{ls}(A) = \sum_{i,j,k} \log \left( |a_{i,j,k}| + \varepsilon \right)\), and \(\varepsilon\) is a small positive value. However, the low-rank recovery model only considers the global correlation along the spectral direction, but ignores the spectral structure in hyperspectral image. To make full use of spatial structure information, we establish a tensor-based non-local low-rank denoising model by introducing the prior information of NSS [45].

**B. TENSOR-BASED NON-LOCAL LOW-RANK DENOISING MODEL**

For a hyperspectral image, we define the full-band patch (FBP), which is stacked by patches at the same location of it over all bands, and the spatial size of each patch is \(6 \times 6\) \((d_h = d_w = 6)\) in this paper, where \(d_h\) and \(d_w\) represent the spatial height and spatial width of the hyperspectral image, respectively. Then, non-local self-similarity prior means that there are many FBPs (full-band patches), which are similar to a given local FBP [25]. In addition, non-local techniques mainly use spatial correlations between small patches of the hyperspectral image. To take advantage of the global spectral correlation and spatial structure priors of hyperspectral images, a non-local method is used to reconstruct a third-order tensor, and the newly constructed tensor retains perfectly the two prior information. A hyperspectral image can be expressed as a third-order tensor \(Y \in \mathbb{R}^{d_h \times d_w \times d_s}\), which is first partitioned into a group of overlapping full-band patches (FBPs) \(\{Y_i\}_{1 \leq i \leq d_h - d_s + 1} \subset \mathbb{R}^{d_h \times d_w \times d_s}\) \((d_h < d_h, d_w < d_w)\). Then, by converting each band of a FBP into a column vector, we can build a group of 2D FBPs \(\{P_i\}_{1 \leq i \leq d_h - d_s + 1} \subset \mathbb{R}^{d_h \times d_w \times d_s}\) \((d_h < d_h, d_w < d_w)\) to represent the hyperspectral image \(Y\). We can now represent all FBPs as a group of 2D patches \(\Omega_Y = \{Y_i \in \mathbb{R}^{d_h \times d_w \times d_s}\}_{i=1}^{N}\), where \(N = (d_h - d_h + 1) \times (d_w - d_w + 1)\) is the number of patches in the whole hyperspectral image. According to the NSS of hyperspectral image, and by performing block matching [11], [25], for a given local FBP \(Y_i\), we can find a set of FBPs similar to it in its non-local neighborhood from \(\Omega_Y\), \(Y_j \in \mathbb{R}^{d_h \times d_w \times d_s}\) \((d_h, d_w, d_s)\) represents a third-order tensor, which is stacked by \(Y_i\) and non-local FBPs similar to it in \(\Omega_Y\). Thus, the global correlation along the spectral direction and the prior information of non-local self-similarity along the space are well preserved. By using the low-rank nature of hyperspectral data and solving the following optimization problem, we can estimate the corresponding true non-local similar FBPs \(L_i\) from the noise image \(Y_i\):

$$\min_{L_i,S_i,N_i} \|L_i\|_s + \lambda P_{ls}(S_i) + \tau \|N_i\|_F^2$$

s.t. \(Y_i = L_i + S_i + N_i\) \(\text{(11)}\)

where \(L_i\) represents clean non-local similar FBPs, \(S_i\) is sparse noise, \(N_i\) represents Gaussian noise. It is obvious that \(L_i\) is a low-rank tensor. We can reconstruct the estimated hyperspectral image by aggregating all reconstructed \(L_i\)'s. The denoising process is portrayed in Figure 1.

**C. A TENSOR-BASED NON-LOCAL LOW-RANK DENOISING ALGORITHM BY USING ADMM**

To solve equation (11), we use the famous ADMM [9], [32], [35], [36], [46], [47]. This method can turn a complex problem into a series of simple problems. By introducing a Lagrangian multiplier \(D\), the augmented Lagrangian function corresponding to the optimization problem in (11) is:

$$L(L_i, S_i, N_i, D) = \|L_i\|_s + \lambda P_{ls}(S_i) + \tau \|N_i\|_F^2$$

$$+ <D, Y_i - (L_i + S_i + N_i)>$$

$$+ \frac{\mu}{2} \|Y_i - (L_i + S_i + N_i)\|_F^2$$ \(\text{(12)}\)

where \(\mu\) is the penalty parameter. For (12), the ADMM is used to update each variable and fix the remaining variables by minimizing the following equivalent function.

$$L(L_i, S_i, N_i, D) = \|L_i\|_s + \lambda P_{ls}(S_i) + \tau \|N_i\|_F^2$$

$$+ <D, Y_i - (L_i + S_i + N_i)>$$

$$+ \frac{\mu}{2} \|Y_i - (L_i + S_i + N_i)\|_F^2$$ \(\text{(13)}\)

In the \((k + 1)\)th iteration, \(L_i^{k+1}, S_i^{k+1}, N_i^{k+1}, D^{k+1}\) are updated as follows:

To update \(L_i^{k+1}\), we have

$$L_i^{k+1} = \arg \min_{L_i} \|L_i\|_s$$

$$+ \frac{\mu}{2} \|Y_i - (L_i + S_i^{k} + N_i^{k}) + \frac{D^{k}}{\mu}\|_F^2$$ \(\text{(14)}\)

