Multishot Compressive Hyperspectral Imaging Based on Tensor Fibered Rank Minimization and Its Primal-Dual Algorithm

Ting Xie, *Member, IEEE*, Xudong Kang, *Senior Member, IEEE*, Renwei Dian, *Member, IEEE*, Tonghan Wang, and Licheng Liu, *Member, IEEE*

***Abstract*—Coded aperture snapshot spectral imaging (CASSI) compresses tens to hundreds of spectral bands of the hyperspectral image (HSI) to a 2-D compressive measurement. For spatially or spectrally rich scenes, the compressive measurement provided by a single snapshot CASSI may not be sufficient. By taking multiple snapshots of the same scene, multishot CASSI leads to a less ill-posed inverse reconstruction problem, making the CASSI system more suitable for spatially or spectrally rich HSI. Considering the strong spectral correlation of HSI and the directional characteristics of mask shifting in multishot CASSI, the mode-1 tensor fibered rank (TFR) minimization is presented for its reconstruction in this article. Specifically, the mode-1 TFR is derived from the tensor singular value decomposition (t-SVD) to the mode-1 t-SVD, and the mode-1 TFR minimization is reduced to a mode-1 tensor nuclear norm minimization problem, to achieve more accurate HSI characterization in multishot CASSI reconstruction. The primal-dual algorithm (PDA) is applied to solve the objective optimization problem, which is flexible. Experimental results on the CAVE, Cuperite, and Urban datasets demonstrate the effectiveness of the proposed method.**

***Index Terms*—Coded aperture snapshot spectral imaging (CASSI), hyperspectral imaging (HSI), primal-dual algorithm (PDA), tensor fibered rank (TFR) minimization.**

I. Introduction

HYPERSPECTRAL images (HSIs) contain tens or hundreds of spectral bands per pixel [1], [2], which provide abundant spectral information and have the potential to help us identify the different substances appearing in the scenes [3], [4], [5], [6]. Based on this, hyperspectral imaging first attracts interest in remote sensing [7], [8] and medical imaging [9], then quickly spreads into many other fields. To obtain HSI, traditional hyperspectral imaging techniques focus on a temporal sequential scanning manner, that is, scanning in the spatial domain or in the spectral domain. This approach makes data collection take a long time. In addition, since a large amount of data is acquired, this also hinders the widespread application of HSI.

With the rapid development of computational imaging, CASSI [10], [11], [12], [13], [14] has been introduced as a novel HSI imaging method, which greatly reduces acquisition time and storage space. CASSI utilizes a coded aperture and one or two dispersive elements to modulate the optical field from a scene [10], [13], [15]. Specifically, in CASSI, each spatial position of the 3-D HSI is modulated by a coded aperture. The coded scene is spectrally horizontally shifted (or vertically shifted) by one or two dispersive prisms [16], then integrated along the spectral dimension and finally captured by the detector, to obtain 2-D compressive measurement [10], [17]. We suppose the coded scene is horizontally shifted in this article. Within the CASSI paradigm, different classes of CASSI have been presented, e.g., single dispersive CASSI (SD-CASSI) [10], dual dispersive CASSI (DD-CASSI) [11], [12], and colored coded aperture [18], etc. Since the hundreds of spectral bands of HSI are compressed to 2-D measurements, resulting in a large amount of information being lost, the compressive measurements provided by a single shot CASSI is insufficient. To handle it, He et al. [15] proposed a subspace-based fusion model by combining CASSI and RGB measurements for HSI reconstruction. However, for spatially or spectrally rich scene, this fusion reconstruction performance still needs to be improved. In order to exploit more information, multishot CASSI [18], [19], [20], [21], [22], [23] has been proposed. Multishot CASSI takes multiple snapshots of the same scene, with each snapshot being encoded differently, which increases the number of the compressive measurements and greatly easing the reconstruction problem, especially for spatially or spectrally rich scenes [19], [20], [21]. On the hardware, the multishot CASSI can use digital micromirror device to change the encoding apertures of each snapshot [21], [23], [24].

HSI reconstruction plays a vital role in compressive hyperspectral imaging by solving the ill-posed inverse problem. In the

Manuscript received 30 November 2023; revised 8 January 2024; accepted 17 January 2024. Date of publication 29 January 2024; date of current version 12 February 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62101188, Grant 62261001, and Grant 62071174, and in part by the Natural Science Foundation of Hunan Province under Grant 2023JJ40450. *(Corresponding author: Licheng Liu.)*

Ting Xie is with the College of Engineering and Design, Hunan Normal University, Changsha 410081, China (e-mail: xieting@hunnu.edu.cn).

Xudong Kang is with the School of Robotics, Hunan University, Changsha 410082, China (e-mail: xudong_kang@163.com).

Renwei Dian is with the School of Robotics and the Key Laboratory of Visual Perception and Artificial Intelligence of Hunan Province, Hunan University, Changsha 410082, China (e-mail: drw@hnu.edu.cn).

Tonghan Wang is with the School of Information Engineering, East China University of Technology, Nanchang 330013, China (e-mail: thwang@ecut.edu.cn).

