Abstract—In this article, we propose a novel nonconvex approach to robust principal component analysis (RPCA) for hyperspectral image (HSI) denoising, which focuses on simultaneously developing more accurate approximations to both rank and columnwise sparsity for the low-rank and sparse components, respectively. In particular, the new method adopts the log-determinant rank approximation and a novel $\ell_2\text{-log}$ norm, to restrict the local low-rank or columnwise sparse properties for the component matrices, respectively. For the $\ell_2\text{-log}$-regularized shrinkage problem, we develop an efficient, closed-form solution, which is named $\ell_2\text{-log}$-shrinkage operator. The new regularization and the corresponding operator can be generally used in other problems that require columnwise sparsity. Moreover, we impose the spatial–spectral total variation regularization in the log-based nonconvex RPCA model, which enhances the global piecewise smoothness and spectral consistency from the spatial and spectral views in the recovered HSI. Extensive experiments on both simulated and real HSIs demonstrate the effectiveness of the proposed method in denoising HSIs.

Index Terms—Hyperspectral image (HSI), low rank, robust principal component analysis (RPCA), sparse.

Manuscript received 25 December 2021; revised 1 May 2022 and 20 July 2022; accepted 14 August 2022. Date of publication 15 September 2022. Date of current version 3 October 2022. This work was supported by the National Science Natural Foundation of China (NSFC) in part by the Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies, Shenzhen 518055, China, and also with the Shandong Province Colleges and Universities Youth Innovation Plan under Grant 2019KJ062 and Grant 2020KJN011; in part by the Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies under Grant 2022B1212010005; and in part by the National Institutes of Health (NIH) under Grant R21AG070909 and Grant U3H NS100606-03. (Corresponding author: Chenglizhao Chen.)

Cheng Peng is with the School of Computer Science and Engineering, Qingdao University, Qingdao 266071, China (e-mail: czl123@163.com).

Yangyong Chen is with the College of Computer Science and Technology, Harbin Institute of Technology, Shenzhen 518055, China, and also with the Guangdong Provincial Key Laboratory, Novel Security Intelligence Technologies, Shenzhen 518055, China.

Xinxing Wu is with the Department of Computer Science, University of Kentucky, Lexington, KY 40506 USA.

Zhao Kang is with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China.

Qiang Cheng is with the Institute of Biomedical Informatics and the Department of Computer Science, University of Kentucky, Lexington, KY 40506 USA.

Digital Object Identifier 10.1109/TGRS.2022.3206783

I. INTRODUCTION

HYPERSPECTRAL imaging is widely used in various applications, such as biomedical imaging, terrain classification, military surveillance, and remote sensing [25], [26], [44], [61], [65]. Despite the broad applications of hyperspectral images (HSIs), clean HSIs are rarely obtained due to unavoidable corruptions by mixed types of noise, such as Gaussian noise, impulse noise, deadlines, and stripes, in the acquisition process [58]. Thus, the heavy noise makes it challenging to process HSIs in various applications, such as classification [31] and unmixing [28]. Due to the adverse effects of noise, there is a pressing need for developing effective algorithms to remove noise from HSIs as a preprocessing step of further HSI applications for enhanced learning performance.

HSIs contain hundreds of bands sampled from the visible and infrared range of the electromagnetic spectrum, forming a three-order tensor structure similar to RBG images with three channels. In this sense, HSIs can be regarded as extensions of RBG images [20]. HSIs contain three dimensions, including two spatial dimensions (along and across the track) and one spectral dimension (wavelength) [7]. In fact, each band of an HSI can be regarded as a grayscale image, and existing image denoising algorithms can be readily adopted independently in a bandwise manner to remove noise from HSIs [17]. It should be noted that there is a stark difference between the tasks of denoising single grayscale images and HSIs. Generally, different bands in an HSI usually have high correlations and spectral redundancies, which do not exist in grayscale images. Unfortunately, simply applying grayscale image denoising algorithms to HSI in a bandwise manner ignores the high correlations between different bands, which omits the special structural information of HSI and thus leads to unsatisfactory denoising performance. Thus, despite the great success of grayscale image denoising algorithms [2], [12], [17], [40], there is still an essential demand in designing specialized denoising methods for HSIs due to their particular characteristics.

In the last decades, a number of HSI denoising methods have been developed [34], [41], [67]. For example, Du et al. [67] proposed a unified probabilistic framework, in which the spatial and spectral dependencies are simultaneously adopted; Elad and Aharon [41] proposed to take both nonlocal similarity and spectral–spatial structure of HSIs into consideration in a sparse representation-based framework. Moreover, methods,
such as principal component analysis (PCA), wavelet shrinkage, anisotropic diffusion, multitask sparse matrix factorization, and tensor decomposition, have been considered for HSI denoising [8], [9], [16], [48], [54], [56], [66]. Most of the above methods require some specific prior knowledge of the noise. Unfortunately, such knowledge is rarely available and generally limited in real-world applications, and thus, the above methods cannot remove all types of noise from HSIs. Hence, more effective methods are urgently in demand.

Recently, low-rank techniques have been developed and shown successfully in various low-rank recovery applications, such as low-rank matrix factorization, low-rank subspace learning, and HSI denoising [4], [35], [37], [38], [39], [46], [50]. For HSI denoising, various low-rank approaches have been attempted, such as low-rank matrix factorization [15], [54], low-rank tensor decomposition [18], [59], low-rank dictionary learning [19], and low-rank regularization methods [55]. The success of these methods in HSI denoising is based on a natural assumption that a scene of a clean HSI is composed of much fewer endmembers than spectral bands and pixels, which reveals the natural low-rank structures of HSI [1], [21]. Thus, low-rank-based techniques have been widely adopted to remove noise from HSIs, among which PCA [6] and low-rank matrix factorization-based methods [10] are typical ones. Unfortunately, these methods are known to suffer from being sensitive to outliers that commonly exist in HSIs [3]. Thus, it is generally difficult for these methods to completely remove outliers, such as stripes, deadlines, and impulse noise from HSIs. To combat this issue, robust principal component analysis (RPCA) models the outliers by separating a sparse component, which significantly improves the robustness. In the original paper of RPCA [3], it is proven that under certain conditions, there is a high probability to correctly separate the low-rank and sparse components from the observed data.

Traditional low-rank matrix recovery (LRMR) models, such as RPCA, adopt the nuclear norm to restrict the low-rank property of the target matrix. Recently, it has been pointed out that the nuclear norm as used in the RPCA model is not accurate in approximating the rank function [10], [24], [36], [51], which may lead to degraded performance in low-rank recovery [35], [36]. To solve this issue, nonconvex approaches are attempted to better approximate the rank function in the RPCA framework [35], which achieves promising performance. Despite the success of nonconvex rank approximations, nonconvex approximations of the sparsity are rarely considered in the RPCA model. In this article, we point out that there is a close connection between the nuclear norm and the $\ell_{2,1}$ norm, i.e., the rank and columnwise sparsity approximations. For nonconvex rank approximations, the improved approximating behaviors actually benefit from the improved approximation to the sparsity of the singular values, which inspires us to improve the approximating behavior of the columnwise sparsity in the RPCA framework with nonconvex approach for more accurate low-rank and sparse components separation in HSI denoising [51]. In particular, we propose a log-based columnwisely sparse approximation, namely, the $\ell_{2,\log}$ norm, that admits some favorable properties, which will be discussed with details in Sections III and V.

While low-rank recovery models are effective in HSI denoising, they only explore the correlation between spectral bands of HSIs with the low-rank constraints while omitting the spatial correlation of local neighboring pixels. Several approaches have been attempted in low-rank models to incorporate spatial information of HSIs, such as wavelets [8], [42], total variation (TV) regularization [23], [47], and sparse representation [65], [68]. These methods follow a common strategy by restricting low-rank structure on the overall HSI with specialized regularization constraints imposed on it. However, the same material from different local areas of HSI may have starkly different spectral signatures, which leads to the increased rank of the overall HSI. Meanwhile, local areas are likely to contain the same material and thus the same spectral signature, which implies the local low-rank property of HSI. This inspires the segmentation of the HSI into overlapping 3-D patches [53], [58], where they are processed sequentially with the RPCA model. Oftentimes, the sparse noise exists in the same location of some bands and naturally forms a local low-rank structure. Such noise is considered as structured sparse noise and is often mathematically treated as part of the low-rank component by spectral low-rank property. Consequently, it is difficult for local low-rank models to remove such noise only with spectral correlation of HSI and it is demanding to exploit spatial constraint for improved denoising performance. In fact, clean HSI favors global piecewise smoothness and spectral consistency from the spatial and spectral viewpoints, respectively, which are destroyed by the structured sparse noise. Thus, in this article, we follow [20] and simultaneously seek the separation of low-rank and sparse matrices with spatial–spectral TV (SSTV) for enhanced spatial–spectral piecewise smoothness and consistency. It is noted that the log-based nonconvex approximations to the rank and columnwise sparsity ensure that the low-rank and sparse components can be more accurately separated than traditional methods. Meanwhile, the SSTV term enhances the global piecewise smoothness and spectral consistency of the recovered HSI to remove potential noise remaining in the low-rank component.

We summarize the key contributions of this article as follows.

1) We propose a nonconvex columnwise sparsity approximation named the $\ell_{2,\log}$ norm, which is more accurate than the widely adopted convex $\ell_{2,1}$ norm in approximating the columnwise sparsity. With the new approximation, we propose a novel nonconvex RPCA model with simultaneous log-based nonconvex approximations to the rank and columnwise sparsity.

2) For the $\ell_{2,\log}$-norm regularized shrinkage problem, we formally provide an efficient closed-form solution. The $\ell_{2,\log}$-norm regularized shrinkage problem and its solution can be generally adopted in various problems that restrict columnwise or rowwise sparsity, such as robustness learning and feature selection.