To simplify the calculation, let \(X_i^{k} = Y_i - (S_i^{k} + N_i^{k}) + \frac{D^{k}}{\mu}\), then (14) can be rewritten as

$$L_i^{k+1} = \arg \min_{L_i} \|L_i\|_s + \frac{\mu}{2} \|L_i - X_i^{k}\|_F^2$$ \(\text{(15)}\)
Solving (15) corresponds to solving, in the frequency domain, the following tensor restoration problem
\[
\mathcal{L}_i^{k+1} = \arg \min_{\tilde{L}_i} \left\| \text{bdig}(\tilde{L}_i) \right\|_* + \frac{\mu}{2} \left\| \tilde{L}_i - \hat{X}_i^k \right\|^2_F \tag{16}
\]
where, \(\tilde{L}_i\) is the FFT along the third dimension of \(\mathcal{L}_i\), and \(\hat{X}_i^k\) is FFT along the third dimension \(X_i^k\). Equation (16) has a closed-form solution and is solvable by the method of singular value threshold [32]. To solve the optimization problem in (16), we can decompose it into \(d_n\) independent minimization problem. Let \(\hat{L}_i^{k+1,(i)}\) represent the \(i\)th frontal slice of \(\hat{L}_i^{k+1}\). Likewise, we can define \(\hat{X}_i^{k,(i)}\), thus
\[
\hat{L}_i^{k+1,(i)} = \arg \min_{\hat{L}_i^{(i)}} \left\| \hat{L}_i^{(i)} \right\|_* + \frac{\mu}{2} \left\| \hat{L}_i^{(i)} - \hat{X}_i^{k,(i)} \right\|^2_F \tag{17}
\]
Before solving this problem, let us give some definitions. If \(r = \text{rank}(X)\) and \(\sigma_i(l = 1, 2, \cdots, r)\) denotes the \(i\)th singular value of \(X\), then, the singular value decomposition (SVD) of \(X\) can be expressed as \(X = U \text{diag} \{\sigma(l)\}_{1 \leq l \leq r} V^T\). Thus, we can define the singular value threshold operator \(\text{SVD}_r(X)\) as:
\[
\text{SVD}_r(X) = U \text{diag} \left\{ \left( \sigma_i - \tau \right)_+ \right\}_{1 \leq i \leq r} V^T \tag{18}
\]
where \((\cdot)_+ = \max(\cdot, 0)\). We have
\[
\hat{L}_i^{k+1,(i)} = \text{SVD}_r \left( \hat{X}_i^{k,(i)} \right), \quad i = 1, 2, \cdots, d_n \tag{19}
\]
and updating of \(\mathcal{L}_i^{k+1}\) is achievable via inverse Fourier transform, namely
\[
\mathcal{L}_i^{k+1} = \text{ifft}(\hat{L}_i^{k+1,(i)}), \quad [1, 3] \tag{20}
\]
To update \(S_i^{k+1}\), the following sub-problem is solved.
\[
S_i^{k+1} = \arg \min_{S_i} \lambda P_{\alpha,S_i}(S_i)
+ \frac{\mu}{2} \left\| y_i - \left( L_i^{k+1} + S_i + N_i^k \right) + \frac{D_k}{\mu} \right\|^2_F \tag{21}
\]
Equation (21) is convertible to the following optimization problem:
\[
S_i^{k+1} = \arg \min_{S_i} \lambda P_{\alpha,S_i}(S_i) + \frac{\mu}{2} \left\| S_i - Q \right\|^2_F \tag{22}
\]
where \(Q = y_i - \left( L_i^{k+1} + N_i^k \right) + \frac{D_k}{\mu}\). It has been proved to have a closed solution [48], thus:
\[
S_i^{k+1} = D_{\alpha,\epsilon}(Q) \tag{23}
\]
where \(\alpha = \frac{\lambda}{\mu}, D_{\alpha,\epsilon}(\cdot)\) is the threshold operator, and it is defined as:
\[
D_{\alpha,\epsilon}(z) =
\begin{cases}
0 & \text{if } c_2 \leq 0 \\
\text{sign}(z) \left( \frac{c_1 + \sqrt{c_2}}{2} \right) & \text{if } c_2 > 0
\end{cases}
\tag{24}
\]
where \(c_1 = |z| - \epsilon, c_2 = (c_1)^2 - 4 (\alpha - \epsilon |z|)\).

For \(N_i^{k+1}\), we have the following sub-problems.
\[
N_i^{k+1} = \arg \min_{N_i} \tau \left\| N_i \right\|^2_F
+ \frac{\mu}{2} \left\| y_i - \left( L_i^{k+1} + S_i^{k+1} + N_i \right) + \frac{D_k}{\mu} \right\|^2_F \tag{25}
\]
Equation (25) is a least square problem, and we can obtain its closed solution as:
\[
N_i^{k+1} = \frac{\mu}{(2\tau + \mu)} \left[ y_i - \left( L_i^{k+1} + S_i^{k+1} + N_i^{k+1} \right) \right] + \frac{D_k}{\mu} \tag{26}
\]
The following formulation can be used to update \(D^{k+1}\):
\[
D^{k+1} = D^k + \mu \left[ y_i - \left( L_i^{k+1} + S_i^{k+1} + N_i^{k+1} \right) \right] \tag{27}
\]
The hyperspectral denoise algorithm proposed in this paper is summarized in algorithm 1, which is described as follows.