Licheng Liu is with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: lichenghnu@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2024.3359321

beginning, different priors have been exploited to character the properties of HSI for CASSI reconstruction, for example, sparsity in transform domain [25], [26], sparsity in over-complete dictionaries [12], [27], [28], the Gaussian mixture model-based methods [29], [30], and the total variation (TV) [21], [31], [32]. In terms of objective optimization problem solving, the most commonly used are the gradient projection for sparse reconstruction (GPSR) [33] and the two-step iterative shrinkage/thresholding (TwIST) [34]. However, the reconstruction results of these methods remain unsatisfactory, such as excessive smoothing phenomenon. To improve the reconstruction performance, more effective priors have been developed, including nonlocal self-similarity [16] based on weighted nuclear norm minimization, and spatial nonlocal self-similarity integrated with global spectral correlation of HSI [35]. To solve these objective optimization problems, the alternating direction method of multiplier (ADMM) [35] has become the most popular method. Nevertheless, those ADMM-based methods [16], [35] have special requirements on the sensing matrix and are mainly designed for the single snapshot CASSI mechanism, which makes them inflexible.

Deep learning methods [36], [37], [38] have shown promising results for CASSI reconstruction. In [36], an end-to-end convolutional neural network (CNN) is proposed for compressive hyperspectral imaging by jointly optimizing the two phases, that is, the measurement phase and the reconstruction phase. A CNN-based method [38] is proposed for coded HSI reconstruction by jointly deep external and internal learning. In [39], the spatial-spectral self-attention-based convolutional network is proposed for CASSI reconstruction by exploiting nonlocal spatial similarity and bandwise spectral correlation. By utilizing UNet to mine the deep image prior and Tucker decomposition to characterize the low-rank prior of the image, a deep prior and low-rank tensor representation (DP-LRTR) [17] is proposed for CASSI reconstruction. However, end-to-end deep learning methods usually take several days to train, and the pretrained network is generally designed for the specific CASSI system. Different CASSI systems have different target scene sizes and encoding apertures, which pose challenges in applying them to other imaging systems.

In order to exploit the inherent tensor structure of high-dimensional data, t-SVD has achieved promising results in HSI reconstruction [40], [41]. Since t-SVD lacks the flexibility to handle different correlations in different HSI modes, the literature [41] generalizes the t-SVD to the mode-k t-SVD to more accurately characterize the HSI structure in hyperspectral mixed noise removal, and accordingly, a new rank is proposed, called TFR. The main steps of the mode-k t-SVD algorithm ($k = 1, 2, 3$) are to perform the fast Fourier transformation (FFT) along the kth dimension of the tensor, and then perform SVD on the matrix formed by each slice of the transformed tensor, to obtain the factor matrix. The TFR is a combination of all mode-k ($k = 1, 2, 3$) TFR, and mode-k TFR is obtained by the rank of the slices after the FFT in the kth dimension of the tensor. Fig. 1 presents the mode-k ($k = 1, 2, 3$) t-SVD for tensor $\mathcal{X}$. Let n_1 represent the spatial height size (mode-1), n_2 denote the spatial width size (mode-2), and

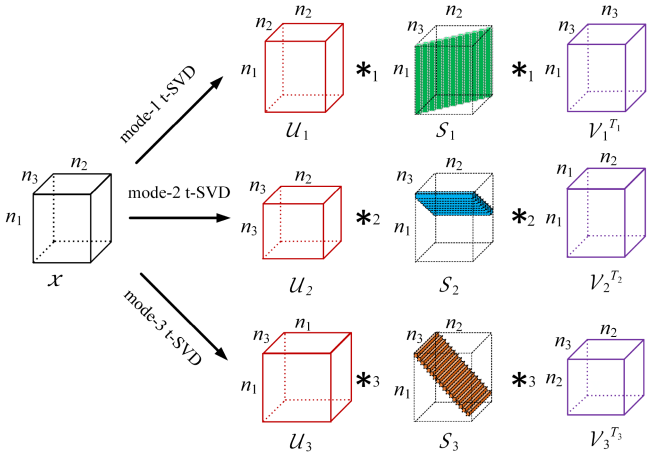


Fig. 1. Description of the mode-k ($k = 1, 2, 3$) t-SVD for HSI $\mathcal{X}$.

n_3 represent the spectral size of HSI (mode-3). According to the main steps of mode-k t-SVD and the description of Fig. 1, different modes of t-SVD characterize different correlations of HSI.

Considering the strong spectral correlation of HSI and the directional characteristics of mask shifting of CASSI, the mode-1 TFR minimization can obtain a more accurate HSI structure characterization, compared with the mode-2 TFR minimization and the mode-3 TFR minimization. First, in CASSI, tens to hundreds of HSI spectral bands are compressed into 2-D measurements, which results in a large loss of spectral information. Zheng et al. [41] indicated that the strong spectral correlation was inadequately shown by the mode-3 TFR, but exactly shown by the mode-1 TFR and mode-2 TFR. Therefore, to reconstruct the HSI from the compressive measurements, if using mode-3 t-SVD to perform the mode-3 TFR minimization constraint on the HSI, the main lost spectral information will not be well recovered and will result in suboptimal performance. Second, in this article, the mask is shifted along the spatial horizontal direction (mode-2), resulting in the appearance of horizontal stripes in the compressive measurements. We found that the mode-1 t-SVD has better horizontal stripes discrimination ability than the mode-2 t-SVD. Using the mode-1 t-SVD to perform the mode-1 TFR minimization has better horizontal stripes removal ability than the mode-2 TFR minimization. The explanation is shown in Section III-B and also demonstrated in experiments.