3) Elegant theoretical analysis results, such as perturbation analysis and boundedness analysis, are provided
to guarantee the desirable properties of the proposed model, which are essentially important for unsupervised learning methods.

4) Superior performance is observed compared with state-of-the-art baseline methods, which confirms the effectiveness of the proposed method.

II. RELATED WORK

In this section, we will briefly review a few techniques that are closely related to our work.

A. Robust Principal Component Analysis

Given data matrix $X$, RPCA assumes that the data can be decomposed into a low rank $L$ and a sparse $S$, which can be mathematically formed as $X = L + S$. To obtain the two components, the classic RPCA aims at solving the following constrained optimization problem [3]:

$$
\min_{L,S} \|L\|_* + \lambda \|S\|_1, \quad \text{s.t. } X = L + S
$$

where $\| \cdot \|_*$ is the nuclear norm that adds all singular values of the input matrix, $\| \cdot \|_1$ is the $\ell_1$ norm that adds the absolute values of all elements of the input matrix, and $\lambda \geq 0$ is a balancing parameter.

B. Spatial–Spectral TV

For HSI, it is natural that two nearby bands are very similar, which indicates spectral consistency. Also, HSIs have spatial correlations. Thus, it is convincing to adopt the TV norm from both spatial and spectral directions for HSI. For an observed tensor cube $\mathcal{M}$, we denote its $(i,j,k)$th element by $(\mathcal{M})_{i,j,b}$, where $i$ and $j$ represent the horizontal and vertical directions, respectively, and $b$ corresponds to the spectral direction. Then, the anisotropic SSTV norm can be formulated as

$$
\|\mathcal{M}\|_{\text{SSTV}} = \|D_x\mathcal{M}\|_1 + \|D_y\mathcal{M}\|_1 + \|D_z\mathcal{M}\|_1
$$

where $D_x$, $D_y$, and $D_z$ perform the first-order discrete differences of $\mathcal{M}$ in the horizontal, vertical, and spectral directions, respectively, which are defined as

$$
\begin{align*}
D_x\mathcal{M} & = \text{vec}([(\mathcal{M})_{i+1,j,b} - (\mathcal{M})_{i,j,b}]) \\
D_y\mathcal{M} & = \text{vec}([(\mathcal{M})_{i,j+1,b} - (\mathcal{M})_{i,j,b}]) \\
D_z\mathcal{M} & = \text{vec}([(\mathcal{M})_{i,j,b+1} - (\mathcal{M})_{i,j,b}])
\end{align*}
$$

with periodic boundary conditions. Here, vec$(\cdot)$ is a linear operator that reshapes a tensor into a vector. With (2), the piecewise smoothness is restricted in both the spatial and spectral directions. It is pointed out that the equal weights for the gradients along different dimensions might not be proper and (2) is further extended to the following anisotropic SSTV regularization:

$$
\|\mathcal{M}\|_{\text{SSTV}} = \tau_x \|D_x\mathcal{M}\|_1 + \tau_y \|D_y\mathcal{M}\|_1 + \tau_z \|D_z\mathcal{M}\|_1
$$

where $\tau_x$, $\tau_y$, and $\tau_z$ are balancing parameters. In this article, we set $[\tau_x, \tau_y, \tau_z] = [1, 1, 0.5]$ as recommended in [9] and [51].

C. Deep Convolutional Neural Networks for HSI Denoising

Recently, convolutional neural networks (CNNs) have been widely adopted in image denoising tasks. For example, the CNN method extracts the intrinsic and different image features and obtains the state-of-the-art performance in natural image denoising [60]. The trainable nonlinear reaction–diffusion (TNRD) [11] is successful for natural images and is further extended for HSI denoising in [52]. However, these methods do not fully exploit the characteristics of spectral redundancy in HSI data. To account for spectral–spatial information in HSIs, various algorithms are developed. For example, a spatial–spectral deep residual CNN is developed in [57] with multiscale and multilevel feature representation; a modified 3-D U-net architecture with rich multiscale information of HSIs encoding is proposed to exploit spatial–spectral correlations in HSIs [14]; a 3-D atrous denoising CNN is developed for HSI denoising, which extracts feature maps along both spatial and spectral dimensions [32]; a novel framework focuses on non-independent and identically distributed (i.i.d.) noise removal from HSIs a deep spatiotemporal Bayesian posterior (DSSBP) structure [64]; and a spatial–spectral gradient learning strategy is adopted in CNN to extract intrinsic and deep features of HSIs [63]. Most of the deep learning methods as mentioned above are based on ResNet [32], [57], U-Net [14], or plain CNN [63], [64]. Although these methods obtain promising performance in HSI denoising, they only exploit local spectral and spatial information from HSIs, which is still inefficient in fully exploiting global and local spectral–spatial information of HSIs. Thus, it is still in urgent need of developing more efficient methods for noise removal from HSIs.

III. SSTV REGULARIZED NONCONVEX RPCA

For the observed HSI $O \in \mathbb{R}^{M \times N \times p}$, it is natural to separate the clean part $L \in \mathbb{R}^{M \times N \times p}$ and noise part $S \in \mathbb{R}^{M \times N \times p}$ as $O = L + S$. Due to the nature of HSIs, usually, the adjacent bands are highly correlated, which leads to the low-rank structure. For HSIs, the spectral signatures of the same local area are more likely to be the same, which inspires us to exploit the local low-rank property of the HSIs. Thus, we divide the HSI into overlapping patches and exploit the patchwise local-low rank structure. Specifically, for the tensor $L$, we first find an $m \times n \times p$ patch cube centralized at location $(i,j)$. Then, we vectorize all patch bands and form its corresponding Casorati matrix $L_{i,j} \in \mathbb{R}^{mn \times p}$, with each column of $L_{i,j}$ being a vectorized patch band. Similarly, we define the Casorati matrices $O_{i,j}$ and $S_{i,j}$ from $O$ and $S$. With these definitions, the target is to separate low rank $L_{i,j}$ and sparse $S_{i,j}$ from the observed matrix $O_{i,j}$. To keep the spatial structural information of $S_{i,j}$, we adopt the $\ell_{2,1}$ norm in the low-rank and sparse separation model [3], [29], which leads to the following patch-based RPCA model [20]:

$$
\min_{L,S} \sum_{i} \sum_{j} \left\{ \|L_{i,j}\|_* + \lambda \|S_{i,j}\|_{2,1} \right\} \\
\text{s.t. } O_{i,j} = L_{i,j} + S_{i,j}
$$

where $\|S_{i,j}\|_{2,1} = \sum_{b=1}^{p} \|S_{i,j}(b)\|_2$ is the $\ell_{2,1}$ norm that restricts columnwisely sparse structure for $S_{i,j}$ with $\| \cdot \|_2$ being
where $A = [\ldots, a_j, \ldots]$ is the input matrix and $\| \cdot \|_2$ is the $\ell_2$ norm. The $\ell_{2, \log}$ norm admits the following properties.

1) $\| A \|_{2, \log} \geq 0$ for any matrix $A$ and $\| A \|_{2, \log} = 0$ if and only if $A = 0$.

2) The $\ell_{2, \log}$ norm is nonconvex, continuous, and differentiable for nonzero columns, where

\[
\frac{\partial}{\partial A} \| A \|_{2, \log} = \begin{bmatrix}
\| a_1 \|_2 (1 + \| a_1 \|_2) & \cdots & a_j \\
\| a_j \|_2 (1 + \| a_j \|_2) & \cdots & \| a_n \|_2 (1 + \| a_n \|_2)
\end{bmatrix}
\]

3) For $\| a_j \|_2 \neq 0$, $\sum_j \log(1 + \| a_j \|_2) < \sum_j \| a_j \|_2$, which implies that the $\ell_{2, \log}$ norm is more accurate than the $\ell_{2, 1}$ norm in approximating the sparsity for large values and noise effects for small values.

4) As will be clear in the remark of this section, the $\ell_{2, \log}$ norm has smaller expectation than the $\ell_2$ norm and is better in estimating noise effects.

5) As will be clear in Section V, the triangle inequality holds for $\ell_{2, \log}$.

Thus, with the $\ell_{2, \log}$ norm, (7) is further developed into

\[
\min_{L, S} \sum_{i,j} \log \det \left( I + (L_{i,j}^T L_{i,j})^{1/2} \right) + \lambda \sum_{i,j} \log(1 + \| S_{i,j} \|_2)
\]

\[
\text{s.t. } O_{i,j} = L_{i,j} + S_{i,j}.
\]

(9)

In the above model, the sparse term helps remove sparse noise from the data, while the low-rank term helps retain the useful structural information that is highly correlated among different bands. It is seen that (9) is a patch-based model, which learns the local low-rank property of HSIs. Although patch-based low-rank recovery methods are successful in HSI denoising [13], [53], [58], they fail to exploit global structural information, i.e., correlations of spatial pixels and spectral bands of HSIs, which may lead to failure in noise removal [20]. To address this issue, we integrate the SSTV with (9) to seek local low-rank and TV properties in both spatial and spectral domains and obtain the following model:

\[
\min_{L, S} \sum_{i,j} \left\{ \| L_{i,j} \|_{\log \det} + \lambda \| S_{i,j} \|_{2, \log} \right\} + \gamma \| L \|_{\text{SSTV}}
\]

\[
\text{s.t. } O_{i,j} = L_{i,j} + S_{i,j}.
\]

(10)

where $\gamma \geq 0$ is a balancing parameter. We name the model in (10) the log-based local low-rank and sparse separation model with spatial–spectral total variation (L$^3$S$^3$TV). It is seen that the L$^3$S$^3$TV model exploits both local and global structures of HSIs, where the local low-rank property of HSIs is sought with the first two terms, and the correlations of spatial pixels and spectral bands are exploited with the SSTV term. For optimization, we will develop an efficient algorithm in Section IV.