\textbf{Algorithm 1 The Hyperspectral Denoise Algorithm}

\textbf{Input:} Noise hyperspectral image \(Y \in \mathbb{R}^{d_y \times d_w \times d_z}\), parameters \(\lambda, \tau\).
\textbf{1. Initialization:} Parameters \(\eta = 1.1, \mu^0 = 10^{-3}, \mu_{\text{max}} = 10^{10}, k = 0\)
\textbf{2. Construct the entire FBP set} \(\Omega_v\)
\textbf{3. Construct the set of similar FBP group set} \(\{\gamma_i\}_{i=1}^K\)
\textbf{For each FBP groups} \(\gamma_i\) \textbf{do}
\textbf{4. Repeat}
\textbf{5. Update} \(L^{k+1}\) by (20)
\textbf{6. Update} \(S^{k+1}\) by (23)
\textbf{7. Update} \(N^{k+1}\) by (26)
\textbf{8. Update Lagrangian multiplier} \(D^{k+1}\) by (27)
\textbf{9. Update} \(\mu^{k+1} \leftarrow \min \left( \eta \mu^{k+1}, \mu_{\text{max}} \right)\)
\textbf{10.} \(k \leftarrow k + 1\)
\textbf{11.} until stopping criterion is satisfied
\textbf{Endfor}
\textbf{12. Aggregate} \(\{\mathcal{L}_i\}_{i=1}^K\) to form the clean hyperspectral image \(\mathcal{L}\).
\textbf{13. Output} Useful signal \(\mathcal{L}\).

It is worth noting that the penalty parameter \(\mu\) adopts an adaptive updating strategy \(\mu^{k+1} \leftarrow \min \left( \eta \mu^{k+1}, \mu_{\text{max}} \right)\) in our algorithm, where \(\mu_{\text{max}}\) represents the upper bound of \(\mu\). This is because ADMM converges very slowly with a fixed \(\mu\) [40].

\textbf{IV. RESULTS OF EXPERIMENTS AND ANALYSIS}

To verify the performance of the proposed hyperspectral image denoising algorithm, we performed experiments using simulated and real data. To analyze the performance of the algorithm quantitatively and qualitatively, seven typical denoising algorithms were selected for comparative experiments. These algorithms include band-wise K-SVD [10], band-wise BM3D [11], BM4D [21] and LRMR [7], which are two-dimensional extended to three-dimensional methods,
and PARAFAC [24], LRTA [28] and tensor dictionary learning (TDL) [29], which are tensor-based methods. The parameters of all the seven algorithms were adjusted according to their descriptions in the corresponding reference papers to ensure their best performance. The algorithm parameters proposed in this paper are discussed in detail in section IV-C below. The spatial size of the patch was important. In fact, the larger the size, the better the accuracy and the longer the computational time will be. In principle, $3 \times 3, 4 \times 4, \ldots, 8 \times 8$ are all fine but we finally chose $6 \times 6$ as the tradeoff between accuracy and computational time for all cases.

The methods were run on a PC with an Intel(R) Core(TM) i9-9900k Processor @3.60 GHz and 64.00 GB of RAM. To evaluate quantitatively the performance of the proposed algorithm, peak signal-to-noise ratio (PSNR), structural similarity index measurement (SSIM) [7] and feature similarity (FSIM) [49] were calculated. For hyperspectral data, we calculated the PSNR, SSIM and FSIM of each band and estimated the means, which were called mean PSNR (MPSNR), mean SSIM (MSSIM) and mean FSIM (MFSIM), respectively. The larger the mean of each of the three measurements, the better the noise reduction performance of the algorithm.

A. SIMULATED DATA EXPERIMENTS

Two public-domain datasets were used in these experiments to evaluate the performance of the proposed algorithm. These include the Washington DC Mall dataset and the Pavia City Center dataset [2]. The former contains $1208 \times 307$ pixels with 191 bands; a sub-image with size of $256 \times 256 \times 191$ was used for our experiment (Figure 2(a)). The latter was acquired by the reflective optics system imaging spectrometer (ROSIS-03), and the first 22 bands (containing all the noisy bands) were removed to simulate a clean image; a sub-image of size $200 \times 200 \times 80$ was extracted for our experiment (Figure 2(b)).

![Figure 2](https://example.com/f2.jpg)

**FIGURE 2.** False color composites of (a) the Washington DC Mall dataset and (b) the Pavia City Center dataset.

In these experiments, Gaussian noise and sparse noise were added to the two reference hyperspectral datasets and the following two cases were considered:

Case 1: We added Gaussian noise of independent and identical distribution to all bands of the two hyperspectral datasets. According to [2] and [4], considering the impact of strong noise on performance of the proposed algorithm, the mean value of Gaussian noise was set to zero and its variances were set to $0.04, 0.06, 0.08$ and $0.1$, respectively. In fact, different variances correspond to different noise levels. Each of these four variance values corresponds to the mean signal noise ratio (SNR) value of all bands of a hyperspectral image, i.e., $0.04, 0.06, 0.08$, and $0.1$ correspond to mean SNR values of $11.12, 7.99, 6.00$, and $4.64$ dB for the Washington DC Mall dataset, and to $13.68, 10.39, 8.19$, and $6.61$ dB for the Pavia City Center dataset.

