A Non-local Rank-Constraint Hyperspectral Images Denoising Method with 3-D Anisotropic Total Variation

Tao Gong[1,2], **Desheng Wen**[1] **and Tianbin He**[1]

[1]Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences,
[2]University of Chinese Academy of Sciences

gongtao2017@opt.cn

Abstract. Hyperspectral Images (HSIs) are usually degraded by many kinds of noise called mixed noise, which greatly limits the subsequent applications of HSIs. Many researches have proved the patch-based low-rank methods and the total variation (TV) based approaches have a good effect on reducing noise in HSIs. Here, we propose a non-local patch based rank-constraint HSIs noise suppression methods with a global 3-D anisotropic total variation (NLRATV). Differing from previous patch-based methods which usually ignore spatial structural information, we add more structural constraints with the non-local similarity across patches for suppressing the structural noise that exists at the same location of many bands. Besides, we utilize the global 3-D anisotropic total variation to ensure its smoothness in spatial and spectral dimensionalities while reconstructing the image. The augmented Lagrange multiplier method is adopted to optimize the proposed algorithm. The real data experiments have proved the superiority of NLRATV in decreasing mixed and dense noise.

1. Introduction

HSIs use tens or hundreds of spectral bands to capture the same ground objects by the high spectral resolution sensors. Compared with classical imaging methods, HSIs contain more spectral information which is greatly useful for substances identification. Therefore, they have been widely applied in biomedicine, agriculture, geology, etc. However, during HSIs acquisition, it's inevitable to suffer a blend of image details and many kinds of noise called mixed noise, containing addition noise, like Gaussian noise, and sparse noise, like stripes and dead lines, etc., which greatly limits HSIs' subsequent applications, like classification [1], unmixing [2], target recognition [3], etc. So HSIs denoising is important for most of their applications.

HSIs contain two spatial dimensions and one spectral dimension. The reflectivity of the same spatial pixel of HSIs in different bands is different, which can be regarded as a spectral characteristic curve of ground objects. The causes of HSIs mixed noise include atmospheric interference, dark current, the non-uniformity of detector response, photon effect, etc. Generally, the noise contaminates the three dimensions of HSIs, resulting in blurry images and distorted spectral curves.

Some excellent gray-level images denoising methods have been directly applied or extended to denoising HSIs band by band, including K-SVD [4], BM3D [5], NCSR [6], etc. But these methods ignore the strong spectral correlation between bands of the HIS. Principal component analysis (PCA) [7] is a typical method for denoising. But it's sensitive to outliers and can just work while images are corrupted by small amount of Gaussian noise. It can't perform well in HSIs denoising because HSIs often suffer mixed noise and dense noise that makes some bands almost overlapped by noise [8-9].

Robust PCA (RPCA) is proposed by Candès et al. in [10], which is more steady to outliers, and is proved highly useful for recovering the low-rank component and removing the sparse noise of HSIs [11-15]. The HSIs are divided into overlapping patches in Low-rank matrix recovery (LRMR) [11], and the authors use the Go Decomposition algorithm [16] to iteratively solve the clean HSI. In [12], authors introduce a new HSIs denoising method, i.e., NAILRMA, which explores different noise intensity in different HSIs bands. To improve the denoising effects, some researchers consider spatio-spectral characteristics. Othman et al. [17] perform the wavelet shrinkage operator in spectral derivative-domain. Group low-rank representation (GLRR) [18] is an efficient denoising method that divides patches into some groups and adds spatial constraint using non-local similarity across patches.

Total variation (TV) is also an effective method and has been used for HSI denoising. Yuan et al. propose a spectral–spatial adaptive total variation (SSAHTV), using spectral and spatial differences to improve denoising effect [19]. He et al. [8] combine the low-rank method and TV regularization (LRTV), where TV is calculated in spatial domain. Aggarwal et al. [20] propose a 3-D spatio-spectral total variation (SSTV) that uses the spatial and spectral smoothness. Liu et al. consider the spatio-spectral smoothness and correlation between bands and propose low-rank constraint SSTV [21]. In [9], He at al. use the patch-based rank-constraint matrix factorization to remove the addition noise and adopt a 3-D anisotropic TV norm to ensure spatio-spectral smoothness. Reference [22] introduce a new TV norm that combines the non-local self-similarity with spatio-spatial correlation and gains a good denoising effect.

Here, we propose a novel noise reduction method, NLRATV, exploiting non-local rank-constraint matrix factorization and 3-D anisotropic total variation. More structural constraints is adopted by adding the non-local similarity between across patches, which helps suppress structural noise existing the same location of many bands, like structural stripes. A 3-D anisotropic total variation is used to preserve the spatio-spectral smoothness. The real datasets experiment results have demonstrated the superiority of the proposed algorithm.

2. Related Work

2.1. Patch Based Rank-Constraint Method

HSIs are usually degraded by mixed noise, including addition noise, like Gaussian noise, and sparse noise, like impulse noise, stripes, and dead lines, etc. The degradation of the HSI can be modeled as