Remark: For the $\ell_{2, \log}$ norm, we have the following conclusions. The following discussions can be generalized to the
other columns and the overall matrix. For a specific column of a matrix, we denote the elements by \(X_1, \ldots, X_d\). Then, for any distributions of \(X_i\)'s, the expectation of the log-based approximation is generally less than the \(\ell_2\)-based approximation, which can be formally analyzed in the following:

\[
E\left(\log \left(1 + \frac{1}{\sum_{i=1}^{d} X_i^2}\right)\right) = \int_0^{+\infty} \log(1 + \sqrt{y}) f_{\sum_{i=1}^{d} X_i^2}^{\ell_2}(y) dy
\]

where \(f_{\sum_{i=1}^{d} X_i^2}^{\ell_2}(y)\) is the probability density function for \(y = \sum_{i=1}^{d} X_i^2\).

Moreover, for essentially small values of \(\sum_{i=1}^{d} X_i^2\), it is natural that such values correspond to noise and the corresponding columns are indeed sparse. Thus, for such small values, it is essentially important that the approximation is close to 0 rather than 1 to distinguish noise effects and useful information. It is noted that \(\log(1 + \sqrt{x}) < \sqrt{x}\) holds for small \(x\), which indicates that the log-based approximation is closer to 0 than the \(\ell_2\)-based approach and thus is more accurate in approximating the real sparsity. Thus, it is expected that the log-based approximation is more accurate in approximating the real sparse indicator of the columns than the \(\ell_2\)-based approach.

\[\text{IV. Optimization}\]

In this section, we will develop an efficient optimization algorithm for (10) based on the augmented Lagrange multiplier (ALM) method. In particular, we first introduce some auxiliary variables to (10) and obtain the following equivalent model:

\[
\min_{C, Z, B, S, A} \sum_{i,j} \left\{ \|C_{i,j}\|_{\log \det} + \lambda \|S_{i,j}\|_{2, \log} + g \|C\|_{\text{SSTV}} \right\} + \frac{\rho}{2} \left\|L_{i,j} - A_{i,j} + \frac{1}{\rho} Z_{C_{i,j}} \right\|_F^2
\]

subject to \(O_{i,j} = L_{i,j} + S_{i,j}\), \(L_{i,j} = A_{i,j}\), \(A = B\), \(C = DB\)

(12)

where \(D = [\tau_1 D_1, \tau_2 D_2, \ldots, \tau_d D_d]\) denotes the TV operator in the spatial and spectral directions and \(DB = [\tau_1 D_1 B, \tau_2 D_2 B, \ldots, \tau_d D_d B] \in \mathcal{R}^{MNP \times 3}\).

Then, we need to optimize the augmented Lagrange function as follows:

\[
\min_{C, Z, B, S, A, Z^o, Z^b, Z^c} \left( \sum_{i,j} \left\{ \|C_{i,j}\|_{\log \det} + \lambda \|S_{i,j}\|_{2, \log} + \frac{\rho}{2} \left\|L_{i,j} - A_{i,j} + \frac{1}{\rho} Z_{C_{i,j}} \right\|_F^2 \right\} + \frac{\rho}{2} \left\|O_{i,j} - L_{i,j} - S_{i,j} + \frac{1}{\rho} Z_{O_{i,j}} \right\|_F^2 \right) + \frac{\rho}{2} \left\|A_{i,j} - \frac{1}{\rho} Z_{A_{i,j}} \right\|_F^2
\]

(11)

where \(C, Z, B, S, A, Z^o, Z^b, Z^c\) and \(\mathcal{M}\) denote \(\{\sum_{a,b,c} (M_{a,b,c}^2)_{a,b,c}\}^{1/2}\) for tensor \(\mathcal{M}\) for ease of notation. Next, we will develop the alternating optimization strategies for each variable.

\[\text{A. } L_{i,j}\text{-Minimization}\]

To optimize \(L_{i,j}\), we have the following subproblem:

\[
\min_{L_{i,j}} \sum_{i,j} \left\{ \|C_{i,j}\|_{\log \det} + \frac{\rho}{2} \left\|L_{i,j} - A_{i,j} + \frac{1}{\rho} Z_{C_{i,j}} \right\|_F^2 \right\} + \frac{\rho}{2} \left\|O_{i,j} - L_{i,j} - S_{i,j} + \frac{1}{\rho} Z_{O_{i,j}} \right\|_F^2.
\]

(15)

The above problem can be solved for each \(L_{i,j}\) independently with

\[
\min_{L_{i,j}} \|L_{i,j}\|_{\log \det} + \frac{\rho}{2} \left\|L_{i,j} - A_{i,j} + \frac{1}{\rho} Z_{C_{i,j}} \right\|_F^2 + \frac{\rho}{2} \left\|O_{i,j} - L_{i,j} - S_{i,j} + \frac{1}{\rho} Z_{O_{i,j}} \right\|_F^2.
\]

(16)

If

\[
\mathcal{X}_{i,j} = \frac{O_{i,j} - S_{i,j} + \frac{1}{\rho} Z_{O_{i,j}} + A_{i,j} - \frac{1}{\rho} Z_{A_{i,j}}}{2}
\]

then with straightforward algebra, the problem (16) is equivalent to

\[
\min_{L_{i,j}} \frac{1}{2\rho} \|L_{i,j}\|_{\log \det} + \frac{1}{2} \left\|L_{i,j} - \mathcal{X}_{i,j}\right\|_F^2.
\]

(17)

For a matrix \(D\), we define \(P(D), Q(D), \) and \(\sigma_i(D)\) to be its left and right singular vectors and the \(i\)th largest singular value, respectively. Then, similar to [27] and [41], (17) admits a closed-form solution with the following operator:

\[
L_{i,j} = D_{\frac{\sigma_i(D)}{\sigma_{i+1}(D)}}(\mathcal{X}_{i,j})
\]

(18)

where \(D_{\sigma_i(D)} = P(D) \text{diag}[\sigma_i(D)](Q(D))^T\), with

\[
\sigma_i^* = \begin{cases} 
\zeta', & \text{if } f_i(\zeta) \leq f_i(0) \text{ and } (1 + \sigma_i(D))^2 > 4\delta \\
0, & \text{otherwise}
\end{cases}
\]

(19)

where

\[
f_i(x) = \frac{1}{2} (x - \sigma_i(D))^2 + \tau \log(1 + x)
\]

and

\[
\zeta = \frac{\sigma_i(D) - 1}{2} + \sqrt{(1 + \sigma_i(D))^2 - 4\delta}.
\]
B. $S_{i,j}$-Minimization

The subproblem associated with optimization of $S_{i,j}$ is

$$\min_{\bar{S}_{i,j}} \sum_{i,j} \lambda \|S_{i,j}\|_{2,\log} + \frac{\rho}{2} \left\|O_{i,j} - L_{i,j} - S_{i,j} + \frac{1}{\rho} Z_{i,j}^0 \right\|_F^2$$

(20)

which can be solved in an elementwise manner with

$$\min_{\bar{S}_{i,j}} \frac{\lambda}{\rho} \|S_{i,j}\|_{2,\log} + \frac{1}{2} \left\|O_{i,j} - L_{i,j} - S_{i,j} + \frac{1}{\rho} Z_{i,j}^0 \right\|_F^2.$$  
(21)

Problem in a format of (21) is an $\ell_2,\log$-regularized shrinkage problem. We formally give Theorem 1 to solve it.

**Theorem 1 ($\ell_2,\log$-Shrinkage Operator):** Given matrix $Y \in \mathbb{R}^{d \times n}$ and a nonnegative parameter $\alpha$, the following problem:

$$\min_{\bar{w} \in \mathbb{R}^{d \times n}} \frac{1}{2} \|Y - W\|_F^2 + \alpha \|W\|_{2,\log}$$

(22)

is called $\ell_2,\log$-regularized shrinkage problem, which admits a closed-form solution in a columnwise manner

$$w_i = \begin{cases} \frac{f_i(\xi)}{\sqrt{\sum_{j \neq i} (s_j^u)^2 - \xi}}, & \text{if } f_i(\xi) \leq \frac{4}{\sqrt{n}}, \\
0, & \text{otherwise} \end{cases}$$

(23)

where $f_i(x) = (1/2)(x - \|y_{i}\|_2)^2 + \alpha \log(1 + x)$, and $\xi = (\|y_{i}\|_2 - 1/2) + ((1 + \|y_{i}\|_2)/4) - \alpha/2$.

**Proof:** It is easy to see that (22) can be solved with respect to each $w_i$ independently. For the subproblem is

$$\min \frac{1}{2} \|y_i - w_i\|_2^2 + \alpha \log(1 + \|w_i\|_2).$$

We may treat $w_i$ as a special matrix and perform thin singular value decomposition (SVD) to it. Then, it is seen that $w_i$ has exactly one singular value, which is $\sigma(w_i) = (w_i^T w_i)^{1/2} = \|w_i\|_2$, where $\sigma(\cdot)$ is the singular value of the input vector. Thus, optimizing $w_i$ is equivalent to

$$\min \frac{1}{2} \|y_i - w_i\|_2^2 + \alpha \log(1 + \sigma(w_i)).$$

(24)

Hence, according to [27] and [41] and Section IV-A, the solution to (24) is obtained with $w_i = u_i \sigma^*(w_i)v_i^T$, where $u_i$ and $v_i$ are left and right singular vectors of $y_i$, respectively, and

$$\sigma^*(w_i) = \begin{cases} \frac{\xi}{\sqrt{\sum_{j \neq i} (s_j^u)^2 - \xi}}, & \text{if } f_i(\xi) \leq f_i(0), \\
0, & \text{otherwise} \end{cases}$$

(25)

with

$$f_i(x) = \frac{1}{4}(x - \sigma(y_i))^2 + \alpha \log(1 + x)$$

and

$$\xi = \frac{\sigma(y_i) - 1}{2} + \sqrt{\frac{(1 + \sigma(y_i))^2}{4} - \alpha}.$$

If $y_i = 0$, then $w_i = 0$ is clearly the optimal solution and (23) is true. Thus, it suffices to consider $y_i \neq 0$. In this case, it is straightforward that $y_i = (y_i/\|y_i\|_2)(y_i/\|y_i\|_2)^T$ is a thin SVD of $y_i$. Here, the notation $[1]$ represents a special row matrix with only one column that is 1. We substitute $u_i = (y_i/\|y_i\|_2)$.