Bear these in mind, in this article, we propose the mode-1 TFR minimization for multishot CASSI and solves the optimization problem by PDA. By setting the coefficients of mode-2 TFR and mode-3 TFR to be zeros, the mode-1 TFR is a special case of TFR to obtain a more accurate HSI structure characterization in multishot CASSI reconstruction. Specifically, the mode-1 TFR is performed by the mode-1 t-SVD. By using the mode-1 t-SVD, the mode-1 TFR minimization is converted into the mode-1 tensor nuclear norm minimization (M1TNN) problem. To avoid matrix inversion of a large matrix, the PDA is utilized to solve the optimization problem, which makes it flexible (we note the proposed method as PDA-M1TNN).

A. Contributions of the Article

The contributions mainly include the following two points.

1) The mode-1 TFR minimization is used to exploit the spatial-spectral information of HSI in multishot CASSI reconstruction by considering the strong spectral correlation of HSI and the mask shifting along the horizontal direction in multishot CASSI.
2) The PDA is used to solve the optimization problem, which avoids matrix inversion of a large matrix and makes itself flexible for any single shot or multishot CASSI.

B. Article Organization

The rest of this article is organized as follows. Section II introduces the tensor algebra framework. The proposed PDA-M1TNN is presented in Section III. Section IV contains the experimental results, and Section V concludes this article.

II. Tensor Algebra Framework

The scalars, vectors, matrices, and tensors in the article are represented by nonbold lowercase letters (such as a), bold lowercase letters (such as $\mathbf{a}$), bold uppercase letters (such as $\mathbf{A}$), and Euler script letters (such as $\mathcal{A}$), respectively. The vectorization of tensor $\mathcal{A}$ is $\mathbf{a}$, represented by $\mathbf{a} = \texttt{vec}(\mathcal{A})$, and the inverse operator of this operator is expressed as $\texttt{ten}(\mathbf{a})$. Hence, we have $\mathcal{A} = \texttt{ten}(\texttt{vec}(\mathcal{A}))$. For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\bar{\mathcal{A}}_k \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is denoted as the FFT of $\mathcal{A}$ along the kth dimension, represented as $\bar{\mathcal{A}}_k = \texttt{fft}(\mathcal{A}, [\,], k)$. $A_1^{(i)} \in \mathbb{R}^{n_2 \times n_3}$, $A_2^{(i)} \in \mathbb{R}^{n_1 \times n_3}$, and $A_3^{(i)} \in \mathbb{R}^{n_2 \times n_3}$ are, respectively, represented as the ith lateral (mode-1), horizontal (mode-2), and frontal (mode-3) slices of $\mathcal{A}$. The following are some definitions, please refer to [41] and [42] for a detailed introduction.

Definition 1: (*t-product* [42]) The t-product $\mathcal{A} * \mathcal{B}$ for $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times l \times n_3}$, is a tensor $\mathcal{C} \in \mathbb{R}^{n_1 \times l \times n_3}$ with tubes

$$\mathcal{C}(i, j, :) = \sum_{t=1}^{n_2} \mathcal{A}(i, t, :) * \mathcal{B}(t, j, :) \tag{1}$$

where $*$ is circular convolution [41].

Definition 2: (*t-SVD* [42]) Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the t-SVD of $\mathcal{A}$ is

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T \tag{2}$$

where $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a f-diagonal tensor, $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors.

Definition 3: (*Tensor tubal rank* [42]) For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tensor tubal rank is represented as $\text{rank}_\text{t}(\mathcal{A})$, which is computed as

$$\text{rank}_\text{t}(\mathcal{A}) = \#\{i, \mathcal{S}(i, i, :) \neq 0\} \tag{3}$$

with $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$.

Definition 4: (*Mode-k t-SVD* [41]) Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Then $\mathcal{A}$ is factored as

$$\mathcal{A} = \mathcal{U}_k *_k \mathcal{S}_k *_k V_k^{T_k}, k = 1, 2, 3 \tag{4}$$

Algorithm 1: Mode-k t-SVD.

Input: $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$.
$\bar{\mathcal{A}}_k \leftarrow \texttt{fft}(\mathcal{A}, [\,], k)$.
for $i = 1$ to n_k **do**
  $[U, S, V] = \texttt{svd}((\bar{A}_k)_k^{(i)})$.
  $(\bar{U}_k)_k^{(i)} \leftarrow U$; $(\bar{S}_k)_k^{(i)} \leftarrow S$; $(\bar{V}_k)_k^{(i)} \leftarrow V$.
end while
$\mathcal{U}_k \leftarrow \texttt{ifft}(\bar{\mathcal{U}}_k, [\,], k)$.
$\mathcal{S}_k \leftarrow \texttt{ifft}(\bar{\mathcal{S}}_k, [\,], k)$.
$\mathcal{V}_k \leftarrow \texttt{ifft}(\bar{\mathcal{V}}_k, [\,], k)$.
Output: $\mathcal{U}_k, \mathcal{S}_k, \mathcal{V}_k$.

where $*_k$ is the mode-k t-product [41]. Algorithm 1 presents this decomposition. When $k = 3$, the mode-3 t-SVD is the t-SVD.

Definition 5: (*Tensor mode-k permutation* [41]) The mode-k permutation of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ denoted as $\texttt{permute}(\mathcal{A}, k)$ is defined as the tensor whose ith mode-3 slice is the ith mode-k slice of $\mathcal{A}$, and we have

$$\texttt{permute}(\mathcal{A}, k) : \mathbb{R}^{n_1 \times n_2 \times n_3} \rightarrow \mathbb{R}^{n_i \times n_j \times n_k} \tag{5}$$

where $i, j, k \in \{1, 2, 3\}$ and $i \neq j \neq k$. The inverse operator is denoted as $\texttt{ipermute}(\mathcal{A}, k)$, which is computed as

$$\texttt{ipermute}(\mathcal{A}, k) : \mathbb{R}^{n_i \times n_j \times n_k} \rightarrow \mathbb{R}^{n_1 \times n_2 \times n_3} \tag{6}$$

and we have $\mathcal{A} = \texttt{ipermute}(\texttt{permute}(\mathcal{A}, k), k)$. It should be known that the mode-k t-SVD of $\mathcal{A}$ is equivalent to the t-SVD of $\texttt{permute}(\mathcal{A}, k)$.