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

where $\mathcal{Y}, \mathcal{X}, \mathcal{S}, \mathcal{N} \in \mathbb{R}^{M \times N \times p}$ denote observed images, denoised HSI, sparse noise, and addition noise, respectively. The main purpose of HSIs denoising is to obtain a clean image $\mathcal{X}$ while preserving the image details. $\mathcal{Y}, \mathcal{X}, \mathcal{S}$, and $\mathcal{N}$ are reshaped into Casorati matrices $\mathbf{Y}$, $\mathbf{X}$, $\mathbf{S}$ and $\mathbf{N} \in \mathbb{R}^{MN \times p}$, where each band is converted into a column vector $\in \mathbb{R}^{MN \times 1}$, and the column vectors are arranged column-wise. The Casorati matrix form of model (1) is showed as follow:

$$\mathbf{Y} = \mathbf{X} + \mathbf{S} + \mathbf{N} \tag{2}$$

Many researchers exploit the strong correlations of adjacent bands in HSIs to improve rank-constraint denoising methods based on RPCA [4-6], [10], which can be written as

$$\min_{\mathbf{X},\,\mathbf{S}} \|\mathbf{X}\|_* + \lambda\|\mathbf{S}\|_1, \quad \text{s.t. } \|\mathbf{Y} - \mathbf{X} - \mathbf{S}\|_F^2 \le \epsilon, \quad \text{rank}(\mathbf{X}) \le r \tag{3}$$

where r is the upper bound of the rank value of $\mathbf{X}$ and λ is a regularization parameter. $\|\cdot\|_*$, $\|\cdot\|_1$ and $\|\cdot\|_F$ stands for nuclear norm, L1-norm and Frobenius norm of a matrix.

Formula (3) can be adopted to denoise a variety of noise in HSIs. The spatial dimensionality of HSIs is much greater than the spectral one ($MN \gg p$), which results in blurring while using model (3) [9]. So the patch based rank-constraint methods are studied and have been proved a good effect on

HSIs denoising [8-9], [23], which divide the HSI into subcubes centered at location (i,j) of size m × m × p and optimize each patch. The degradation model and its Casorati matrix form is as follow:

$$\mathcal{Y}_{(i,j)} = \mathcal{X}_{(i,j)} + \mathcal{S}_{(i,j)} + \mathcal{N}_{(i,j)}, \quad \mathbf{Y}_{(i,j)} = \mathbf{X}_{(i,j)} + \mathbf{S}_{(i,j)} + \mathbf{N}_{(i,j)} \tag{4}$$

The patch based rank-constraint denoising model is as follow:

$$\min_{\mathbf{X}_{(i,j)},\mathbf{S}_{(i,j)}} \left\|\mathbf{X}_{(i,j)}\right\|_* + \lambda\left\|\mathbf{S}_{(i,j)}\right\|_1$$
$$\text{s.t.} \quad \left\|\mathbf{Y}_{(i,j)} - \mathbf{X}_{(i,j)} - \mathbf{S}_{(i,j)}\right\|_{\mathrm{F}}^2 \le \epsilon, \quad \operatorname{rank}\left(\mathbf{X}_{(i,j)}\right) \le r \tag{5}$$

2.2. SSTV Model

The rank-constraint RPCA can't efficiently remove the Gaussian noise and the structural sparse noise which is usually regarded as low-rank part [23]. Thus researchers resort to total variation (TV) norm [24] based regularization which is able to capture the piecewise smoothness structure and extend it to HSIs denoising [8], and [25-26]. There are isotropic and anisotropic TV norm [27]. The former usually blurs images, while the latter preserves the textures. So the latter is researched more [8], and [27-28]. The 2-D TV norm is calculated band by band, ignoring spectral information of HSIs. So researches focus on 3-D anisotropic TV (3DATV) norm or spatial-spectral TV (SSTV) [9], [19-24] and [26]. The representation of them is as follow:

$$\|\mathcal{X}\|_{3DATV} = \tau_i\|\boldsymbol{D}_i\mathcal{X}\|_1 + \tau_j\left\|\boldsymbol{D}_j\mathcal{X}\right\|_1 + \tau_b\|\boldsymbol{D}_b\,\mathcal{X}\|_1, \text{ and } \begin{bmatrix}\boldsymbol{D}_i\mathcal{X}\\ \boldsymbol{D}_j\mathcal{X}\\ \boldsymbol{D}_b\mathcal{X}\end{bmatrix} = \begin{bmatrix}\mathcal{X}(i+1,j,b) - \mathcal{X}(i,j,b)\\ \mathcal{X}(i,j+1,b) - \mathcal{X}(i,j,b)\\ \mathcal{X}(i,j,b+1) - \mathcal{X}(i,j,b)\end{bmatrix} \tag{6}$$

where $\boldsymbol{D}_i$, $\boldsymbol{D}_j$, $\boldsymbol{D}_b$ denotes finite-difference operators, and $\tau_i = \tau_j = 1$, $\tau_b = 0.5$, [28].

3. Proposed Method

3.1. Non-local Rank-Constraint HSIs Denoising Method with 3-D Anisotropic Total Variation

The strong correlations of adjacent bands is the most important characteristic of HSIs, which is greatly helpful to HSI denoising. We propose a non-local patch based rank-constraint HSIs noise suppression methods with a global 3-D anisotropic total variation (NLRATV). Differing previous algorithms, the local similarity in the patches, the non-local similarity across patches, and a global SSTV norm is collaborated to improve the HSIs denoising.

First, we extract subcubes from original HSI and reshape them into Casorati matrices also named as patches. The step length and the spatial size and the number of subcubes are l and m, and $\mathrm{K} = \left\lceil\frac{M-m}{l}\right\rceil \times \left\lceil\frac{N-m}{l}\right\rceil$ respectively, where $\lceil * \rceil$ rounds down to the nearest integer. We define an operator $\mathcal{T}_{(n)}(*)$ to extract a subcube $\in \mathbb{R}^{m\times m\times p}$ located at (i,j) and reshape it into a patch $\mathbf{Y}_{(n)} \in \mathbb{R}^{\mathrm{m}^2\times \mathrm{p}}$, i.e.

$$\mathbf{Y}_{(n)} = \mathcal{T}_{(n)}(\mathcal{Y}), \quad n = 1, 2, \cdots, \mathrm{K} \tag{7}$$

After these patches are extracted, we calculate the Euclidean distance across two patches, i.e.