Case 2: We added Gaussian noise and speckle noise to all bands of the two hyperspectral datasets. The mean value of Gaussian noise was set to zero and its variance was fixed at $0.02$. According to [4] and [40], speckle noise was generated by adding salt-and-pepper noise to the hyperspectral image, and the percentage of sparse noise was set to $0.05, 0.1$ and $0.15$. Through comparative analysis, we believe that these three values can lead to effective verification of the performance of the proposed algorithm. Similarly, different percentages of sparse noise correspond to different noise levels. Each of these three values corresponds to the mean signal noise ratio (SNR) value of all bands of a hyperspectral image, i.e., $0.05, 0.10$, and $0.15$ correspond to mean SNR values of $3.07, 1.99$, and $1.55$ dB for the Washington DC Mall dataset, and $5.07, 3.49$, and $2.82$ dB for the Pavia City Center dataset.

For Case 1, the results of applying different de-noising algorithms to the Washington DC Mall and the Pavia City Center datasets (Table 1) show that the proposed algorithm was superior to the other seven algorithms in terms of each performance index. Therefore, our method performed best and that it can remove Gaussian noise effectively. Table 2 shows values of the three image quality indicators for the different methods in Case 2. It can be seen from this table that the tensor-based methods achieved better results than K-SVD, BM3D, and BM4D, and that LRMR provides the higher values than the first six algorithms for MPSNR, MSSIM, and MFSIM since it was a method based on low-rank matrix recovery. We can also see that the proposed algorithm was superior to the other algorithms. These results show that our proposed algorithm can remove mixture of Gaussian noise and sparse noise effectively.

Figures 3 and 4 show the results of the different algorithms in denoising the Washington DC Mall dataset in the two simulated cases. Figures 5 and 6 show the results of the different algorithms in denoising the Pavia City Center dataset in the two simulated cases. The original band images in Figures 3 and 5 were contaminated with Gaussian noise (the mean value was set to zero and variance to $0.1$). As can be seen from these two images, the K-SVD and BM3D methods lost some information because they did not take into account the differences between different pixels. The tensor-based LRTA and PARAFAC methods cannot remove Gaussian noise completely. Denoising was over-smoothed by the BM4D method. The image restored by TDL properly...
removes Gaussian noise while finely preserves the structure underlying the image. The proposed algorithm and LRMR achieved better results than the other algorithms as images restored by the two methods preserve the edge and detailed information, proving that they can remove Gaussian noise effectively. The original band images in Figures 4 and 6 were simultaneously contaminated with Gaussian noise (with mean value of zero and variance of 0.02) and sparse noise (set to 0.15). From the two figures, it can be seen that the K-SVD, BM3D and BM4D methods almost failed to remove noise. Combined with the preceding content, it can be concluded that they have no ability to remove salt-and-pepper noise. The tensor-based LRTA and PARAFAC methods were capable of removing sparse noise, but some structural information was lost. TDL method performs comparatively better in structure preserving than LRTA and PARAFAC methods. Although LRMR obtains much better results than TDL, some impulse noise still exist in the image. The algorithm proposed in this paper had better performance, which shows that it can remove mixed noise effectively.

### B. REAL DATA EXPERIMENTS

We used the Indian Pines dataset [40] for verification of the proposed algorithms performance. This dataset was acquired by the AVIRIS sensor in 1992. It consists of 145 × 145 pixels with 220 spectral bands. After removing water absorption

### TABLE 1. Average performance of different denoising algorithms in case 1.

| Dataset          | Variance | Metrics          | K-SVD | BM3D | LRTA  | PARAFAC | BM4D | TDL | LRMR | Proposed |
|------------------|----------|------------------|-------|------|-------|---------|------|-----|------|----------|
| Washington DC Mall | 0.04     | MPSNR            | 30.2675 | 32.0266 | 40.1536 | 31.0678 | 38.8895 | 30.7849 | 39.7697 | 40.1794 | 40.7970 |
|                  |          | MSSIM            | 0.8763  | 0.9163 | 0.9855 | 0.9030  | 0.9809  | 0.9819  | 0.9875  | 0.9892  |
|                  |          | MFSIM            | 0.9532  | 0.9515 | 0.9272 | 0.9420  | 0.9885  | 0.9920  | 0.9916  | 0.9939  |
|                  | 0.06     | MPSNR            | 27.8133 | 29.8120 | 37.2922 | 30.8941 | 36.2009 | 38.4468 | 37.3703 | 38.6870 |
|                  |          | MSSIM            | 0.8002  | 0.8667 | 0.9697 | 0.9000  | 0.9650  | 0.9642  | 0.9772  | 0.9824  |
|                  |          | MFSIM            | 0.9000  | 0.9242 | 0.9337 | 0.9401  | 0.9794  | 0.9902  | 0.9853  | 0.9899  |
|                  | 0.08     | MPSNR            | 26.2285 | 28.3674 | 35.1881 | 30.7481 | 34.3360 | 36.2610 | 35.4838 | 37.1098 |
|                  |          | MSSIM            | 0.7292  | 0.8208 | 0.9540 | 0.8971  | 0.9469  | 0.9683  | 0.9635  | 0.9751  |
|                  |          | MFSIM            | 0.8180  | 0.8988 | 0.9378 | 0.9584  | 0.9603  | 0.9676  | 0.9785  | 0.9854  |
|                  | 0.1      | MPSNR            | 25.0878 | 27.3069 | 33.8120 | 30.6744 | 32.9256 | 34.9114 | 34.1502 | 35.9263 |
|                  |          | MSSIM            | 0.6651  | 0.7786 | 0.9401 | 0.8946  | 0.9276  | 0.9570  | 0.9546  | 0.9677  |
|                  |          | MFSIM            | 0.8406  | 0.8755 | 0.9651 | 0.9377  | 0.9856  | 0.9725  | 0.9719  | 0.9806  |