Definition 6: (*Mode-k TFR* [41]) Let $\mathcal{A} = \mathcal{U}_k *_k \mathcal{S}_k *_k V_k^{T_k}$ be the Mode-k t-SVD of $\mathcal{A}$. The mode-k TFR denoted as $\text{rank}_{f_k}(\mathcal{A})$, is computed as the number of nonzero mode-k fibers of $\mathcal{S}_k$, and the mode-3 TFR becomes tensor tubal rank.

Definition 7: (*TNN* [42]) Let $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$ be the t-SVD of $\mathcal{A}$. The TNN of $\mathcal{A}$ is computed as the average of the nuclear norm of all the frontal slices of $\bar{\mathcal{A}}$, namely, $\|\mathcal{A}\|_* = \frac{1}{n_3} \sum_{i,j} \bar{\mathcal{S}}(i, i, j)$.

Definition 8: (*Mode-k TNN [41]*) For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the mode-$k$ TNN, denoted as $\|\mathcal{A}\|_{\text{TNN}_k}$, is computed as the sum of singular values of all the mode-k slices of $\bar{\mathcal{A}}_k$, and the mode-3 TNN is the TNN.

III. Proposed Method

In this section, the snapshot hyperspectral imaging model is first introduced, then the mode-1 TFR minimization introduced for multishot CASSI and the optimization problem solved by the PDA is presented. Specifically, the mode-1 TFR is derived from the t-SVD to the mode-1 t-SVD, and accordingly, the mode-1 TFR minimization is reduced to a mode-1 tensor nuclear norm minimization problem. The PDA is introduced to solve the objective optimization problem.

A. Snapshot Hyperspectral Imaging Model

In snapshot compressive imaging systems, the measurement $\boldsymbol{b_1}$ of a single snapshot for the HSI $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is modeled

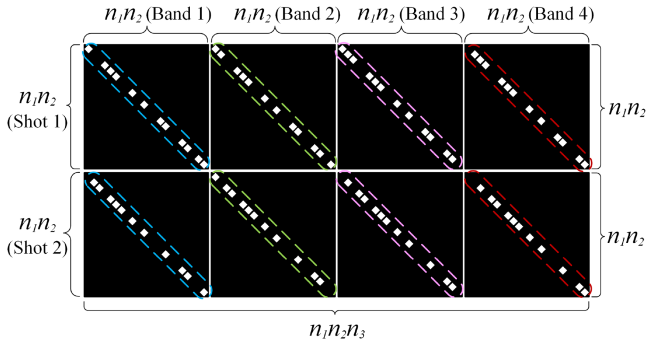


Fig. 2. Sensing matrix $\boldsymbol{H}$ for $n_3 = 4$ and $s = 2$.

as

$$\boldsymbol{b}_1 = \boldsymbol{H}_1 \boldsymbol{x} \tag{7}$$

where $\boldsymbol{H}_1 \in \mathbb{R}^{n_1 n_2 \times n_1 n_2 n_3}$ is the sensing matrix and the coding scheme is the same as [16] with the horizontally shifted random binary mask, $\boldsymbol{x} \in \mathbb{R}^{n_1 n_2 n_3}$ is the vectorized of HSI $\mathcal{X}$, expressed as $\boldsymbol{x} = \mathtt{vec}(\mathcal{X})$, and n_1, n_2, and n_3, respectively, represent the width, height, and number of spectral bands of HSI. In (9), the $\boldsymbol{H}_1$ has a special structure and could be formed as

$$\boldsymbol{H}_1 = \left[\boldsymbol{D}_{11}, \boldsymbol{D}_{12}, \quad \cdots, \quad \boldsymbol{D}_{1n_3}\right] \tag{8}$$

where $\boldsymbol{D}_{1l} \in \mathbb{R}^{n_1 n_2 \times n_1 n_2}$ $(1 \le l \le n_3)$ are diagonal matrices. This CASSI system is the same as the DD-CASSI system [11], [12], which encodes spatial and spectral information about a scene using the dual-disperser architecture. When taking multiple snapshots, the CASSI system for multishot can be formed as

$$\boldsymbol{b} = \boldsymbol{H}\boldsymbol{x} \tag{9}$$

where $\boldsymbol{b} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_s]^T \in \mathbb{R}^{n_1 n_2 s}$, $\boldsymbol{H} = [\boldsymbol{H}_1, \ldots, \boldsymbol{H}_s]^T \in \mathbb{R}^{n_1 n_2 s \times n_1 n_2 n_3}$, and s represents snapshot times. Correspondingly, (9) can be rewritten as

$$\begin{bmatrix} \boldsymbol{b_1} \\ \boldsymbol{b_2} \\ \vdots \\ \boldsymbol{b_s} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_1 \\ \boldsymbol{H}_2 \\ \vdots \\ \boldsymbol{H}_s \end{bmatrix} \boldsymbol{x}, \tag{10}$$

Fig. 2 shows the sensing matrix $\boldsymbol{H}$ for $n_3 = 4$ and $s = 2$. In addition, For all n_3 bands at the position (m, n, j), $m = 1, \cdots n_1$, $n = 1, \ldots, n_2$, and $j = 1, \cdots s$, the discrete form of one pixel in the jth snapshot is expressed as

$$\mathcal{Y}_{m,n,j} = \sum_{l=1}^{n_3} \mathcal{M}_{m,n,l,j} \mathcal{X}_{m,n,l} \tag{11}$$

where $\mathcal{M} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times s}$ denotes the discrete coded apertures for multishot CASSI.