$$\boldsymbol{\mathcal{D}}_{(nt)} = \left\|\mathbf{Y}_{(n)} - \mathbf{Y}_{(t)}\right\|_{\mathrm{F}}^2 \quad n, t = 1, 2, \cdots, \mathrm{K} \tag{8}$$

For each patch, we find k similar patches which are the least Euclidean distance. The indices of similar patches for the patch n are stored in $\mathcal{g}_n = \{\mathcal{g}_{n1}, \mathcal{g}_{n2}, \cdots, \mathcal{g}_{nk}\}$ to help image reconstruction. The similar patches of the patch n are reshaped into a group $\mathcal{G}_{(n)}(\mathcal{Y}) \in \mathbb{R}^{k\mathrm{m}^2\times \mathrm{p}}$, i.e.

$$\mathcal{G}_{(n)}(\mathcal{Y}) = \left[\boldsymbol{Y}_{(\mathcal{g}_{n1})}^T, \boldsymbol{Y}_{(\mathcal{g}_{n2})}^T, \cdots, \boldsymbol{Y}_{(\mathcal{g}_{n3})}^T\right]^T \quad n = 1, 2, \cdots, \mathrm{K} \tag{9}$$

which utilizes the local similarity and non-local similarity, bringing extral structural information.

Based on above analysis, we propose a NLRATV model to denoise HSIs, i.e.

$$\min_{\mathcal{X},\mathcal{S}} \sum_{n=1}^{K} \left(\left\| \mathcal{G}_{(n)}(\mathcal{X}) \right\|_* + \lambda \left\| \mathcal{G}_{(n)}(\mathcal{S}) \right\|_1 \right) + \tau \|\mathcal{X}\|_{SSTV}$$

$$s.t. \left\| \mathcal{G}_{(n)}(\mathcal{Y}) - \mathcal{G}_{(n)}(\mathcal{X}) - \mathcal{G}_{(n)}(\mathcal{S}) \right\|_F^2 \le \epsilon, \ \operatorname{rank}\left(\mathcal{G}_{(n)}(\mathcal{X}) \right) \le r \tag{10}$$

3.2. Optimization Procedure

To be convenient for optimization operation, the function (10) can be reformulated as

$$\min_{\mathcal{X},\mathcal{S}} \sum_{n=1}^{K} \left(\left\| \mathcal{G}_{(n)}(\mathcal{L}) \right\|_* + \lambda \left\| \mathcal{G}_{(n)}(\mathcal{S}) \right\|_1 \right) + \tau \|\mathcal{U}\|_1 \quad s.t., \mathcal{G}_{(n)}(\mathcal{L}) = \mathcal{G}_{(n)}(\mathcal{J}),$$
$$\mathcal{J} = \mathcal{X}, \mathcal{U} = \boldsymbol{D}\mathcal{X}, \ \left\| \mathcal{G}_{(n)}(\mathcal{Y}) - \mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{S}) \right\|_F^2 \le \epsilon, \ \operatorname{rank}\left(\mathcal{G}_{(n)}(\mathcal{L}) \right) \le r \tag{11}$$

where $\boldsymbol{D} = \left[\tau_i \boldsymbol{D}_i, \tau_j \boldsymbol{D}_j, \tau_b \boldsymbol{D}_b \right]$ denotes the SSTV, and $\mathcal{L}, \mathcal{J} \in \mathbb{R}^{M \times N \times p}$, and $\mathcal{U} \in \mathbb{R}^{M \times N \times p \times 3}$ are auxiliary variables. Formula (11) is able to be efficiently optimized using the augmented Lagrange multiplier (ALM) [29] and can be translated into the following function:

$$\begin{aligned} min\, L(\mathcal{L}, \mathcal{S}, \mathcal{J}, \mathcal{X}, \mathcal{U}, \Lambda_1, \Lambda_2, \Lambda_3, \Upsilon) = & \min_{\mathcal{L}, \mathcal{S}, \mathcal{J}, \mathcal{X}, \mathcal{U}, \Lambda_1, \Lambda_2, \Lambda_3, \Upsilon} \sum_{n=1}^{K} \left(\left\| \mathcal{G}_{(n)}(\mathcal{L}) \right\|_* + \lambda \left\| \mathcal{G}_{(n)}(\mathcal{S}) \right\|_1 \right. \\ & + \langle \Lambda_1, \mathcal{G}_{(n)}(\mathcal{Y}) - \mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{S}) \rangle + \frac{\mu}{2} \left\| \mathcal{G}_{(n)}(\mathcal{Y}) - \mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{S}) \right\|_F^2 \\ & + \langle \Lambda_2, \mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{J}) \rangle + \frac{\mu}{2} \left\| \mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{J}) \right\|_F^2 \Big) + \tau \|\mathcal{U}\|_1 + \langle \Upsilon, \mathcal{U} - \boldsymbol{D}\mathcal{X} \rangle \\ & + \frac{\mu}{2} \|\mathcal{U} - \boldsymbol{D}\mathcal{X}\|_F^2 + \langle \Lambda_3, \mathcal{J} - \mathcal{X} \rangle + \frac{\mu}{2} \|\mathcal{J} - \mathcal{X}\|_F^2 \ \text{ s.t.}, \operatorname{rank}\left(\mathcal{G}_{(n)}(\mathcal{L}) \right) \le r. \end{aligned} \tag{12}$$

where μ is the penalty parameter, and Λ_1, Λ_2, Λ_3, and Υ are the Lagrange multipliers. $\Upsilon = [\Upsilon_1, \Upsilon_2, \Upsilon_3]$, and $\mathcal{U} = [\mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3]$. We divide formula (12) into two subproblems and optimize them iteratively: updating ($\mathcal{L}$,$\mathcal{S}$) and updating ($\mathcal{J}$,$\mathcal{X}$,$\mathcal{U}$) with other parameters fixed respectively.