### TABLE 2. Average performance of different denoising algorithms in case 2.

| Dataset          | Sparse percentage | Metrics          | K-SVD | BM3D | LRTA  | PARAFAC | BM4D | TDL | LRMR | Proposed |
|------------------|-------------------|------------------|-------|------|-------|---------|------|-----|------|----------|
| Washington DC Mall | 0.05    | MPSNR            | 17.5283 | 17.5334 | 23.8066 | 30.8242 | 18.6624 | 31.8594 | 32.4719 | 33.0757 |
|                  |          | MSSIM            | 0.4984  | 0.5025 | 0.6510 | 0.9139  | 0.5184  | 0.9375  | 0.7412  | 0.9430  |
|                  |          | MFSIM            | 0.8234  | 0.8300 | 0.8235 | 0.9558  | 0.8448  | 0.9722  | 0.9640  | 0.9731  |
|                  | 0.1      | MPSNR            | 14.6713 | 14.5392 | 21.9644 | 27.0468 | 15.5274 | 27.3406 | 32.0581 | 32.4213 |
|                  |          | MSSIM            | 0.3135  | 0.3104 | 0.5693 | 0.8508  | 0.3297  | 0.8749  | 0.7275  | 0.8945  |
|                  |          | MFSIM            | 0.7253  | 0.7317 | 0.7890 | 0.9326  | 0.7406  | 0.9511  | 0.9507  | 0.9572  |
|                  | 0.15     | MPSNR            | 12.9805 | 12.7775 | 21.7650 | 24.2700 | 13.7293 | 24.5303 | 31.4963 | 31.9364 |
|                  |          | MSSIM            | 0.2192  | 0.2139 | 0.5834 | 0.7778  | 0.2622  | 0.8140  | 0.6928  | 0.8437  |
|                  |          | MFSIM            | 0.6526  | 0.5587 | 0.8050 | 0.9070  | 0.9063  | 0.9303  | 0.9415  | 0.9409  |
|                  | 0.05     | MPSNR            | 17.5841 | 17.6022 | 25.6793 | 29.7448 | 19.7336 | 30.9157 | 31.7925 | 32.8082 |
|                  |          | MSSIM            | 0.4650  | 0.4693 | 0.7180 | 0.8810  | 0.5473  | 0.9158  | 0.8941  | 0.9050  |
|                  |          | MFSIM            | 0.8087  | 0.8168 | 0.8289 | 0.9253  | 0.8238  | 0.9463  | 0.9360  | 0.9511  |
|                  | 0.08     | MPSNR            | 14.7176 | 14.5994 | 24.4609 | 26.5511 | 16.5890 | 27.6160 | 30.5473 | 32.0789 |
|                  |          | MSSIM            | 0.2775  | 0.2743 | 0.6905 | 0.8162  | 0.3738  | 0.8368  | 0.8839  | 0.8376  |
|                  |          | MFSIM            | 0.7030  | 0.7106 | 0.8240 | 0.8987  | 0.7889  | 0.9112  | 0.9106  | 0.9223  |
|                  | 0.1      | MPSNR            | 13.0418 | 12.8451 | 22.8023 | 24.0426 | 17.7971 | 24.6948 | 29.3587 | 31.9417 |
|                  |          | MSSIM            | 0.1867  | 0.1814 | 0.6388 | 0.7433  | 0.1998  | 0.8109  | 0.8572  | 0.8632  |
|                  |          | MFSIM            | 0.6257  | 0.6331 | 0.8087 | 0.8720  | 0.6401  | 0.8834  | 0.8864  | 0.8910  |
bands 104-108, 150-163 and 220, the final dataset we used contained 200 bands (Fig. 7). The image was corrupted mainly by Gaussian noise and impulse noise.
We studied the influence of parameters in the proposed method. To evaluate and analyze these optimal parameters, we used the Washington DC Mall dataset and the Pavia City Center dataset as an example in Case 2 (the mean Gaussian noise was zero and its variance was 0.02, and sparse noise was set to 0.15), and used MPSNR as the evaluation measure of the analysis. To obtain $L_i$ accurately, the choices for parameter $\lambda$ used in equation (11) were $\{10^{-4}, 5 \times 10^{-4}, 10^{-3}, 5 \times 10^{-3}, 10^{-2}, 5 \times 10^{-2}, 0.1, 0.5, 1\}$, and the choices for $\tau$ were $\{10^{-4}, 5 \times 10^{-4}, 10^{-3}, 5 \times 10^{-3}, 10^{-2}, 5 \times 10^{-2}, 0.1, 0.5, 1, 1.5\}$. Figures 11 and 12 show the MPSNR values of the proposed algorithm, as related to regularization parameters $\lambda$ and $\tau$ on the Washington DC Mall dataset and the Pavia City Center dataset in Case 2. As can be seen from Figures 11 and 12, when the values of these two parameters were small, the value of MPSNR was almost constant. When $\tau$ changes from 0.005 to 1.5, the value of MPSNR first increases rapidly with the increase of $\lambda$, then decreases, and finally tends to be stable. In addition, the advantage of the proposed method becomes apparent when the noise level increases. As we can see, the best performance of the proposed algorithm was obtained when $\lambda = 0.005$ and $\tau = 0.05$ for the Washington DC Mall dataset, and when $\lambda = 0.01$ and $\tau = 0.1$ for the Pavia City Center dataset. At the same time, we set $\lambda = 0.005$ and $\tau = 0.05$ for the Indian Pines dataset. The penalty parameter $\mu$ adopts an adaptive updating strategy $\mu_{k+1} \leftarrow \min (\eta \mu_{k+1}, \mu_{\text{max}})$, where $\mu_{\text{max}}$ represents the upper bound of $\mu$. We empirically set $\eta = 1.1, \mu^0 = 1e-3, \mu_{\text{max}} = 1e10$.