$\sigma(y_i) = \|y_i\|_2$, and $v_i = [1]$ into the above equations, which leads to (23) and concludes the proof.

For ease of notation, we denote the $\ell_2,\log$-shrinkage operator of (23) as $T_{\log}(Y)$, generating the solution to (20)

$$S_{i,j} = T_{\log}(O_{i,j} - L_{i,j} + Z_{i,j}^0/\rho).$$

(26)

It is seen that the log-based shrinkage problem admits a closed-form solution, which is more efficient than the existing approaches that adopt iterative optimization strategies such as [39] and [44].

C. $A$-Minimization

The subproblem associated with optimization of $A$ is

$$\min_{\bar{A}} \sum_{i,j} \alpha \|L_{i,j} + A_{i,j} + \frac{1}{\rho} Z_{i,j}^0\|_F^2 + \rho \|A - B + \frac{1}{\rho} Z^B\|_F^2.$$  
(27)

We define $1_{i,j}$ to be an indicator function, which returns 1 if the condition in the subscript is satisfied and 0 otherwise. Then, the above problem for $A$ can be rewritten in an elementwise manner for each $(A)_{a,b,c}$ as follows:

$$\arg \min_{(A)_{a,b,c}} \left( (A)_{a,b,c} - (B)_{a,b,c} + \frac{1}{\rho} (Z^B)_{a,b,c} \right)^2 + \left( (L)_{a,b,c} - (A)_{a,b,c} + \frac{1}{\rho} (Z^A)_{a,b,c} \right)^2 \times \sum_{i,j} 1_{(A)_{a,b,c} = (A)_{a,b,c}}.$$  
(28)

where $\sum_{i,j} 1_{(A)_{a,b,c} = (A)_{a,b,c}}$ counts the number of times that $(A)_{a,b,c}$ is overlapped in (27). It is seen that the problem is quadratic in $(A)_{a,b,c}$, which admits a closed-form solution with the following first-order optimality condition. Thus, we have the following closed-form solution for $A$:

$$(A)_{a,b,c} = \frac{1}{1 + \sum_{i,j} 1_{(A)_{a,b,c} = (A)_{a,b,c}}} \left( (B)_{a,b,c} - \frac{1}{\rho} (Z^B)_{a,b,c} + \sum_{i,j} 1_{(A)_{a,b,c} = (A)_{a,b,c}} \left( (L)_{a,b,c} + \frac{1}{\rho} (Z^A)_{a,b,c} \right) \right).$$

(29)

D. $B$-Minimization

The subproblem associated with optimization of $B$ is

$$\min \frac{\rho}{2} \left\|C - DB + \frac{1}{\rho} Z^C\right\|_F^2 + \frac{\rho}{2} \left\|A - B + \frac{1}{\rho} Z^B\right\|_F^2$$

(30)

which can be solved with the following equation:

$$(D^T D + I)B = D^T (C + Z^C/\rho) + (A + Z^B/\rho)$$

(31)

which can be efficiently solved by the fast Fourier transform (FFT)

$$B = F^{-1} \left[ \frac{F(D^T (C + Z^C/\rho) + (A + Z^B/\rho))}{1 + F(\tau, D)^2 + F(\tau, D)^2 + F(\tau, D)^2} \right].$$

(32)
Algorithm 1 HSI Restoration via \( L^3S^2TV \) Model

Input: Observed HSI \( O \in \mathbb{R}^{M \times N \times p} \), patch size \( m \times n \), stopping criterion \( \epsilon \), maximum number of iterations \( t_{\text{max}} \), balancing parameters \( \lambda, \gamma, \tau \), parameters \( \rho_{\text{max}}, \rho, \kappa \).

1. Initialize: \( \mathcal{L}^{(0)}, S^{(0)}, A^{(0)}, B^{(0)}, (Z^{(0)}), (Z^{(0)}), (Z^{(0)}), (Z^{(0)}), e^{(0)}, (Z^{(0)}), t = 1 \).
2. Repeat
3. Update all patches \( ((\mathcal{L}_{i,j})^{(t)}), (S_{i,j})^{(t)} \) by Eq. (18) and Eq. (26), respectively;
4. Update \( (Z^{(t)}), (Z^{(t)}), C^{(t)} \) by Eq. (29), (32) and (34), respectively;
5. Update \( (Z^{(t)}), (Z^{(t)}), (Z^{(t)}), (Z^{(t)}), (Z^{(t)}), (Z^{(t)}), (Z^{(t)}), (Z^{(t)}), (Z^{(t)}), e^{(t)}, (Z^{(t)}), t = 1 \).
6. Update \( \rho^{(t)} := \min(\rho^{(t-1)}, \kappa \rho_{\text{max}}) \).
7. Check the convergence condition: \( \max \{ \|O_{i,j} - \tilde{L}_{i,j}^{(t)} - S_{i,j}^{(t)}\|_{\infty}, \|L^{(t)} - A^{(t)}\|_{\infty}, \|A^{(t)} - B^{(t)}\|_{\infty}, \|C^{(t)} - D^{(t)}\|_{\infty} \} \leq \epsilon \) or \( t \geq t_{\text{max}} \).
8. \( t = t + 1 \).

Output: Denoised image \( \mathcal{L} = L^{(t)} \); \( \mathcal{L} \).

E. C-Minimization

The subproblem associated with optimization of \( C \) is

\[
\min_{C} \gamma \|C\|_1 + \frac{\rho}{2} \left\| C - DB + \frac{1}{\rho} ZC \right\|_F^2. \tag{33}
\]

The above problem can be solved with the soft-shrinkage operator in an elementwise manner, which leads to

\[
C_{i,j} = \max \left( (DB - ZC/\rho)_{i,j} - \gamma/\rho, 0 \right). \tag{34}
\]

F. Updating of \( Z^{(t)}, Z^{(t)}, Z^{(t)}, Z^{(t)}, \) and \( \rho \)

We update the following variables in a standard way:

\[
\begin{align*}
Z^{(t)} &= \mathcal{L}^{(t)} + \rho(O_{i,j} - \mathcal{L}_{i,j} - S_{i,j}) \\
Z^{(t)} &= \mathcal{L}^{(t)} + \rho(A_{i,j} - \mathcal{L}_{i,j} - S_{i,j}) \\
Z^{(t)} &= DB + \rho(A - B) \\
Z^{(t)} &= DB + \rho(C - DB) \\
\rho &= \rho \kappa \\
\end{align*}
\tag{35}
\]

where \( \kappa > 1 \) is a parameter that keeps \( \rho \) increasing along with the optimization iterations. We summarize the above optimization strategy in Algorithm 1.

To analyze the complexity of the proposed algorithm, we first consider the complexity of updating each patch at each iteration as follows. To update \( \mathcal{L}_{i,j} \) with (18), the major complexity comes from the computation of SVD, which has a complexity of \( O(\min(mn^2, m^2n)) \). To update \( S_{i,j} \) with (26), the complexity is \( O(mn) \). To update \( A_{i,j} \) and \( B_{i,j} \) with (29), (32) and (34), the complexity for each variable is \( O(mn \log(mn)) \), and \( O(mn \log(mn)) \), respectively. To update \( Z^{(t)}, Z^{(t)}, Z^{(t)}, Z^{(t)}, Z^{(t)}, Z^{(t)}, Z^{(t)} \), and \( Z^{(t)} \) by (35), the complexity is \( O(mn) \). In summary, the overall complexity for each patch at each iteration is \( O(\min(mn^2, m^2n) + mn \log(mn)) \). Thus, the overall complexity of the proposed algorithm is \( O(MN \min(n_{\text{max}}(\min(mn^2, m^2n) + mn \log(mn)))) = O(MN \min(mn) + \log(mn))) \text{max}) \). Often times, for the path size, we have \( m = m \), and thus, the overall complexity can be reduced to \( O(MN n_{\text{max}}) \).

V. Perturbation Analysis

In the objective function given in (10), the log-determinant and \( \ell_{2,\log} \) norm are nonconvex. In general, the algorithm of \( L^3S^2TV \) can only find a local minimum. Inspired by [30], we can formulate the estimation error between the unknown global minimum and the computationally obtained local minimum as estimation noise. This formulation will need Lemma 1.

Lemma 1: The \( \ell_{2,\log} \)-norm satisfies the triangle inequality.

Proof: Let \( A \) and \( B \) be two matrices of the same size. Denote the \( j \)th column of matrix \( X \) by \( X_j \). Then, we have

\[
\|A\|_{2,\log} + \|B\|_{2,\log} = \sum_j \log(1 + \|A_j\|_2) + \log(1 + \|B_j\|_2)
\]

\[
= \sum_j \log(1 + \|A_j\|_2 + \|B_j\|_2 + \|A_j\|_2 \|B_j\|_2)
\]

\[
\geq \sum_j \log(\|A_j + B_j\|_2) = \|A + B\|_{2,\log}.
\]

By using this property of the \( \ell_{2,\log} \)-norm, we can bound the estimation noise at a certain level as follows. Without loss of generality, we consider an arbitrary patch of (10), and then, the conclusion can be generalized to the overall objective.