3.2.1. Updating ($\mathcal{L}$, $\mathcal{S}$). The function for each group $\mathcal{G}_{(n)}(\mathcal{L})$ and $\mathcal{G}_{(n)}(\mathcal{S})$ can be simplified as

$$\begin{aligned} \arg\min_{\mathcal{G}_{(n)}(\mathcal{L}), \mathcal{G}_{(n)}(\mathcal{S})} & \left\| \mathcal{G}_{(n)}(\mathcal{L}) \right\|_* + \lambda \left\| \mathcal{G}_{(n)}(\mathcal{S}) \right\|_1 + \langle \Lambda_1, \mathcal{G}_{(n)}(\mathcal{Y}) - \mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{S}) \rangle \\ & + \frac{\mu}{2} \left\| \mathcal{G}_{(n)}(\mathcal{Y}) - \mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{S}) \right\|_F^2 + \langle \Lambda_2, \mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{J}) \rangle \\ & + \frac{\mu}{2} \left\| \mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{J}) \right\|_F^2 \quad \text{s.t.}, \operatorname{rank}\left(\mathcal{G}_{(n)}(\mathcal{L}) \right) \le r. \end{aligned} \tag{13}$$

The subproblems for variables ($\mathcal{G}_{(n)}(\mathcal{L})$, $\mathcal{G}_{(n)}(\mathcal{S})$) are shown as follow:

$$\begin{aligned} \mathcal{G}_{(n)}(\mathcal{L}): \arg\min_{\operatorname{rank}\left(\mathcal{G}_{(n)}(\mathcal{L}) \right) \le r} & \left\| \mathcal{G}_{(n)}(\mathcal{L}) \right\|_* + \frac{\mu}{2} \left\| \mathcal{G}_{(n)}(\mathcal{Y}) - \mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{S}) + \frac{\Lambda_1}{\mu} \right\|_F^2 \\ & + \frac{\mu}{2} \left\| \mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{J}) + \frac{\Lambda_2}{\mu} \right\|_F^2 = \arg\min_{\operatorname{rank}\left(\mathcal{G}_{(n)}(\mathcal{L}) \right) \le r} \left\| \mathcal{G}_{(n)}(\mathcal{L}) \right\|_* \\ & + \mu \left\| \mathcal{G}_{(n)}(\mathcal{L}) - \left(\mathcal{G}_{(n)}(\mathcal{Y}) - \mathcal{G}_{(n)}(\mathcal{S}) + \mathcal{G}_{(n)}(\mathcal{J}) \right)/2 - (\Lambda_1 - \Lambda_2)/(2\mu) \right\|_F^2 \end{aligned} \tag{14}$$

$$\mathcal{G}_{(n)}(\mathcal{S}): \arg\min_{\mathcal{G}_{(n)}(\mathcal{S})} \lambda\|\mathcal{G}_{(n)}(\mathcal{S})\|_1 + \frac{\mu}{2}\left\|\mathcal{G}_{(n)}(\mathcal{Y}) - \mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{S}) + \frac{\Lambda_1}{\mu}\right\|_F^2 \tag{15}$$

The function (14) can be solved by the singular value shrinkage operator [30-31], i.e.

$$\mathcal{G}_{(n)}(\mathcal{L}) = \mathfrak{D}_{1/2\mu}\left(\left(\mathcal{G}_{(n)}(\mathcal{Y}) - \mathcal{G}_{(n)}(\mathcal{S}) + \mathcal{G}_{(n)}(\mathcal{J})\right)/2 + (\Lambda_1 - \Lambda_2)/(2\mu)\right) \tag{16}$$

where $\mathfrak{D}_{1/2\mu}(*) = \mathrm{diag}\left\{\max\left(\sigma_i - \frac{1}{2\mu}, 0\right)\right\}$, and σ_i denotes the singular value of the matrix $*$. Function (15) is able to be optimized by the soft-thresholding (shrinkage) operator [9], [29]:

$$\Re_{\lambda/\mu}(x) = \begin{cases} x - \lambda/\mu, & \text{if } x > \lambda/\mu \\ x + \lambda/\mu, & \text{if } x < -\lambda/\mu \\ 0, & otherwise \end{cases} \tag{17}$$

that is, $\mathcal{G}_{(n)}(\mathcal{S})$ can be formulated as

$$\mathcal{G}_{(n)}(\mathcal{S}) = \Re_{\lambda/\mu}\left(\mathcal{G}_{(n)}(\mathcal{Y}) - \mathcal{G}_{(n)}(\mathcal{L}) + \Lambda_1/\mu\right) \tag{18}$$

3.2.2. Updating $(\mathcal{J}, \mathcal{X}, \mathcal{U})$: The formula for $(\mathcal{J}, \mathcal{X}, \mathcal{U})$ is simplified as

$$\begin{aligned} \arg\min(\mathcal{J}, \mathcal{X}, \mathcal{U}) = \min_{\mathcal{J},\mathcal{X},\mathcal{U}} & \sum_{n=1}^{K}\left(\langle \Lambda_2, \mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{J})\rangle + \frac{\mu}{2}\left\|\mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{J})\right\|_F^2\right) \\ & + \tau\|\mathcal{U}\|_1 + \langle \Upsilon, \mathcal{U} - \boldsymbol{D}\mathcal{X}\rangle + \frac{\mu}{2}\|\mathcal{U} - \boldsymbol{D}\mathcal{X}\|_F^2 + \langle \Lambda_3, \mathcal{J} - \mathcal{X}\rangle + \frac{\mu}{2}\|\mathcal{J} - \mathcal{X}\|_F^2 \end{aligned} \tag{19}$$