B. Mode-1 TFR for Multishot CASSI

The mode-1 TFR minimization can obtain a more accurate HSI structure characterization, compared with the mode-2 TFR minimization and the mode-3 TFR minimization by considering the strong spectral correlation of HSI and directional characteristics of mask shifting of CASSI. Literature [41] has indicated that the strong spectral correlation is inadequately shown by the mode-3 TFR, but exactly shown by the mode-1 TFR and mode-2 TFR. Furthermore, in this article, the mask is shifted along the spatial horizontal direction (mode-2), resulting in the appearance of horizontal stripes in the compressive measurements. Fig. 3 shows the two snapshots measurements for the Urban dataset. As observed, there are numerous horizontal bands in each snapshot measurement. We found that the mode-1 t-SVD has better horizontal stripes discrimination ability than mode-2 t-SVD. This is because the main step of the mode-2 t-SVD is to perform FFT along the spatial horizontal direction (mode-2), while the mode-1 t-SVD is to perform FFT along the spatial vertical direction (mode-1). By performing FFT along the spatial vertical direction, horizontal stripes can be adequately detected, and then performing rank minimization through SVDs of lateral slices to describe spatial-spectral correlation can make horizontal stripes removed to a certain extent. To further illustrate it, we applied the Fourier transform of a clean image and a clean image contaminated by horizontal stripes, as shown in Fig. 4. As observed, the magnitude spectrums of FFT of mode-2 of the clean image and the noisy image have almost no difference. While compared with Fig. 4(b) and (e), i.e., the magnitude spectrums of FFT of mode-1 of the clean image and the noisy image, they are different, and the FFT of mode-1 can adequately detect the horizontal stripes, while for the FFT along the mode-2, this capability is lacking. Therefore, using the mode-1 t-SVD to perform the mode-1 TFR minimization has better horizontal stripes removal ability than the mode-2 TFR minimization.

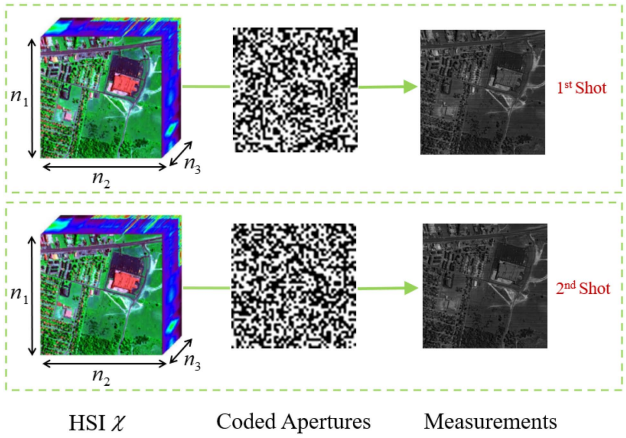


Fig. 3. Illustration of the two snapshot measurements for Urban dataset.

C. PDA-M1TNN for Multishot CASSI

To recover the HSI from the multishot CASSI compressive measurements, we consider the mode-1 TFR minimization to character the spatial-spectral correlation. Since the direct minimization of mode-1 TFR is an NP-hard problem, the mode-1 TNN is considered as the convex relaxation of the mode-1 TRF, and the objective optimization problem for multishot CASSI is

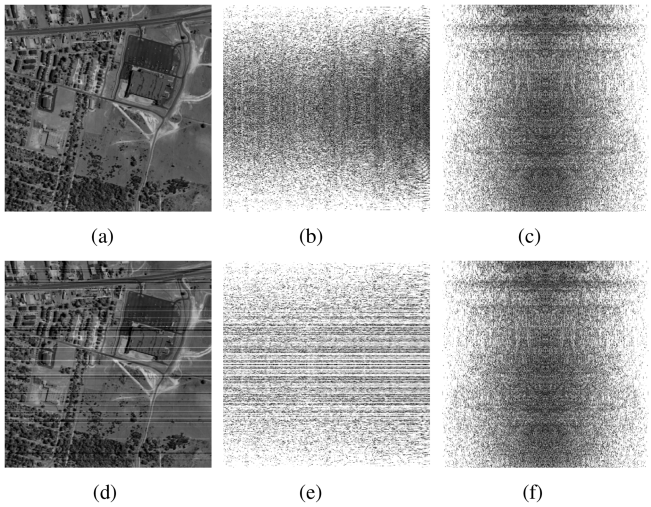


Fig. 4. Magnitude comparison of the FFT spectrum. (a) Clean image. (b) Magnitude spectrum of the FFT of each column of image (a). (c) Magnitude spectrum of FFT of each row of image (a). (d) Noisy image contaminated by the horizontal stripes to image (a). (e) Magnitude spectrum of FFT of each column of image (d). (f) Magnitude spectrum of FFT of each row of image (d). (a) Clean image. (b) $\texttt{fft}((a), n, 1)$. (c) $\texttt{fft}((a), n, 2)$. (d) Noisy image. (e) $\texttt{fft}((d), n, 1)$. (f) $\texttt{fft}((d), n, 2)$.