Theorem 2: Given a nontrivial observation data matrix \( O \), let \( L^* \) and \( S^* = O - L^* \) be a pair of global solution minimizing the objective in \( L \) given in (10). For any sufficiently small \( \epsilon > 0 \), there exists an \( \delta > 0 \) such that for any nonnegative observation matrix \( \bar{O} = O + E \) with \( \|E\|_{2,\log} \leq \delta \), we have

\[
J_L(L) := \|L^* - L\|_{\log \det} + \lambda \|L^* - L\|_{2,\log} + \mu \|L^* - L\|_{SSTV} < \epsilon
\]

where \( L \) is a local solution of \( \bar{O} \) estimated by the algorithm for solving (10).

Proof: First, define the main objective function of (10) to be \( R(L; O) \). Without loss of generality, we will consider \( \lambda = 1 \) and other \( \lambda \) values can be similarly treated. Now, take a sufficiently small \( \epsilon \) that is \( 0 < \epsilon < \|L^* - O\|_{2,\log} \), and let \( \delta_1 = R(L^*; O) \). Note that \( \delta_1 > 0 \), because \( \delta_1 = 0 \) is equivalent to \( O = 0 \), violating the condition on \( O \). Necessarily, we have \( \epsilon < \delta_1 \).

Let \( \mathcal{G} \) be the open set of all \( L \) close to \( L^* \)

\[
\mathcal{G} := \{ L \mid |J_L(L) - L^*| < \epsilon \}
\]

Also, define a set \( B := \{ L \mid \|L\|_{2,\log} \leq \epsilon \} \), where \( \epsilon \) is a positive constant. Let \( \mathcal{G} := \{ L \mid J_L(L) \geq \epsilon \} \cap B \). Because \( \mathcal{G} \) is bounded and closed and the norms are continuous, we have \( \min_{L \in \mathcal{G}} R(L; O) = (R(L^*; O) - \delta_1) > 0 \).

For \( L \notin \mathcal{G} \) and \( \max(L) > \epsilon \), it is straightforward to verify that \( \|O - \hat{L}\|_{2,\log} = \min \left( \log(1 + \sum (O_{i,j} - \hat{L}_{i,j})^2)^{1/2} \right) > \log(1 + \max(L) - \max(O)) > \epsilon + \delta_1 \). Here, \( \max(A) \) operates on the set of all elements of a matrix \( A \). Therefore, \( R(L; O) - R(L^*; O) > \|O - \hat{L}\|_{2,\log} - R(L^*; O) > \epsilon + \delta_1 - \delta_1 = \epsilon \). Now, take \( \delta = \min(\epsilon, \delta_1)/3 \), and then, we have \( R(L; O) - R(L^*; O) > 3 \delta > 0 \), for any \( L \notin \mathcal{G} \).

For any \( \hat{O} \) such that \( \|O - \hat{O}\|_{2,\log} < \delta \), by the triangle inequality of \( \ell_{2,\log} \) norm, we have \( R(L; \hat{O}) + \|O - \hat{O}\|_{2,\log} \geq \)}
\[ R(\hat{L}; O) \geq R(L^*; O) - \|O - \hat{O}\|_{2,\log}^2 \]

where

\[ R(L; O) = R(L^*; O) - \|O - \hat{O}\|_{2,\log}^2 - R(L^*; \hat{O}) \]

\[ \geq R(L; O) - \|O - \hat{O}\|_{2,\log} - R(L^*; \hat{O}) \]

\[ \geq R(L; O) - \|O - \hat{O}\|_{2,\log} - 2\|O - \hat{O}\|_{2,\log} \]

\[ > 3\delta - 2\delta = \delta. \]

All solutions that are not in \( \mathcal{G} \) will not be a minimizer of the objective function. This concludes the proof.

In this section, we conduct extensive experiments to testify the effectiveness of the proposed method. In particular, we compare our method with seven state-of-the-art HSI denoising methods, including the block-matching and 4-D filtering (BM4D) algorithm [33], the convex approach to RPCA [3], GoDec-based LRMR [58], SSTV [5], noise-adjusted iterative low-rank matrix approximation (NAILRMA) [22], factorization-based nonconvex RPCA (U-FFP) [35], total variation regularized low-rank matrix factorization (LRTV) [23], and local low-rank matrix recovery with global spatial–spectral total variation (LLRTV) [20].

VI. EXPERIMENTS

In this section, we conduct extensive experiments to testify the effectiveness of the proposed method. In particular, we compare our method with seven state-of-the-art HSI denoising methods, including the block-matching and 4-D filtering (BM4D) algorithm [33], the convex approach to RPCA [3], GoDec-based LRMR [58], SSTV [5], noise-adjusted iterative low-rank matrix approximation (NAILRMA) [22], factorization-based nonconvex RPCA (U-FFP) [35], total variation regularized low-rank matrix factorization (LRTV) [23], and local low-rank matrix recovery with global spatial–spectral total variation (LLRTV) [20]. We follow a strategy similar to the literature [20], [22] and scale the gray value of each band of HSIs to the interval \([0, 1]\) before denoising. After denoising, we restore the recovered images to the original gray value level, which facilitates numerical calculation and visualization. In the experiments, we tune the parameters for each method such that they are manually adjusted to the best according to the default strategy.

A. Simulated HSI Data Experiments

In this test, we conduct experiments on simulated datasets to quantitatively compare all methods in HSI denoising. In particular, we use the ground truth (GT) image of the Indian pines dataset to generate the synthetic data, which has a size of \(145 \times 145 \times 224\). We treat the synthetic dataset as the GT of the simulated HSIs. Then, we use the GT of the simulated HSIs to generate noisy HSIs, where we artificially add noise to the GT under six conditions, resulting in six noisy HSI datasets. We describe how we add noise to the GT to obtain the noisy datasets as follows.

1) Case 1: We add zero-mean Gaussian noise to each band of the GT, where all bands are fixed to have the same noise intensity. In this case, we set the noise variance to be 0.1.

2) Case 2: Based on Case 1, we further add some deadlines to bands 81–120. In each of these bands, we randomly add 3–10 deadlines, where each deadline has a random width of 1-3 columns.

3) Case 3: Similar to Case 1, we add the Gaussian noise to the GT, except that the noise variance is set to 0.14 in this case. Then, we add some stripes in bands 161–190. In each of these bands, we randomly select 20–40 columns to add stripes. For each column, we randomly select a value within \([-0.25, 0.25]\) and add it to all pixels of this column.

4) Case 4: Based on Case 2, we further add some stripes in bands 161–190 in the same way as in Case 3.

5) Case 5: First, we randomly add zero-mean Gaussian noise with an SNR value between 15 and 25 dB to each band, where the average SNR value of the noise in all bands is 20.43 dB. Then, we add salt and pepper impulse noise to the dataset by randomly corrupting 20% pixels in each band. Finally, we randomly add impulse noise with intensity between 0.0196 and 0.0784 to each band separately, where the average intensity of the impulse noise in all bands is 0.0492.

6) Case 6: First, we randomly add zero-mean Gaussian noise with an SNR value between 45 and 55 dB to each band, where the average SNR value of the noise in all bands is 49.75 dB. Then, we add impulse noise to the dataset in the same way as in Case 5.

7) Case 7: In this case, all types of noises in cases 1–6 are added to the GT. First, we add Gaussian and impulse noise to the GT in the same way as Case 6. Then, we further add some deadlines to bands 71–80 in the same way as in Case 2. Finally, we randomly select 10–20 columns in each of bands 161–175 to add stripes. For each column of these stripes, we randomly select a value within \([-0.25, 0.25]\) and add it to all pixels.

In the rest of this section, we will conduct extensive experiments to test the proposed method both quantitatively and qualitatively. The detailed comparison results and discussions are presented in the following.

To quantitatively evaluate the proposed method, we adopt three widely used evaluation metrics to compare all methods, including the peak signal-to-noise ratio (PSNR) [27], structural similarity (SSIM) [49], relative global scale (ERGAS) [45], information fidelity criterion (IFC) [43], and feature similarity (FSIM) [62]. Better performance is obtained with lower values for ERGAS and higher values for the other four metrics. Since metrics, including PSNR, SSIM, and FSIM, are applied to every single band, we report the averaged values over all bands, which are denoted as mean PSNR (MPSNR), mean SSIM (MSSIM), and mean FSIM (MFSIM), respectively. We apply all methods to the above generated noisy datasets and report the detailed denoising performance in Table I. It is seen that the proposed method and LRTDTV are among the top-tier methods, where they obtain almost all the top two denoising performances. In particular, the proposed method achieves the top one and two performances in 22 and 32 out of 35 cases, respectively, which shows its superiority to other baseline methods. In other cases, the proposed method is
Table I: Quantitative Evaluation of Different Methods in Different Noise Cases of Indian Pines Dataset

| Noise | Metric | BM4D | LMR | NAILRMA | U-FFP | RPCA | LRTV | LRTDTV | SSTV | LLRGTV | Ours |
|-------|--------|------|------|---------|-------|------|------|--------|------|--------|------|
| MPSNR | 37.945 | 37.032 | 37.234 | 37.522 | 32.279 | 39.316 | 40.569 | 32.857 | 40.697 | 42.809 | |
| Case 1 | ERGAS | 30.490 | 33.902 | 32.872 | 39.174 | 58.783 | 26.722 | 23.958 | 57.775 | 22.860 | 17.417 |
| | IFP | 3.024 | 3.263 | 3.272 | 3.046 | 2.977 | 3.923 | 3.046 | 2.977 | 3.923 | 3.119 |
| | FSIM | 0.961 | 0.945 | 0.943 | 0.919 | 0.892 | 0.975 | 0.897 | 0.855 | 0.976 | 0.988 |
| | TIME(s) | 419.690 | 285.903 | 166.732 | 29.905 | 377.144 | 404.123 | 520.790 | 178.315 | 56.515 | 47.145 |
| MSE | 33.547 | 35.203 | 35.008 | 34.008 | 32.018 | 33.276 | 32.003 | 34.154 | 34.303 | |
| Case 2 | MSE | 34.608 | 33.825 | 34.101 | 32.599 | 29.710 | 37.491 | 38.037 | 30.431 | 37.079 | 37.184 |
| | Case 3 | MSE | 36.972 | 0.906 | 0.902 | 0.864 | 0.835 | 0.792 | 0.797 | 0.784 | 0.939 | 0.973 |
| | TIME(s) | 422.276 | 209.775 | 318.858 | 69.162 | 359.402 | 5084.611 | 255.418 | 107.334 | 52.979 | 62.724 |
| Case 4 | MSE | 37.967 | 35.368 | 37.622 | 37.315 | 35.952 | 41.421 | 41.871 | 33.471 | 39.241 | 41.629 |
| | Case 5 | MSE | 37.617 | 33.084 | 31.366 | 30.073 | 27.376 | 34.306 | 35.514 | 27.998 | 35.322 | 35.750 |
| | Case 6 | MSE | 36.972 | 0.906 | 0.902 | 0.864 | 0.835 | 0.792 | 0.797 | 0.784 | 0.939 | 0.973 |
| | TIME(s) | 443.260 | 716.812 | 308.602 | 108.123 | 242.520 | 5904.672 | 76.533 | 140.405 | 61.833 | 79.114 |

The top three performances are highlighted in red, blue, and green, respectively.