The subproblems for variables $(\mathcal{J}, \mathcal{X}, \mathcal{U})$ is as follow:

$$\mathcal{J}: \quad \arg\min_{\mathcal{J}} \frac{\mu}{2}\left\|\mathcal{J} - \mathcal{X} + \frac{\Lambda_3}{\mu}\right\|_F^2 + \sum_{n=1}^{K}\left(\frac{\mu}{2}\left\|\mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{J}) + \frac{\Lambda_2}{\mu}\right\|_F^2\right) \tag{20}$$

$$\mathcal{X}: \quad \arg\min_{\mathcal{X}} \frac{\mu}{2}\left\|\mathcal{U} - \boldsymbol{D}\mathcal{X} + \frac{\Upsilon}{\mu}\right\|_F^2 + \frac{\mu}{2}\left\|\mathcal{J} - \mathcal{X} + \frac{\Lambda_3}{\mu}\right\|_F^2 \tag{21}$$

$$\mathcal{U}: \quad \arg\min_{\mathcal{X}} \tau\|\mathcal{U}\|_1 + \frac{\mu}{2}\left\|\mathcal{U} - \boldsymbol{D}\mathcal{X} + \frac{\Upsilon}{\mu}\right\|_F^2 \tag{22}$$

According to function (20), we can get the approximate solution of $\mathcal{J}$ [9], i.e.

$$\mathcal{J} = \left((\mathcal{X} - \Lambda_3/\mu) + \sum_{n=1}^{K}\mathcal{G}_{(n)}^T\left(\mathcal{G}_{(n)}(\mathcal{L}) + \Lambda_2/\mu\right)\right)\left(1 + \sum_{n=1}^{K}\mathcal{G}_{(n)}^T\mathcal{G}_{(n)}\right)^{-1} \tag{23}$$

where, $\mathcal{G}_{(n)}^T$is the inverse transform of $\mathcal{G}_{(n)}$ defined in formula (9). $\mathcal{G}_{(n)}^T$ contains two operators. First, we translate groups into patches by calculating the weighted sum of patches which have the same index, based on the stored indices in $\mathscr{g}_n$. Second, we reshape the patches into the form of HSI.

The fast Fourier transform (FFT) is able to approximate $\mathcal{X}$ in function (21), i.e.

$$\mathcal{X} = \mathcal{F}^{-1}\left[\mathcal{F}\left(\left(\mathcal{J} + \frac{\Lambda_3}{\mu}\right) + \boldsymbol{D}^T\left(\mathcal{U} + \frac{\Upsilon}{\mu}\right)\right)\left(1 + \left(\mathcal{F}(\tau_i\boldsymbol{D}_i)\right)^2 + \left(\mathcal{F}(\tau_j\boldsymbol{D}_j)\right)^2 + \left(\mathcal{F}(\tau_b\boldsymbol{D}_b)\right)^2\right)^{-1}\right] \tag{24}$$

Function (22) is also optimized using soft-thresholding (shrinkage) operator, i.e.

$$\mathcal{U}_1 = \Re_{\frac{\tau}{\mu}}\left(\tau_i\boldsymbol{D}_i\mathcal{X} - \frac{\Upsilon_1}{\mu}\right), \ \mathcal{U}_2 = \Re_{\frac{\tau}{\mu}}\left(\tau_j\boldsymbol{D}_j\mathcal{X} - \frac{\Upsilon_2}{\mu}\right), \ \mathcal{U}_3 = \Re_{\frac{\tau}{\mu}}\left(\tau_b\boldsymbol{D}_b\mathcal{X} - \frac{\Upsilon_3}{\mu}\right) \tag{25}$$

After the variables ($\mathcal{L}, \mathcal{S}, \mathcal{J}, \mathcal{X}, \mathcal{U}$) are all be solved, the Lagrange multipliers $\Lambda_1, \Lambda_2, \Lambda_3$, and Υ will be updated:

$$\Lambda_1 = \Lambda_1 + \mu\left(\mathcal{G}_{(n)}(\mathcal{Y}) - \mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{S})\right), \Lambda_3 = \Lambda_3 + \mu(\mathcal{J} - \mathcal{X})$$
$$\Lambda_2 = \Lambda_2 + \mu\left(\mathcal{G}_{(n)}(\mathcal{L}) - \mathcal{G}_{(n)}(\mathcal{J})\right), \Upsilon = \Upsilon + \mu(\mathcal{U} - \boldsymbol{D}\mathcal{X}) \quad (26)$$

The implementation steps of NLRATV is showed in Algorithm 1.