modeled as

$$\arg\min_{\mathcal{X}} \frac{1}{2}\|\boldsymbol{b}-\boldsymbol{H}\boldsymbol{x}\|_2^2+\lambda\|\mathcal{X}\|_{\text{TNN}_1} \tag{12}$$

where λ is the balancing factor and mode-1 TNN is denoted as $\|\cdot\|_{\text{TNN}_1}$. To solve the problem (12), the widely used ADMM [16] needs to calculate the inverse of matrix $\boldsymbol{H}^T\boldsymbol{H}+\rho\boldsymbol{I}$, where ρ is a constant. However, since the $\boldsymbol{H}\in\mathbb{R}^{n_1n_2\,s\times n_1n_2n_3}$ is a large and fat matrix, $(\boldsymbol{H}^T\boldsymbol{H})$ will be a huge matrix. In addition, it would not get $\boldsymbol{H}\boldsymbol{H}^T$ to be a diagonal matrix for multi-shot CASSI. Therefore, calculating the matrix inversion, i.e., $(\boldsymbol{H}^T\boldsymbol{H}+\rho\boldsymbol{I})^{-1}$ is difficult in ADMM. To solve this problem, we use the PDA [43] to solve (12), avoiding the matrix inversion of the large matrix $\boldsymbol{H}\boldsymbol{H}^T$, and the solution process is as follows.

Equation (12) can be written as the following general form:

$$\arg\min_{\mathcal{X}} f(\boldsymbol{H}\boldsymbol{x})+g(\mathcal{X}) \tag{13}$$

where $f(\boldsymbol{H}\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{b}-\boldsymbol{H}\boldsymbol{x}\|_2^2$ and $g(\mathcal{X})=\lambda\|\mathcal{X}\|_{\text{TNN}_1}$ are convex functions and $\boldsymbol{H}\in\mathbb{R}^{n_1n_2\,s\times n_1n_2n_3}$ is the measurement matrix. Also, the concept of *proximity operator* of index γ of function R is introduced as follows:

$$\text{prox}_{\gamma R}(\boldsymbol{v}):\mathbb{R}^N\to\mathbb{R}^N:\boldsymbol{v}\mapsto\arg\min_{\boldsymbol{u}} R(\boldsymbol{u})+\frac{1}{2\gamma}\|\boldsymbol{u}-\boldsymbol{v}\|_2^2. \tag{14}$$

By using duality, (13) can be rewritten as the following saddle point problem:

$$\min_{\mathcal{X}}\max_{\boldsymbol{y}}\langle\boldsymbol{H}\boldsymbol{x},\boldsymbol{y}\rangle+g(\mathcal{X})-f^*(\boldsymbol{y}) \tag{15}$$

where f^* represents the convex conjugate of function f. In addition, the proximity operator of f^* can be computed as

$$\text{prox}_{\gamma f^*}(\boldsymbol{v})=\boldsymbol{v}-\gamma\text{prox}_{\gamma^{-1}f}(\gamma^{-1}\boldsymbol{v}). \tag{16}$$

Algorithm 2: Solve (12) by PDA.

Initialize: Set $\boldsymbol{x}^0$, $\boldsymbol{y}^0$, and parameters $\tau>0$, $\sigma>0$.
while *A stopping criterion is not satisfied* **do**
  Update $\boldsymbol{y}^{k+1}$ via (20);
  Update $\mathcal{X}^{k+1}$ via (21);
  $\overline{\boldsymbol{x}}^{k+1}=\texttt{vec}(\mathcal{X}^{k+1}+(\mathcal{X}^{k+1}-\mathcal{X}^k))$.
end while

By solving dual procedure and primal procedure with fixed steps alternating, the PDA solves (13) as

$$\boldsymbol{y}^{k+1}=\text{prox}_{\sigma f^*}\left(\boldsymbol{y}^k+\sigma\boldsymbol{H}\overline{\boldsymbol{x}}^k\right) \tag{17}$$

$$\mathcal{X}^{k+1}=\text{prox}_{\tau g}\left(\texttt{ten}\left(\boldsymbol{x}^k-\tau\boldsymbol{H}^T\boldsymbol{y}^{k+1}\right)\right) \tag{18}$$

$$\overline{\boldsymbol{x}}^{k+1}=\texttt{vec}\left(\mathcal{X}^{k+1}+\left(\mathcal{X}^{k+1}-\mathcal{X}^k\right)\right) \tag{19}$$

where $\boldsymbol{y}^{k+1}$ denotes the value of $\boldsymbol{y}$ in the $(k+1)$th iteration, and σ and τ denote fixed steps. The PDA achieves convergence with assumptions $\tau,\sigma>0$ and $\tau\sigma(\sigma_1(\boldsymbol{H}))^2\le 1$. Here, $\sigma_1(\boldsymbol{H})$ is the largest singular value of $\boldsymbol{H}$. Next, we show how to solve (17) and (18).