Also quite competitive and obtains the top third performances. Compared with LRTDTV, the proposed method has significant improvements in MPSNR, ERGAS, and IFP. For example, the proposed method improves the performance by about two and six in MPSNR and ERGAS in Case 1, respectively. Compared with the other baseline methods, the improvements of the proposed method are more significant. For example, BM4D, LRM, and NAILRMA are among the second-tier methods. Compared with them, the proposed method has improvements by about 5 in MPSNR and 1 in IFP in almost all cases. In ERGAS, the proposed method improves the performance by at least ten in almost all cases, and the improvement can be even about 80 on Cases 2 and 4. In MSSIM and MSFIM, the proposed method improves the performance by about 0.05 in almost all cases. Regarding time cost, the proposed method is quite competitive, where it is among the fastest two methods in most cases. Compared with LRTDTV, which obtains the second best denoising performance, it is observed that the proposed method is generally several times faster, which confirms the effectiveness as well as efficiency of the proposed method. These observations suggest the effectiveness and superior performance of the proposed method to the baseline methods from a quantitative perspective.

To evaluate the proposed method from a qualitative perspective, we further show some visual results for detailed comparison. For all methods in comparison, without loss of generality, we show the denoised 206th band of Case 3 in Fig. 2. It is seen that among all methods, the NAILRMA, LRM, RPCA, U-FFP, SSTV, and LLRGTV fail to remove the mixed types of noise. BM4D has relatively better performance with the majority noise removed; however, it has over-smoothing effects in the recovered image, where the edges between smooth regions are blurred and some adjacent smooth regions are merged. LRTV and LRTDTV are among the top methods, where it is seen that they eliminate the majority noise and well capture the detailed information of regions. However, the edge information is damaged in these images and shows inferior
performance compared with the proposed method. Among all methods, the proposed method obtains the best denoising performance, where it is seen that the restored image contains smooth regions, clean edge structures, and eliminates noise. These observations confirm the effectiveness of the proposed method from a visual perspective.

In Table I, we have compared all methods with the averaged performance over all bands in MPSNR, MSSIM, and ERGAS. To better compare the methods in our experiment, in this test, we further compare their performances on each individual band. Without loss of generality, we show the results of Case 3 in terms of PSNR and SSIM in Fig. 3. It is noted that the proposed method achieves the top performance in PSNR with significant improvements in almost all bands. In SSIM, the proposed method is quite competitive among all methods and has at least the top second performance in almost all bands. This suggests that the proposed method not only achieves promising performance on the overall dataset but also in each band, which is essentially important in real-world applications.

To further assess the effectiveness of the proposed method, in this test, we show some results of the spectral signatures in the restored images. In particular, we show the spectral signature curves of the restored images in Case 3 at the location of pixel (140, 90) across all bands. To better investigate the quality of the restored images by each method, we show the spectral signature curves of both the restored images and the GT in Fig. 4. For the recovered images with high quality, their spectral signature curves are expected to be very close to those of the GT. Among the baseline methods, the LLRGTV, LRTV, and LRTDTV have the best performances, where we observe that their curves are the closest ones to the original curve. This confirms the relatively superior performance of these methods among all baseline methods. Compared with these methods, the proposed method has even better performance. Specifically, the curve obtained by LLRGTV is more zigzag than that by our method, especially in bands 20–100, implying that the proposed method better retains spectral consistency between nearby bands in the recovered HSI. The curve obtained by LRTDTV does not have zigzag effects, but it fits the original curve less closely compared with the curve obtained by our method. LRTV has relatively better performance than LRTDTV and LLRGTV, but inferior performance to the proposed method, especially in the first 100 bands. For the other methods, the curves of the recovered HSIs have significant differences from the original HSI, which shows obvious inferior performance to the proposed method. These observations confirm the effectiveness and superior performance of the proposed method.

B. Experiments on Real-World HSI Datasets

In this test, we conduct experiments on three real-world HSI datasets, including the Hyperspectral Digital Imagery Collection Experiment (HYDICE) urban dataset, the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) Indian pines dataset, the EO-1 Hyperion Australia dataset, and Reflective Optics System Imaging Spectrometer (ROSIS) University of Pavia dataset. For a visual illustration, we show some
respectively. Compared with other methods, BM4D and RPCA
and 139th bands obtained by each method in Figs. 6 and 7,
obtained by each method.

tune their parameters such that the best visual performance is
optimized. For RPCA, we tune its parameter around the theoret-
ical optimal value to obtain the best visual performance. For
all methods are evaluated from a visual quality perspective,
straightforward to evaluate the methods quantitatively. Thus,
of the dataset.

This dataset is heavily corrupted with stripes, deadlines,
× 210 bands of images with a size of 307 × 307 pixels.

examples of the false-color HSI in Fig. 5. In the rest of this
section, we will present brief descriptions of the datasets and
the detailed experimental results.

1) HYDICE Urban Dataset: The HYDICE dataset contains
210 bands of images with a size of 307 × 307 pixels. This
data set is heavily corrupted with stripes, deadlines,
atmosphere, water absorption, and other unknown types of
noises. In our experiment, we test the methods using all bands
of the dataset.

Since the real-world datasets lack clean bands, it is not
straightforward to evaluate the methods quantitatively. Thus,
all methods are evaluated from a visual quality perspective,
which suggests us to tune the parameters for the methods as
follows. For RPCA, we tune its parameter around the theoret-
ically optimal value to obtain the best visual performance. For
the other methods, we follow a common strategy and manually
tune their parameters such that the best visual performance is
observed for each method.

Without loss of generality, we show the results of the 108th
and 139th bands obtained by each method in Figs. 6 and 7,
respectively. Compared with other methods, BM4D and RPCA
have inferior performance on this dataset, where it is observed
that both the 108th and 139th bands restored by BM4D and
RPCA have strong fringe noise effects. The restored images
by the NAILRMA, LRMR, U-FFP, and LRTDTV also have
fringe effects in both bands, but they are much lighter than
those restored by BM4D and RPCA. SSTV removes almost
all noise in band 108, but we can still observe the fringe
effects and some detail information is missing in band 139.
Among all methods, LRTGTV, LRTV, and our method are
the only ones that well remove the fringes from the HSIs.
It is seen that LRTGTV and LRTV generate clearly visible
images and appear effective in removing noise. Unfortunately,
both LRTGTV and LRTV fail to retain rich detail information
from the noisy HSIs as our method does. For example, in the
amplified regions, it is seen that the region is oversmoothed
and structural detail information is missing. Moreover, the
images generated by LRTV appear darker than the other
images as well as the original one. In summary, on the
HYDICE dataset, the proposed method removes the noise
and well retains local details, while the baseline methods
fail, suggesting superior performance and effectiveness of the
proposed method in HSI denoising.

2) AVIRIS Indian Pines Dataset: The AVIRIS Indian pines
data set consists of 220 bands of images with a size of
145 × 145 pixels. In this dataset, some bands are severely
damaged by mixed Gaussian and impulse noise, which makes
it challenging to remove noise from this dataset. Among all
the bands, we select two typical ones, including bands 11 and
150, to show the denoising performance of all methods. The
restored images of these two bands are shown in Figs. 8 and 9.
All parameters are tuned in a way that follows Section VI-B1
on this dataset.

In Fig. 8, it is seen that the original band contains light
noise. In the amplified area, it is seen that the proposed method
removes the noise and the edge is clearly observed. In contrast,
in the bands restored by U-FFP, SSTV, LLRGTV, LRTV, and LRTDTV, the edges are difficult to observe. In the bands restored by BM4D and LRMR, although the edge information is not missing, we can still observe noise effects in the dark zone. Among all the baseline methods, the NAILRMA and RPCA have relatively the best performances. However, we can still observe light noise effects in the dark zone and the edges are not as clear as in that restored by our method. In Fig. 9, it is seen that the original band contains heavy noise. It is observed BM4D, RPCA, and SSTV fail to remove noise from band 150 where we can still observe heavy noise effects in the restored bands. Among all the other methods, as is observed in the amplified area, the proposed method is the best in retaining the rich edge information. Although methods, such as LRTV and LRTDTV have better performance in retaining edge information in band 150 than band 11, the edges in the restored bands are blurred and the smoothness of smooth zones can be further improved. These observations show that the proposed method has the best performance and is the only one that obtains satisfactory results in both bands 11 and 150, which suggests its effectiveness on this dataset.