2) COMPUTATIONAL TIME COMPARISON
To evaluate further the proposed algorithm, we compared its running time with those of the other seven algorithms. Table 3 reports the running times of applying the different methods to the Indian Pines dataset, and it shows that the proposed method ran much faster than PARAFAC and BM4D, but slower than five of the other methods. The latter result

![Algorithm](image)

**FIGURE 10.** Denoising results of different algorithms on the Indian Pines dataset: (a) Original band 219, (b) K-SVD, (c) BM3D, (d) LRTA, (e) PARAFAC, (f) BM4D, (g) TDL, (h) LRMR, (i) the proposed algorithm.

obtain better results than the first three algorithms. However, there were some over-smoothened areas. The BM4D method resulted in some loss of boundary information. The TDL and LRMR preserve the structure underlying the image. However, the images restored by TDL and LRMR have some stripes. The method proposed here obtained the best results; it retained the boundary and structure information of the image. This proves that our method can deal with both Gaussian noise and sparse noise. As can be seen from Figure 10, the K-SVD, BM3D, LRTA, and BM4D methods failed to remove noise, and after denoising the image has almost no available information. The PARAFAC method achieved better results than those four methods, but heavy noise can be removed, and after denoising the image has almost no available information. The PARAFAC method achieved better results than those four methods, but heavy noise can be also observed. The TDL preserves the structure of image. However, the images restored by TDL remains some unexpected sharp noises. LRMR method obtains good results except in the marginal area. This was because the patch-based method may result in the loss of some interdimensional information. The method proposed here has obtained the best denoising results whereby the boundary information was best preserved.

C. DISCUSSION
1) CHOICE OF PARAMETERS
We studied the influence of parameters in the proposed algorithm, which include two regularization parameters $\lambda$ and $\tau$, and one penalty parameter $\mu$. The parameter $\lambda$ was the one that constrains sparse noise and $\tau$ was used to limit Gaussian noise. It was clear that we should consider how to select these parameters properly. To evaluate and analyze these optimal parameters, we used the Washington DC Mall dataset and the Pavia City Center dataset as an example in Case 2 (the mean Gaussian noise was zero and its variance was 0.02, and sparse noise was set to 0.15), and used MPSNR as the evaluation measure of the analysis. To obtain $L_i$ accurately, the choices for parameter $\lambda$ used in equation (11) were $\{10^{-4}, 5 \times 10^{-4}, 10^{-3}, 5 \times 10^{-3}, 10^{-2}, 5 \times 10^{-2}, 0.1, 0.5, 1\}$, and the choices for $\tau$ were $\{10^{-4}, 5 \times 10^{-4}, 10^{-3}, 5 \times 10^{-3}, 10^{-2}, 5 \times 10^{-2}, 0.1, 0.5, 1, 1.5\}$. Figures 11 and 12 show the MPSNR values of the proposed algorithm, as related to regularization parameters $\lambda$ and $\tau$ on the Washington DC Mall dataset and the Pavia City Center dataset in Case 2. As can be seen from Figures 11 and 12, when the values of these two parameters were small, the value of MPSNR was almost constant. When $\tau$ changes from 0.005 to 1.5, the value of MPSNR first increases rapidly with the increase of $\lambda$, then decreases, and finally tends to be stable. In addition, the advantage of the proposed method becomes apparent when the noise level increases. As we can see, the best performance of the proposed algorithm was obtained when $\lambda = 0.005$ and $\tau = 0.05$ for the Washington DC Mall dataset, and when $\lambda = 0.01$ and $\tau = 0.1$ for the Pavia City Center dataset. At the same time, we set $\lambda = 0.005$ and $\tau = 0.05$ for the Indian Pines dataset. The penalty parameter $\mu$ adopts an adaptive updating strategy $\mu_{k+1} \leftarrow \min (\eta \mu_{k+1}, \mu_{\text{max}})$, where $\mu_{\text{max}}$ represents the upper bound of $\mu$. We empirically set $\eta = 1.1, \mu^0 = 1e-3, \mu_{\text{max}} = 1e10$.