1) Solve the subproblem $\boldsymbol{y}$ in (17): Based on (16), (17) is updated as

$$\begin{aligned}\boldsymbol{y}^{k+1}&=\text{prox}_{\sigma f^*}\left(\boldsymbol{y}^k+\sigma\boldsymbol{H}\overline{\boldsymbol{x}}^k\right)\\&=\boldsymbol{y}^k+\sigma\boldsymbol{H}\overline{\boldsymbol{x}}^k-\sigma\text{prox}_{\sigma^{-1}f}\left(\sigma^{-1}\left(\boldsymbol{y}^k+\sigma\boldsymbol{H}\overline{\boldsymbol{x}}^k\right)\right)\\&=\frac{\boldsymbol{y}^k+\sigma\boldsymbol{H}\overline{\boldsymbol{x}}^k-\sigma\boldsymbol{b}}{1+\sigma}.\end{aligned} \tag{20}$$

2) Solve the subproblem $\mathcal{X}$ in (18): Let $\mathcal{Z}^k=\texttt{ten}(\boldsymbol{x}^k-\tau\boldsymbol{H}^T\boldsymbol{y}^{k+1})$, then (18) becomes

$$\begin{aligned}\mathcal{X}^{k+1}&=\arg\min_{\mathcal{X}}\frac{1}{2}\|\mathcal{X}-\mathcal{Z}^k\|_2^2+\lambda\tau\|\mathcal{X}\|_{\text{TNN}_1}\\&=\mathcal{U}*_1\mathcal{S}_{\text{tnn}}^{\lambda\tau}*_1\mathcal{V}^{T_1}\end{aligned} \tag{21}$$

where $\mathcal{Z}^k=\mathcal{U}*_1\mathcal{S}*_1\mathcal{V}^{T_1}$, $\overline{\mathcal{S}}_1=\texttt{fft}(\mathcal{S},[\,],1)$, $(\overline{\mathcal{S}}_1)_{\text{tnn}}^{\lambda\tau}=\max(\overline{\mathcal{S}}_1-\lambda\tau,0)$, and $\mathcal{S}_{\text{tnn}}^{\lambda\tau}=\texttt{ifft}((\overline{\mathcal{S}}_1)_{\text{tnn}}^{\lambda\tau},[\,],1)$. The whole algorithm procedure is summarized in Algorithm 2.

IV. Experimental Results

Several experiments are conducted to evaluate the effectiveness of the PDA-M1TNN in this section. The PDA-M1TNN is compared with DeSCI [16], Twist with TV constraint [34] (called Twist-TV) and DP-LRTR [17] quantitatively and visually. Since DeSCI has special requirements on the sensing matrix and is mainly designed for the single snapshot CASSI in the original paper, we only compare the DeSCI on the single snapshot CASSI. The mean peak signal-to-noise ratio (MPSNR) and the mean structure similarity (MSSIM) [44] are used to evaluate the reconstruction performance. All tests are running in Intel(R) Core(TM) i9-10900 K CPU @ 3.70 GHz. For each shot, the coding aperture follows a Bernoulli random distribution consisting of $\{0, 1\}$ with equal probability. The CAVE, Cuprite, and Urban are selected for experiments. Particularly, five images

TABLE I
QUANTITATIVE RESULTS OF DIFFERENT SNAPSHOTS TIMES WITH $s = 1, 2, 4,$ AND 6 ON THE CAVE DATASET

Times	Method	MPSNR					MSSIM				
		balloons	toy	face	hairs	oil-painting	balloons	toy	face	hairs	oil-painting
1	Twist-TV	**37.32**	27.12	34.47	30.88	26.15	**0.985**	0.912	0.965	0.913	0.740
	DeSCI	36.41	19.10	32.95	23.85	26.35	0.978	0.852	0.950	0.878	0.772
	DP-LRTR	35.95	**30.20**	**37.24**	**37.31**	**31.15**	0.960	**0.917**	**0.970**	**0.962**	**0.899**
	PDA-M1TNN	24.43	28.86	29.34	29.88	26.22	0.835	0.894	0.916	0.900	0.820
2	Twist-TV	**40.64**	28.94	37.61	33.05	27.96	**0.991**	0.938	0.979	0.940	0.778
	DP-LRTR	38.48	34.77	39.42	38.50	35.05	0.978	**0.951**	**0.980**	0.968	0.941
	PDA-M1TNN	38.33	**36.80**	**40.38**	**39.90**	**35.81**	0.970	0.944	0.976	**0.980**	**0.947**
4	Twist-TV	**43.81**	31.74	40.53	35.68	29.78	**0.994**	0.961	0.986	0.957	0.820
	DP-LRTR	39.03	35.57	39.99	38.79	35.49	0.975	0.955	0.981	0.964	0.945
	PDA-M1TNN	43.38	**40.87**	**43.96**	**42.95**	**39.91**	0.987	**0.978**	**0.988**	**0.991**	**0.976**
6	Twist-TV	**45.92**	34.43	42.52	38.00	31.54	**0.995**	0.974	0.990	0.969	0.858
	DP-LRTR	39.74	36.14	42.30	40.81	37.26	0.982	0.964	0.987	0.982	0.961
	PDA-M1TNN	45.45	**43.41**	**45.78**	**44.83**	**42.41**	0.992	**0.988**	**0.992**	**0.994**	**0.985**

The best results in each case are in bold.

are selected from the CAVE dataset for illustration, including *balloons*, *toy*, *face*, *hairs,* and *oil-painting*.

Before the experiments, the value of the test dataset is converted to the interval [0, 255] in our method. The parameter for balancing data fidelity and regularization items is set to 0.1 in Twist-TV. For DeSCI, its parameter settings are consistent with those described in the original paper. For DP-LRTR, the iteration number is set to 5000, and the rank factor ρ is set separately, to obtain good results. When the number of snapshots are 1, 2, and 4, the ρ is set to 0.4. When the number of snapshots is 6, the ρ is set to 0.4. For the proposed PDA-M1TNN, the λ is set to 5000. For other parameters of our method are empirically set as $\sigma = 0.01$, $\tau = 0.1$, and maximum number of iterations $K = 1000$.