3) EO-1 Hyperion Australia Dataset: The Hyperion Australia image was captured on December 4, 2010, with the original size of $3858 \times 256 \times 242$. We follow the strategy in the literature and remove the overlapping bands between visual near-infrared and shortwave infrared ranges. As a result, we use a subregion of size $200 \times 200 \times 150$ in our experiment. Among all the bands, we select two typical ones, including bands 47 and 89, to show the denoising performance of all methods. The restored images of these two bands are shown in Figs. 10 and 11, respectively. All parameters are tuned in a way that follows Section VI-B1 on this dataset. In Fig. 10, it is seen that images obtained by BM4D and SSTV still have some noise effects, while those obtained by NAILRMA, LRMR, RPCA, and LLRGTV have strong block or over smoothing effects. Among all baseline methods, it is seen that LRTV, U-FFP, and LRTDTV have relatively better performance. In the amplified region, it is seen that LRTV loses the detail information, whereas the proposed method recovers the detail information better than U-FFP and LRTDTV. For example, the proposed method recovers the detail information with clearer edges than U-FFP and LRTDTV. In Fig. 11, it is seen that BM4D, SSTV, and LRTDTV still have noise effects in the recovered images, while LRMR and RPCA have block effects. In the images recovered by NAILRMA, LLRGTV, and LRTV, the detail information is less clean than that recovered by the proposed method. The U-FFP has relatively the best performance among all baseline methods, but it has
Fig. 11. Restoration results on Hyperion dataset. (a) Original 89th band of Hyperion. (b)-(k) Restored results of the 89th band by different methods. The figure is better viewed in a zoomed-in PDF. (a) Original. (b) BM4D. (c) NAILRMA. (d) LRMR. (e) RPCA. (f) U-FFP. (g) SSTV. (h) LLRGTV. (i) LRTV. (j) LRTDTV. (k) Ours.

Fig. 12. Restoration results on Pavia University dataset. (a) Original 2nd band of Hyperion. (b)-(k) Restored results of the 2nd band by different methods. The figure is better viewed in a zoomed-in PDF. (a) Original. (b) BM4D. (c) NAILRMA. (d) LRMR. (e) RPCA. (f) U-FFP. (g) SSTV. (h) LLRGTV. (i) LRTV. (j) LRTDTV. (k) Ours.

Fig. 13. Restoration results on Pavia University dataset. (a) Original 4th band of Hyperion. (b)-(k) Restored results of the 4th band by different methods. The figure is better viewed in a zoomed-in PDF. (a) Original. (b) BM4D. (c) NAILRMA. (d) LRMR. (e) RPCA. (f) U-FFP. (g) SSTV. (h) LLRGTV. (i) LRTV. (j) LRTDTV. (k) Ours.

Fig. 14. Spectrum of pixel (40, 90) in the restoration results of HYDICE dataset, respectively.

relatively more noise than the proposed method in the smooth regions of the recovered image. These observations confirm the effectiveness of the proposed method on this dataset.

4) ROSIS University of Pavia Dataset: This dataset contains 103 spectral bands of 610 × 340 spatial pixels. We follow the literature and use a subset of the data with a size of 300 × 300 × 103 in the experiment. Among all the bands, we report the visual results on bands 2 and 4 in Figs. 12 and 13 to compare the denoising performance of all methods. All parameters are tuned in a way that follows Section VI-B1 on this dataset.

In Fig. 12, it is seen that the images obtained by NAILRMA, RPCA, LLRGTV, LRTV, and LRTDTV have significantly brighter effects. This implies that the corresponding methods not only remove noise but also overly change the intensity values of the underlying clean bands, which brings significant errors in the restored images. For this observation, we have more discussions in Section VI-B5. Despite the overly changed intensity values, we can still observe the noise effects in some bands, such as the one recovered by the LRTDTV. Meanwhile, the other methods do not have such an obvious issue in overly changing the intensity value of original HSIs. In the bands recovered by BM4D, LRMR, and U-FFP, we can still observe some noise effects. The bands recovered by SSTV and our method have relatively better performance among all methods. Compared with the SSTV, the proposed method generates the recovered band with clearer edge and structural information. Similar observations can be found in Fig. 12. These observations suggest the effectiveness of the proposed method in HSI denoising.

5) Spectral Signatures and Mean DN Profiles: To further assess the effectiveness of the proposed method on real-world datasets, we show the comparison results of the spectral signatures of all methods. Without loss of generality, we show the results on the HYDICE dataset. In particular, we show the curves of the original and the restored images at the location of pixel (40, 90) across all bands in Fig. 14. It is
observed that there are several sudden changes, i.e., the peaks and valleys, on the curve of original data, which corresponds to the stripes and deadlines or other types of outliers. It is expected that such outliers are significantly removed on the spectral signature curves of the restored images. It is seen that the proposed method has the best performance in removing such outliers among all methods. Moreover, although methods, such as LRTDTV, LRTV, and LLRGTV, have relatively better denoising performance among the baseline methods as shown in Figs. 6 and 7, the spectral signature curves of these methods significantly deviate from the curve of original data, which implies that these methods not only remove noise from the bands but also overly change the intensity values of the underlying clean bands and thus brings significant errors in the restored images. These observations suggest that the proposed method has a better performance in removing noise and recovering the true intensity values of the HSIs.

For a more detailed comparison, we show the horizontal and vertical mean profiles of the restored images of band 152 in the HYDICE Urban dataset in Figs. 15 and 16, where horizontal and vertical axes represent the row and column numbers and the corresponding mean DN values, respectively. It is seen that the LRTV significantly changes the intensity value of the HSI. Among the other methods, BM4D, LRMR, RPCA, U-FFP, SSTV, and LRTDTV do not remove the stripes from the horizontal direction. Among all the baseline methods, the NAILRMA and LLRGTV have the most competitive performance. However, as can be seen in the highlighted parts in Figs. 15 and 16, they have over smooth effects from the horizontal direction and change the intensity values slightly more than the proposed method. In summary, the experimental results on these real-world datasets confirm the effectiveness of the proposed method and its superiority to baseline methods.

C. Parameter Sensitivity

In the above tests, we have confirmed the effectiveness of the proposed method in HSI denoising. In this test, we further show how the parameters affect the denoising performance of the proposed method. For a quantitative illustration and without loss of generality, we use the synthetic data in Case 3. In particular, we report the performance of the proposed method in three metrics with respect to different combinations of \( \lambda \) and \( \gamma \) in Fig. 17. It is seen that the proposed method achieves relatively good performance when \( \lambda \) is nearby 1 and \( \gamma \) is small. For such \( \lambda \) and \( \gamma \) values, it is seen that the proposed method achieves a good performance with a wide range of combinations of parameters. Similar observations can be found in other cases. Such behavior is indeed important for unsupervised learning method to be potentially applied in real-world applications. These observations suggest that we adopt such values in real-world applications.

D. Convergence Study

For the comprehensive optimization strategy, it is generally not straightforward to provide theoretical results on the convergence. Thus, in this test, we empirically testify the convergence of the proposed method. Without loss of generality, we first conduct experiments using the synthetic data in Case 3, where the parameters are set to be \( \lambda = 1.3 \) and \( \gamma = 0.0022 \). For other values, generally, we can observe similar patterns.

First, we show the convergence behavior of the proposed method in variable sequence by showing three different errors that measure how the constraints are met. Without loss of generality, we show how the updates of variables gradually satisfy the constraints by plotting the sequences of errors, including \( \|O^{(t)} - L^{(t)} - S^{(t)}\|_F^{\infty} \), \( \|L^{(t)} - A^{(t)}\|_F^{\infty} \), and \( \|A^{(t)} - B^{(t)}\|_F^{\infty} \) in Fig. 18. It is seen that the sequences of errors converge within a few numbers of iterations, indicating that the variable sequences gradually meet the constraints and converge. Moreover, to better show the convergence behavior, we also plot the sequences

\[
\begin{align*}
\|O^{(t+1)} - L^{(t+1)} - S^{(t+1)}\|_F^{\infty} \\
\|O^{(t)} - L^{(t)} - S^{(t)}\|_F^{\infty} \\
\|L^{(t+1)} - A^{(t+1)}\|_F^{\infty} \\
\|L^{(t)} - A^{(t)}\|_F^{\infty}
\end{align*}
\]

and

\[
\begin{align*}
\|A^{(t+1)} - B^{(t+1)}\|_F^{\infty} \\
\|A^{(t)} - B^{(t)}\|_F^{\infty}
\end{align*}
\]

in Fig. 18, which show the relative changes between consecutive elements of sequences of errors. It is seen that the relative changes are smaller than 1 after the first few iterations. To further show such behaviors on real-world datasets, we show some curves on the Indian Pines dataset in Fig. 19, where
we set $\lambda = 1$ and $\gamma = 0.0001$. It is seen that a similar pattern to the synthetic dataset is observed on the real-world dataset, which indicates that the sequences of errors indeed converge with at least a linear convergence rate, which is satisfactory and suggested efficiency in real-world applications.

Then, we show the denoising performance of the intermediate variable generated by the proposed method with respect to the iteration numbers in a way similar to [10]. In particular, we plot the MPSNR, MSSIM, and ERGAS values with respect to the iteration numbers in Fig. 20. It is seen that when the iteration number increases, the recovered HSI has gradually improved quality in all metrics, which suggests the effectiveness of using the nonconvex RPCA part in our model. For this purpose, we further conduct experiments to show the significance of exploiting spatial–spectral information for HSI denoising. In particular, we show the results in Fig. 21. It is seen that when $\gamma > 0$ or $\gamma = 0$, our model falls back to (9), where the SSTV term is not adopted. For the two approaches, we report their highest performance by tuning balancing parameters. It is seen that the proposed model with integrated SSTV term has significantly improved denoising performance than the pure nonconvex RPCA model of (9). The only difference between the two models lies in the usage of the SSTV norm, and thus, it is natural to believe that such a difference leads to the difference in denoising performance. These observations confirm the significance and effectiveness of exploiting spatial–spectral information by integrating an SSTV term in our model.