![MPSNR](image)

**FIGURE 11.** MPSNR of the proposed algorithm, as related to regularization parameters $\lambda$ and $\tau$ on the Washington DC Mall dataset in Case 2: (a) variance is 0.02 and percentage of sparse noise is 0.05; (b) variance is 0.02 and percentage of sparse noise is 0.10; (c) variance is 0.02 and percentage of sparse noise is 0.15.
TABLE 3: Computational time (seconds) of different methods on the Indian Pines dataset.

| Method  | K-SVD | BM3D | LRTA | PARAPAC | BMID | TDL | LRM  | proposed |
|---------|-------|------|------|---------|------|-----|------|----------|
| Time(s) | 20.691 | 0.0711 | 1.2486 | 638.0962 | 115.1182 | 7.2878 | 52.7459 | 68.1821 |

was mainly due to the use of non-local methods and SVD in solving low-rank problems, which led to the slowness of the methods. This is worth comparing with the optimal performance we achieved. In future work, we will optimize our algorithm to make it run faster.

V. CONCLUSIONS
In this paper, we proposed a tensor-based low-rank sparse prior denoising model. The main results of this study can be summarized as follows:

1. Hyperspectral data tensor was reconstructed by utilizing the non-local self-similarity property of hyperspectral data. The global correlation of the reconstructed data along the spectral direction and the prior information of non-local self-similarity along the space were well preserved.

2. A non-local low-rank sparse model was constructed based on the low-rank and the prior information of non-local self-similarity of hyperspectral data. This low-rank model mainly used spectral information, while non-local self-similarity obtains spatial information of hyperspectral data. Therefore, our model fully considers both spectral and spatial information.

3. The $\ell_0$ term was replaced by a non-convex log-sum term, which is sparser than the regularization term $\ell_1$.

4. A tensor-based non-local low-rank denoising algorithm was proposed by using the alternating direction multiplier method. Two experiments using simulated data and one experiment using real data showed that our proposed algorithm can remove Gaussian noise and sparse noise (which include streaks, dead lines, and pulses) effectively and better than other denoising methods.

There is no doubt that the proposed algorithm still has room for improvement. In the future, we can try to incorporate more prior information into our proposed model to improve further the performance of the denoising method. Furthermore, further improving the running speed of the proposed algorithm is our future research direction.

REFERENCES

[1] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, “Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 5, no. 2, pp. 354–379, Apr. 2012.

[2] L. Zhuang and 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.

[3] J. Li, Q. Yuan, H. Shen, and L. Zhang, “Noise removal from hyperspectral image with joint spectral–spatial distributed sparse representation,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 9, pp. 5425–5439, Sep. 2016.

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

[5] 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.

[6] 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.

[7] 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.

[8] Z. Wu, Q. Wang, Z. Wu, and Y. Shen, “Total variation-regularized weighted nuclear norm minimization for hyperspectral image mixed denoising,” J. Electron. Imag., vol. 25, no. 1, Feb. 2016, Art. no. 013037.

[9] 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.

[10] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” IEEE Trans. Signal Process., vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[11] 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.

[12] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, “Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Vancouver, BC, Canada, Jun. 2016, pp. 2080–2088.
[13] E. J. Candes and Y. Plan, “Matrix completion with noise,” Proc. IEEE, vol. 98, no. 6, pp. 925–936, Jun. 2010.

[14] Z. Zhou, X. Li, J. Wright, E. Candes, and Y. Ma, “Stable principal component pursuit,” in Proc. IEEE Int. Symp. Inf. Theory, Austin, TX, USA, Jun. 2010, pp. 1518–1522.

[15] E. J. Candes, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” J. ACM, vol. 58, no. 3, Jun. 2011, Art. no. 11.

[16] J. Marral, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, “Non-local sparse models for image restoration,” in Proc. IEEE 12th Int. Conf. Comput. Vis., Kyoto, Japan, Sep. 2009, pp. 2272–2279.

[17] 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. (CVPR), San Diego, CA, USA, Jun. 2005, pp. 60–65.

[18] Y. Zhang, L. Zhai, G. Shi, and Y. Shi, “Nonlocally centralized sparse representation for image restoration,” IEEE Trans. Image Process., vol. 22, no. 4, pp. 1620–1630, Apr. 2013.

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

[20] Y. Qian and M. Ye, “Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 6, no. 2, pp. 499–515, Apr. 2013.

[21] 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, Jan. 2013.

[22] R. Zhu, M. Dong, and J.-H. Xue, “Spectral nonlocal restoration of hyperspectral images with low-rank property,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 8, no. 6, pp. 3062–3067, Jun. 2015.

[23] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, “Nonlocal patch tensor sparse representation for hyperspectral image super-resolution,” IEEE Trans. Image Process., vol. 28, no. 6, pp. 3034–3047, Jun. 2019.