Algorithm 1 Algorithm for NLRATV

Input: $\mathcal{Y} \in \mathbb{R}^{M\times N\times p}$, rank r, patch size m, patch number k in a group, ϵ, λ, μ, τ
Output: $\mathcal{X} \in \mathbb{R}^{M\times N\times p}$
Initialization: $\mathcal{L}^{(0)}$, $\mathcal{X}^{(0)}$, $\mathcal{S}^{(0)}$, $\mathcal{J}^{(0)}$, $\mathcal{U}^{(0)} = 0$, $\Lambda_1^{(0)}, \Lambda_2^{(0)}, \Lambda_3^{(0)}, \Upsilon^{(0)} = 0$, $\mu^{(0)} = 10^{-2}, \mu_{\max} = 10^6$, $\rho = 1.5$, maxiter $= 50$, $\epsilon = 10^{-6}$, $iter = 1$.
Step1: Get $\mathcal{G}_{(n)}$for (n = 1, 2, ⋯, K) using (9).
Step2: Repeat until convergence or $iter >$ maxiter
Update all $\left(\mathcal{G}_{(n)}^{(iter+1)}(\mathcal{L}), \mathcal{G}_{(n)}^{(iter+1)}(\mathcal{S})\right)$ groups using (16) and (18).
Update $\left(\mathcal{J}^{(iter+1)}, \mathcal{X}^{(iter+1)}, \mathcal{U}^{(iter+1)}\right)$ using (23), (24) and (25).
Update Lagrange multipliers $\Lambda_1{}^{(iter+1)}$, $\Lambda_2{}^{(iter+1)}$, $\Lambda_3{}^{(iter+1)}$, and $\Upsilon^{(iter+1)}$ using (26).
Check the convergence conditions:
$\mu = \mu \times \rho, \quad \mu \in \left(\rho\mu^{(0)}, \mu_{\max}\right), \ iter = iter + 1.$
$\max\left\{\left\|\mathcal{G}_{(n)}^{(iter+1)}(\mathcal{Y}) - \mathcal{G}_{(n)}^{(iter+1)}(\mathcal{L}) - \mathcal{G}_{(n)}^{(iter+1)}(\mathcal{S})\right\|_\infty, \left\|\mathcal{G}_{(n)}^{(iter+1)}(\mathcal{L}) - \mathcal{G}_{(n)}^{(iter+1)}(\mathcal{J})\right\|_\infty, \left\|\mathcal{U}^{(iter+1)} - \boldsymbol{D}\mathcal{X}^{(iter+1)}\right\|\right\} \le \epsilon.$

4. Experimental Results

We conduct several real data experiments to compare NLRATV with other typical algorithm in HSI denoising. There are six HSI denoising algorithms for comparison, i.e., LRMR [11], GLRR [18], LRTV [8], SSTV [20] and LLRSSTV [9]. LRMR and LRTV are patch-based HSIs noise removal methods using RPCA. SSTV is an efficient denoising approach based on 3-D TV norm. GLRR utilize the non-local similarity in denoising model. LLRSSTV combine the patch-based rank-constraint method and SSTV norm. We get all the codes of these algorithms from the authors except GLRR, and the code of GLRR is implemented by us.

We use two datasets, i.e., EO-1 Hyperion dataset and the AVIRIS Indian Pines dataset (figure 1), to test above algorithms and the proposed method. The used EO-1 dataset $\in \mathbb{R}^{200\times200\times166}$ is a subcube of original image $\in \mathbb{R}^{400\times1000\times242}$ after removing the water absorption bands and the Indian Pines dataset $\in \mathbb{R}^{145\times145\times200}$ was acquired by the NASA AVIRIS instrument over the Indian Pines test site in 1992 [8]. During the experiments, the block scale and step length are set as $m = 20$ and $l = 10$. The rank of EO-1 Hyperion dataset and AVIRIS Indian Pines dataset is set as $r = 4$ and $r = 10$. The patch number in a group is $k = 4$ [18]. $\epsilon = 10^{-6}$, $\lambda = 0.2$, $\mu = 0.01$, and $\tau = 0.005$ are set according to previous experiment results [8-9], [11].

4.1. Results in EO-1 Hyperion Dataset

This dataset is mainly degraded by stripes and deadlines and even some of them are structural. Denoised images of the 1st and 165th band are showed in figure 2 and figure 3 respectively. The experiments showed in figures 2 and 3 proved that the proposed NLRATV obviously has the best effect on mixed noised removal. LRMR can not remove some stripes which exists at the same location in many spectral bands. GLRR has a better effect on stripes because it exploits more spatial information while combining the local similarity and the non-local similarity, but it's not able to remove noise completely. With using the spatial smoothness band by band, LRTV improves the

performance of HSI denoising method. But it usually destroys some image details. SSTV introduces a 3-D spatio-spectral TV which can remove more Gaussian noise while significantly preserves image details, but it's not very helpful for reducing stripes noise. LLRSSTV is an effective denoising method which combines the local similarity and SSTV and can remove mixed noise, including Gaussian noise, impulse noise, stripes, and dead lines, etc. The figures 2(g) and 3(g) show the proposed NLRATV can wipe off the mixed noise and protect the details simultaneously. Compared with LLRSSTV, NLRATV has a better performance in removing mixed noise and protecting details.

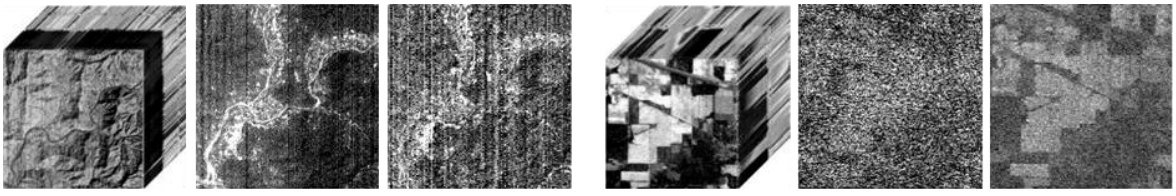

Figure 1. Figure (a) is the EO-1 dataset, and (b) and (c) are the 1th and 165th band of it; (d) is the Indian Pines dataset, and (e) and (f) are the 105th and 149th band of it.

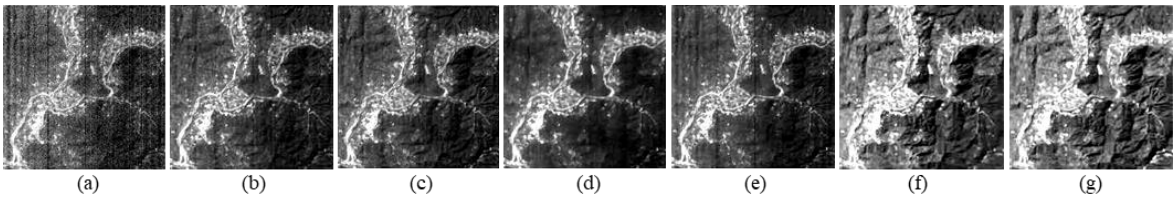

Figure 2. The denoising results of band 1 for EO-1 Hyperion dataset. (a) Original. (b) LRMR. (c) GLRR. (d) LRTV. (e) SSTV. (f) LLRSSTV. (g) NLRATV.