Then, we conduct experiments to show the significance of the nonconvex RPCA part in our model. For this purpose, we compare our model with “RPCA+SSTV,” which refers to the model of convex RPCA with an integrated SSTV term. We show the comparison results of the two approaches in Fig. 22. It is seen that our method has significantly improved performance compared with the RPCA+SSTV approach, which suggests the effectiveness of using the nonconvex approach in HSI denoising applications.

Finally, since $\ell_{2,\text{log}}$ is the key contribution of our work, we further conduct experiments to show the significance of the proposed $\ell_{2,\text{log}}$ term in HSI denoising. In particular, we compare our model with two models that replace the $\ell_{2,\text{log}}$ norm with the $\ell_1$ and $\ell_1$ norms, which are denoted as “logdet+$\ell_1$+SSTV” and “logdet+$\ell_{2,1}$+SSTV,” respectively. We show the comparison results of the three models in Fig. 23. It is seen that our method has significantly improved performance compared with the $\ell_1$ and $\ell_{2,1}$ norm-based approaches in all metrics, which suggests the effectiveness of using the proposed $\ell_{2,\text{log}}$ term. Thus, it is convincing to claim the effectiveness of the proposed method in HSI denoising.
In this article, we propose a novel method for HSI denoising, named $L^1$S-TV. Unlike existing low-rank models that only focus on developing more accurate rank approximation for low-rank component recovery, we propose to simultaneously adopt nonconvex approximations to both the rank and the columnwise sparsity for more accurate separation of the low-rank and sparse components. In particular, we propose a log-based columnwisely sparse approximation, named the $\ell_2^\log$ norm, which is more accurate than the widely used convex approach, i.e., $\ell_2^1$ norm. For its associated shrinkage problem, we developed an efficient optimization strategy that is formally presented in a theorem. The $\ell_2^\log$ norm can be generally used in various problems that restrict columnwise sparsity. Moreover, we impose the SSTV regularization in the log-based nonconvex RPCA model, which enhances the global piecewise smoothness and spectral consistency from the spatial and spectral views in the recovered HSI. Extensive experiments on both simulated and real HSIs demonstrate the effectiveness of the proposed method in denoising HSIs.

REFERENCES

[1] J. M. Bioucas-Dias and A. Plaza, “An overview on hyperspectral unmixing: Geometrical, statistical, and sparse regression based approaches,” in Proc. IEEE Int. Geosci. Remote Sens. Symp., Jul. 2011, pp. 1135–1138.

[2] S. Cai, Z. Kang, M. Yang, X. Xiong, C. Peng, and M. Xiao, “Image denoising via improved dictionary learning with global structure and local similarity preservations,” Symmetry, vol. 10, no. 5, p. 167, May 2018.

[3] E. J. Candès, X. Li, Y. Ma, and J. A. Wright, “Robust principal component analysis?” J. ACM, vol. 58, no. 3, pp. 1–37, 2011.

[4] C. Cao, J. Yu, C. Zhou, K. Hu, F. Xiao, and X. Gao, “Hyperspectral image denoising via subspace-based nonlocal low-rank and sparse factorization,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 12, no. 3, pp. 973–988, Mar. 2019.

[5] S. H. Chan, R. Keshabeh, K. B. Gibson, P. E. Gill, and T. Q. Nguyen, “An augmented Lagrangian method for total variation video restoration,” IEEE Trans. Image Process., vol. 20, no. 11, pp. 3097–3111, Nov. 2011.

[6] C.-I. Chang and Q. Du, “Interference and noise-adjusted principal components analysis,” IEEE Trans. Geosci. Remote Sens., vol. 37, no. 5, pp. 2387–2396, Sep. 1999.

[7] Y. Chang, L. Yan, H. Fang, and C. Luo, “Anisotropic spectral–spatial total variation model for multispectral remote sensing image destriping,” IEEE Trans. Image Process., vol. 24, no. 6, pp. 1852–1866, Jun. 2015.

[8] G. Chen and S.-E. Qian, “Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage,” IEEE Trans. Geosci. Remote Sens., vol. 49, no. 3, pp. 973–980, Mar. 2011.

[9] Y. Chen, W. He, N. Yokoya, T.-Z. Huang, and X.-L. Zhao, “Nonlocal tensor-ring decomposition for hyperspectral image denoising,” IEEE Trans. Geosci. Remote Sens., vol. 58, no. 2, pp. 1348–1362, Oct. 2019.

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

[11] Y. Chen and T. Pock, “Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1256–1272, Jun. 2016.

[12] K. Dubov, 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.
[35] C. Peng, Y. Chen, Z. Kang, C. Chen, and Q. Cheng, “Robust principal component analysis: A factorization-based approach with linear complexity,” Inf. Sci., vol. 513, pp. 583–599, Mar. 2020.

[36] C. Peng, Z. Kang, H. Li, and Q. Cheng, “Subspace clustering using log-determinant rank approximation,” in Proc. 21st ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, New York, NY, USA, Aug. 2015, pp. 925–934.

[37] C. Peng, Q. Zhang, Z. Kang, C. Chen, and Q. Cheng, “Kernel two-dimensional ridge regression for subspace clustering,” Pattern Recognit., vol. 113, May 2021, Art. no. 107749.

[38] C. Peng, Y. Zhang, Y. Chen, Z. Kang, C. Chen, and Q. Cheng, “Log-based sparse nonnegative matrix factorization for data representation,” Knowl.-Based Syst., vol. 251, Sep. 2022, Art. no. 109127.

[39] C. Peng, Z. Zhang, Z. Kang, C. Chen, and Q. Cheng, “Nonnegative matrix factorization with local similarity learning,” Inf. Sci., vol. 562, pp. 325–346, Jul. 2021.

[40] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, “Image denoising using scale mixtures of Gaussians in the wavelet domain,” IEEE Trans. Image Process., vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[41] 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.

[42] B. Rasti, J. R. Sveinsson, and M. O. Ulfarsson, “Wavelet-based sparse reduced-rank regression for hyperspectral image restoration,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 10, pp. 6688–6698, Oct. 2014.

[43] H. R. Sheikh, A. C. Bovik, and G. d. Veciana, “An information fidelity criterion for image quality assessment using natural scene statistics,” IEEE Trans. Image Process., vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[44] K. C. Tiwari, M. K. Arora, and D. Singh, “An assessment of independent component analysis for detection of military targets from hyperspectral images,” Int. J. Appl. Earth Observ. Geospat., vol. 13, no. 5, pp. 730–740, Oct. 2011.

[45] L. Wald, Data Fusion: Definitions and Architectures: Fusion of Images of Different Spatial Resolutions. Paris, France: Les Presses de l’École des Mines, 2002.

[46] M. Wang, Q. Wang, and J. Chanussot, “Tensor low-rank constraint and $\ell_0$ total variation for hyperspectral image mixed noise removal,” IEEE J. Sel. Topics Signal Process., vol. 15, no. 3, pp. 718–733, Apr. 2021.

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

[48] Y. Wang, R. Niu, and X. Yu, “Anisotropic diffusion for hyperspectral imagery enhancement,” IEEE Sensors J., vol. 10, no. 3, pp. 469–477, Mar. 2010.

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

[50] W. Wei, L. Zhang, Y. Jiao, C. Tian, C. Wang, and Y. Zhang, “Intracluster structured low-rank matrix analysis method for hyperspectral denoising,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 2, pp. 866–880, Feb. 2019.

[51] T. Xie, S. Li, and B. Sun, “Hyperspectral images denoising via non-convex regularized low-rank and sparse matrix decomposition,” IEEE Trans. Image Process., vol. 29, pp. 44–56, 2020.

[52] W. Xie and Y. Li, “Hyperspectral imagery denoising by deep learning with trainable nonlinearity function,” IEEE Geosci. Remote Sens. Lett., vol. 14, no. 11, pp. 1963–1967, Nov. 2017.

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

[54] F. Xu, X. Bai, and J. Zhou, “Non-local similarity based tensor decomposition for hyperspectral image denoising,” in Proc. IEEE Int. Conf. Image Process. (ICIP), Sep. 2017, pp. 1890–1894.

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

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

[57] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, “Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 2, pp. 1205–1218, Sep. 2018.

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

[59] H. Zhang, L. Liu, W. He, and L. Zhang, “Hyperspectral image denoising with total variation regularization and nonlocal low-rank tensor decomposition,” IEEE Trans. Geosci. Remote Sens., vol. 58, no. 5, pp. 3071–3084, May 2019.

[60] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising,” IEEE Trans. Image Process., vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[61] L. P. Zhang et al., “Compression of hyperspectral remote sensing images by tensor approach,” Neurocomputing, vol. 147, pp. 358–363, Sep. 2015.

[62] 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.

[63] Q. Zhang, Q. Yuan, J. Li, X. Liu, H. Shen, and L. Zhang, “Hybrid noise removal in hyperspectral imagery with a spatial–spectral gradient network,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 10, pp. 7317–7329, Oct. 2019.

[64] Q. Zhang, Q. Yuan, J. Li, F. Sun, and L. Zhang, “Deep spatio-spectral Bayesian posterior for hyperspectral image non-i.i.d. noise removal,” ISPRS J. Photogramm. Remote Sens., vol. 164, pp. 125–137, Jun. 2020.

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

[66] Y. Zheng, T. Huang, X. Zhao, Y. Chen, and W. He, “Double-factor-regularized low-rank tensor factorization for mixed noise removal in hyperspectral image,” IEEE Trans. Geosci. Remote Sens., vol. 58, no. 12, pp. 8450–8464, Dec. 2020.

[67] P. Zhong and R. Wang, “Multiple-spectral-band CRFs for denoising junk images,” Int. J. Appl. Earth Observ. Geoinf., vol. 29, pp. 44–56, 2020.

[68] L. Zhang 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.