[24] 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.

[25] Q. C. Qiao, Z. Lu, T.-Z. Huang, and J. Zhou, “Denoising and Completion of 3D data via multilinear tools for hyperspectral images denoising by intrinsic tensor sparsity regularization,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 1692–1700.

[26] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” J. Fourier Anal. Appl., vol. 14, no. 5, pp. 877–905, 2008.

[27] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng, “Patch group based nonlocal self-similarity prior learning for image denoising,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Santiago, Chile, Dec. 2015, pp. 244–252.

[28] M.-D. Iordache, J. M. Bioucas-Dias, A. Plaza, and B. Somers, “MUSIC-CSR: Hyperspectral unmixing via multiple signal classification and collaborative sparse regression,” IEEE Trans. Geosci. Remote Sens., vol. 56, no. 10, pp. 4364–4382, Jul. 2014.

[29] R. Wang, H.-C. Li, A. Pizurica, J. Li, A. Plaza, and W. J. Emery, “Hyperspectral unmixing using double reweighted sparse regression and total variation,” IEEE Geosci. Remote Sens. Lett., vol. 14, no. 7, pp. 1146–1150, Jul. 2017.

[30] P. Gong, C. Zhang, Z. Lu, J. Zhang, and J. Ye, “A general iterative shrinking and thresholding algorithm for non-convex regularized optimization problems,” in Proc. ICML, 2013, pp. 37–45.

[31] 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.

GUIXI WANG was born in Chengdu, Sichuan, China, in 1985. He received the B.S. and M.S. degrees in resources and environment from Sichuan Agricultural University, Ya’an, China, in 2009 and 2012, respectively, and the Ph.D. degree in earth exploration and information techniques from the Chengdu University of Technology, Chengdu, China, in 2017. Since 2015, he has been a Research Associate with NIMTIT, Chengdu. His research interests include image processing, environmental monitoring, and analysis.

HONGWEI HAN received the B.S. degree in mathematics education from Yan’an University, Yan’an, in 2002, and the M.S. degree in electronics and communication engineering from the Chengdu University of Technology, Chengdu, in 2009, where he is currently pursuing the Ph.D. degree in geomatics. His research interests include hyperspectral remote sensing image processing and partial differential equation and its application.

VOLUME 8, 2020
EMMANUEL JOHN M. CARRANZA was born in Legazpi City, Philippines, in 1962. He started his career as an Exploration Geologist/Geochemist at Mines and Geosciences Bureau, Philippines, from 1983 to 2001. He was a Researcher, from 2001 to 2003, and an Associate Professor with the Department of Earth Systems Analysis, International Institute of Geo-Information Science and Earth Observation, The Netherlands, from 2003 to 2012. He was an Associate Professor, from 2013 to 2016, and has been an Adjunct Associate Professor with the School of Earth and Environmental Sciences, James Cook University, Australia, since 2017. He was a Visiting Professor with the Institute of Geosciences, State University of Campinas, Brazil, from 2015 to 2017. He has also been a Professor of geological sciences with the University of KwaZulu-Natal, South Africa, since 2017. His fields of expertise are geochemistry (for mineral exploration, ore genesis, and geo-environmental studies), spatial mathematics/statistics (for predictive modeling of mineral resources and geological hazards), and remote sensing for geological/mineral exploration. He has written a book Geochemical Anomaly and Mineral Prospectivity Mapping in GIS. He is the Editor-in-Chief of Natural Resources Research journal and an Associate Editor of Ore Geology Reviews, the Journal of Geochemical Exploration, and Geochemistry: Exploration, Environment, Analysis.

SI GUO was born in Chengdu, Sichuan, China, in 1987. She received the B.S. degree in geophysics and the M.S. and Ph.D. degrees in sedimentology from the Chengdu University of Technology, Chengdu, in 2009, 2011, and 2015, respectively. Since 2009, she has been an Engineer with the Reservoir Prediction and CNPC Chuanqing Drilling Engineering Company.

KE GUO is currently a Professor and the Ph.D. Supervisor with the Chengdu University of Technology. He is an Academic and the Technical Leader of Sichuan, and enjoys special subsidies from the State Council Government. He is the Head of the Teaching Team of mathematics geology in Sichuan and the Scientific Research and Innovation Team of high-level resources and environment in Sichuan. He has long been engaged in the teaching and scientific research of quantitative evaluation and prediction of resources and environment.

KEYAN XIAO is currently a Researcher with the Institute of Mineral Resources, Chinese Academy of Geological Sciences, and the Director of the Metallogenic Prospect Division of this institute. He is a Foreign Academician of the Russian Academy of Natural Sciences. He is also the Deputy Director of the MIR Key Laboratory of Metallogeny and Mineral Resource Assessment, Institute of Mineral Resources of China. He is a Distinguished Professor of Chutian Scholar with Yangtze University, an Adjunct Professor with the China University of Geosciences and the China University of Mining and Technology, and a leading scientific and technological talent of Ministry of Natural Resources of China. His fields of expertise are mathematical geology and evaluation of mineral resources. He is an Editorial Board Member of Acta Geoscientica Sinica and Mineral Deposits and a Guest Editor of Ore Geology Reviews.