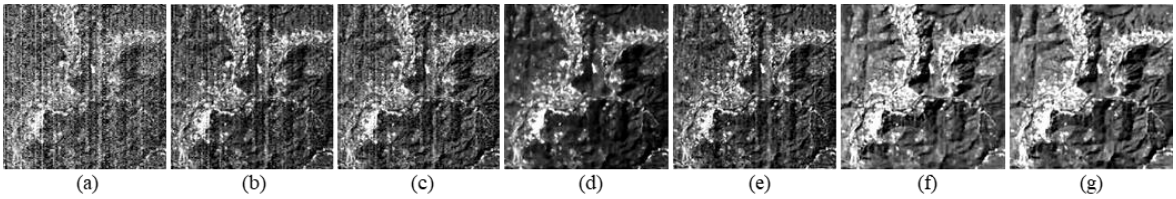

Figure 3. The denoising results of band 165 for EO-1 Hyperion dataset. (a) Original. (b) LRMR. (c) GLRR. (d) LRTV. (e) SSTV. (f) LLRSSTV. (g) NLRATV.

4.2. Results in AVIRIS Indian Pines Dataset

Some bands of this dataset are severely overlapped by dense noise, so that they are quite difficult to be restored, such as the band 105 and band 149, etc. As showed in figure 4 and 5, we can hardly see the details in the original images. Although facing extreme corruption, NLRATV still restore some details of the ground objects. Figures 4(b) and 5(b) prove that LRMR can't remove the dense noise and preserve the image textures well so that the recovered results are still too blurring. After analysing figure 4(b)-4(d) and 5(b)-5(d), we can find that patch-based rank-constraint methods are indeed beneficial to HSI denoising. The restorations reconstructed from the original images with dense noise by SSTV in figure 4(e) and 5(e) are distorted, perhaps mainly because it ignores the strong correlations of adjacent bands without the help of rank-constraint methods. LLRSSTV and its improved counterpart, i.e., NLRATV, have a good effect on dense noise removal and protecting details. After comparing the figure 4(g) and 5(g) with 4(f) and 5(f) respectively, we find NLRATV is capable of reconstructing more image details and relatively stronger in the ability to remove the dense noise.

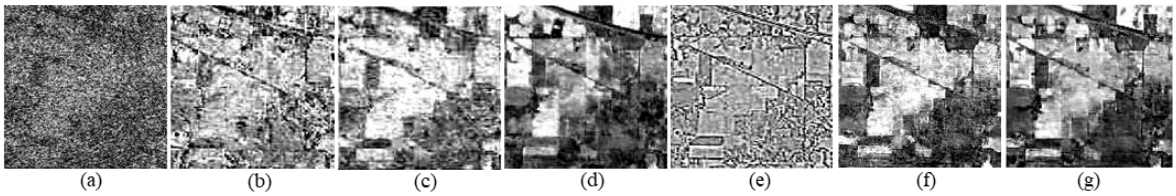

Figure 4. The denoising results of band 105 for Indian Pines dataset. (a) Original. (b) LRMR. (c) GLRR. (d) LRTV. (e) SSTV. (f) LLRSSTV. (g) NLRATV.

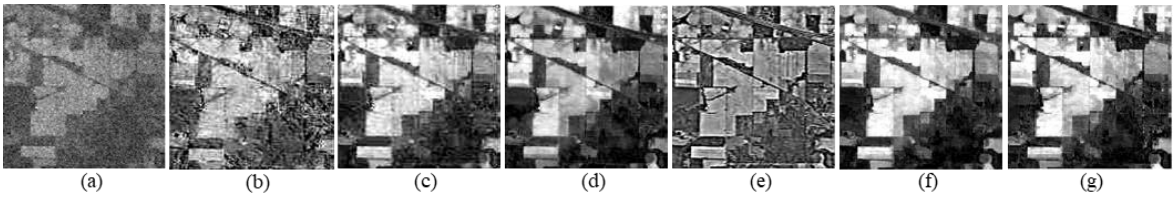

Figure 5. The denoising results of band 149 for Indian Pines dataset. (a) Original. (b) LRMR. (c) GLRR. (d) LRTV. (e) SSTV. (f) LLRSSTV. (g) NLRATV.

5. Conclusion

Hyperspectral images are contaminated by mixed noise, including addition noise and sparse noise, etc., which greatly limits their subsequent applications, like object detection, classification, unmixing, etc. Here, we have proposed a novel mixed and dense noise suppression method using rank constraint matrix factorization, non-local structural information, i.e., non-local similarity, and the global 3-D anisotropic total variation, which is efficient on suppressing the structural noise that exists at the same position in many bands. A global 3-D anisotropic TV norm is exploited to ensure the image smoothness in spatial and spectral domain and preserve image details. Several real data experiments have demonstrated NLRATV is more efficient on mixed and dense noise removal.

Though the proposed method has gained a good effect on HSIs denoising, it is worth more researches to explore further improvement. We translate the patches into matrices and perform low-rank operations on matrices. It can be extended to the tensor form to explore the 3-D rank-constraint methods for HSIs noise reduction. And the weight of three dimensions in 3-D total variation deserves more quantitative analysis. Besides, we can make a further study on the non-local similarity to discover its more powerful strength.

References

[1] Ailong M, Yanfei Z, and Bei Z, Hongzan J and Liangpei Z 2016 Semisupervised subspace-based DNA encoding and matching classifier for hyperspectral remote sensing imagery *IEEE Trans. Geosci. Remote Sens.* **54** 4402–18

[2] José M B, Antonio P and Nicolas D et al. 2012 Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.* **5** 354–79

[3] Shuo Y and Zhenwei S 2016 Hyperspectral image target detection improvement based on total variation *IEEE Trans. Image Process* **25** 2249–58

[4] Michael E and Michal A 2006 Image denoising via sparse and redundant representations over learned dictionaries *IEEE Trans. Image Process* **15** 3736–45

[5] Kostadin D, Alessandro F, Vladimir K and Karen E 2007 Image denoising by sparse 3-D transform-domain collaborative filtering *IEEE Trans. Image Process* **16** 2080–95

[6] Weisheng D, Lei Z, Guangming S and Xin L 2013 Nonlocally centralized sparse representation for image restoration *IEEE Trans. Image Process* **22** 1620–30

[7] Guangyi C and Shen-En Q 2011 Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage *IEEE Trans. Geosci. Remote Sens.* **49** 973–80

[8] Wei H, Hongyan Z, Liangpei Z and Huanfeng S 2016 Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration *IEEE Trans. Geosci. Remote Sens.* **54** 178–88
[9] Wei H, Hongyan Z, Huanfeng S and Liangpei Z 2018 Hyperspectral image denoising using local low-rank matrix recovery and global spatial–spectral total variation *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **11** 713–29
[10] Emmanuel J C, Xiaodong L, Yi M and John W 2011 Robust principal component analysis? *J. ACM* **58**
[11] Hongyan Z, Wei H, Liangpei Z, Huanfeng S and Qiangqiang Y 2014 Hyperspectral image restoration using low-rank matrix recovery *IEEE Trans. Geosci. Remote Sens.* **52** 4729-42
[12] Wei H, Hongyan Z, Liangpei Z and Huanfeng S 2015 Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.* **8** 3050–61
[13] Rui Z, Mingzhi D and Jing-Hao X 2015 Spectral nonlocal restoration of hyperspectral images with low-rank property *IEEE Trans. Geosci. Remote Sens.* **8** 3062-67
[14] Ting H, Hongyan Z, Huanfeng S and Liangpei Z 2014 Robust registration by rank minimization for multiangle hyper/multispectral remotely sensed imagery *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.* **7** 2443–57
[15] Wei He, Hongyan Z, Liangpei Z and Huanfeng S 2014 A noise-adjusted iterative randomized singular value decomposition method for hyperspectral image denoising *IEEE Geoscience and Remote Sensing Symposium* 1536–39
[16] Tianyi Z and Dacheng T 2011 Godec: randomized low-rank sparse matrix decomposition in noisy case *28th* International *Conference on Machine Learning* 33–40
[17] Hisham O and Shen-En Q 2006 Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage *IEEE Trans. Geosci. Remote Sens.* **44** 397–408
[18] Mengdi W, Jing Y, Jing-Hao X and Weidong S 2016 Denoising of hyperspectral images using group low-rank representation *IEEE Trans. Geosci. Remote Sens.* **9** 4420-27
[19] Qiangqiang Y, Liangpei Z, and Huanfeng S 2012 Hyperspectral image denoising employing a spectral–spatial adaptive total variation model *IEEE Trans. Geosci. Remote Sens.* **50** 3660–77
[20] Hemant Kumar A and Angshul M 2016 Hyperspectral image denoising using spatio-spectral total variation *IEEE Geosci. Remote Sens. Lett.* **13** 442–6
[21] Hongyi L, Peipei S, Qian Du, Zebin W, and Zhihui W 2019 Hyperspectral image restoration based on low-rank recovery with a local neighborhood weighted spectral spatial total variation model *IEEE Trans. Geosci. Remote Sens.* **57** 1409-1422
[22] Le S, Byeungwoo J, Zebin W and Liang X 2018 Hyperspectral denoising via cross total variation-regularized unidirectional nonlocal low-rank tensor approximation *25th IEEE International Conference on Image Processing* 2900–4
[23] TANER I 2019 Hyperspectral image denoising using group low-rank and spatial-spectral total variation *IEEE Access* **7** 52095-109
[24] Leonid I R, Stanley O and Emad F 1992 Nonlinear total variation based noise removal algorithms *Physica D* **60** 259–68
[25] Yi C, Luxin Y, Houzhang F and Chunan L 2015 Anisotropic spectral-spatial total variation model for multispectral remote sensing image destriping *IEEE Trans. Image Process* **24** 1852–66
[26] Qiangqiang Y, Liangpei Z and Huanfeng S 2012 Hyperspectral image denoising employing a spectral-spatial adaptive total variation model *IEEE Trans. Geosci. Remote Sens.* **50** 3660–77
[27] Stanley H C, Ramsin K, Kristofor B G, Philip E G and Truong Q N 2011 An augmented Lagrangian method for total variation video restoration *IEEE Trans. Image* Process **20** 3097–111

[28] Le S, Byeungwoo J, Yuhui Z and Zebin W 2017 Hyperspectral image restoration using low-rank representation on spectral difference image *IEEE Geoscience and Remote Sensing Letters* **14** 1151-55
[29] Zhouchen L, Minming C and Yi M 2009 The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices *arXiv*:1009.5055
[30] Jian-Feng C, Emmanuel J C and Zuowei S 2010 A singular value thresholding algorithm for matrix completion *SIAM J. Optim.* **20** 1956–82
[31] Emmanuel J Candes, Carlos A S and Joshua D T 2013 Unbiased risk estimates for singular value thresholding and spectral estimators *IEEE Trans. Signal Process* **61** 4643–57