A. Results on the CAVE Dataset

CAVE dataset[1] includes 32 scenes. Each scene is of size of $512 \times 512 \times 31$. The quantitative results of MPSNR and MSSIM in different snapshot times with $s = 1, 2, 4,$ and 6 on the CAVE dataset are shown in Table I. As observed, Twist-TV obtains the best results on the *balloons* image when the snapshot times are 1, 2, 4, and 6. This is because *balloons* have good image smoothness. Except for balloon images, the DP-LRTR gets the best results both on MPSNR and MSSIM in the other four images when the number of snapshot is 1. But as the snapshot times increases, the proposed PDA-M1TNN outperforms Twist-TV and DP-LRTR. Specifically, the PDA-M1TNN obtains the best results on the *toy*, *face*, *hairs,* and *oil-painting* when the snapshot times are 2, 4, and 6. The PDA-M1TNN outperforms the DP-LRTR by 4.44 dB on MPSNR when the number of snapshot is 4 by average the results on five images. Besides, when the number of snapshots is changed from 1 to 2, the MPSNR in our PDA-M1TNN increases by 10.50 dB by average the results on five images. As the snapshots times increases, the reconstruction performance is further improved. These illustrate the advantage of the proposed method in multishot CASSI.

The visual results of toy image by different methods are shown in Fig. 5. As observed from Fig. 5(a), the recovery result of Twist-TV exhibits excessive smoothness, with many details lost. The recovery result of DP-LRTR exhibits spectral distortion as shown in the zoom in result of Fig. 5(b). Apparently, the reconstruction result of the PDA-M1TNN has a better visual effect. Fig. 6 shows a comparison of the reconstruction of the different methods in spectral dimensions along different spatial pixel positions, which is represented by the center of the circle in Fig. 6(a). The reconstruction results are recovered from four snapshots compressive measurements and the selected three spatial pixels represent three different substances. The spectral signatures of the three different spatial pixel locations are, respectively, shown in Fig. 6(b)–(d). In addition, the correlation coefficients between the results of reconstructed methods and original data are also included. As observed, the PDA-M1TNN recovers the spectrum with high fidelity. These further demonstrate the effectiveness of the PDA-M1TNN.

B. Results on the Cuperite Dataset

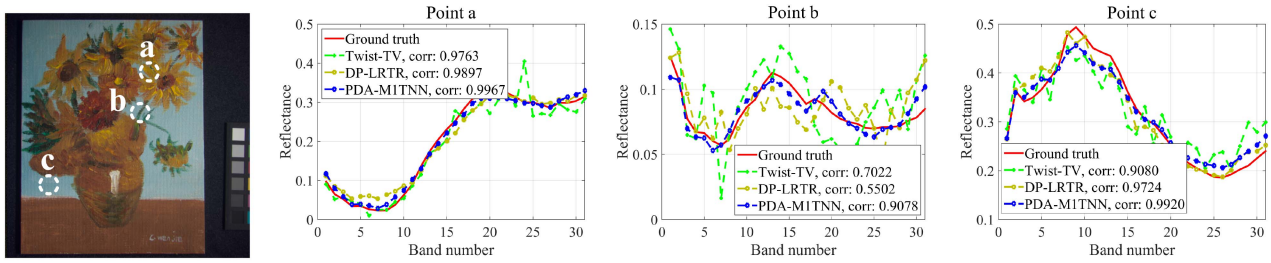
The original data of Cuperite have 224 bands, which range from 370 to 2480 nm. The size of Cuperite in the experiments is $250 \times 191 \times 188$ by removing the contaminated bands.

The quantitative results of the competing methods at different snapshot times with $s = 1, 2, 4,$ and 6 on the Cuperite dataset are shown in Table II. As observed, the PDA-M1TNN outperforms Twist-TV and DP-LRTR in all evaluation metrics when the snapshot times are 2, 4, and 6. DP-LRTR gets the best results when

[1][Online]. Available: https://cave.cs.columbia.edu/repository/Multispectral

(a) (b) (c) (d)

Fig. 5. Reconstruction results in the toy. The red, green, and blue in color image are, respectively, composed of bands 31, 14, 5. (a) Twist-TV. (b) DP-LRTR. (c) PDA-M1TNN. (d) Original.



(a) (b) (c) (d)

Fig. 6. Spectral signatures comparisons of different methods. (a) Three spatial pixel positions. (b) Point a ($x = 302, y = 127$). (c) Point b ($x = 282, y = 208$). (d) Point c ($x = 88, y = 362$).

TABLE II
QUANTITATIVE RESULTS OF DIFFERENT SNAPSHOTS TIMES WITH $s = 1, 2, 4,$ AND 6 ON THE CUPERITE DATASET

Index	Method	s			
		1	2	4	6
MPSNR	Twist-TV	22.47	24.24	26.11	25.35
	DeSCI	21.47	/	/	/
	DP-LRTR	**30.43**	32.94	36.26	36.49
	PDA-M1TNN	24.32	**34.68**	**38.35**	**40.07**
MSSIM	Twist-TV	0.727	0.738	0.767	0.764
	DeSCI	0.677	/	/	/
	DP-LRTR	**0.920**	0.925	0.945	0.951
	PDA-M1TNN	0.885	**0.926**	**0.953**	**0.964**

The best results in each case are in bold.

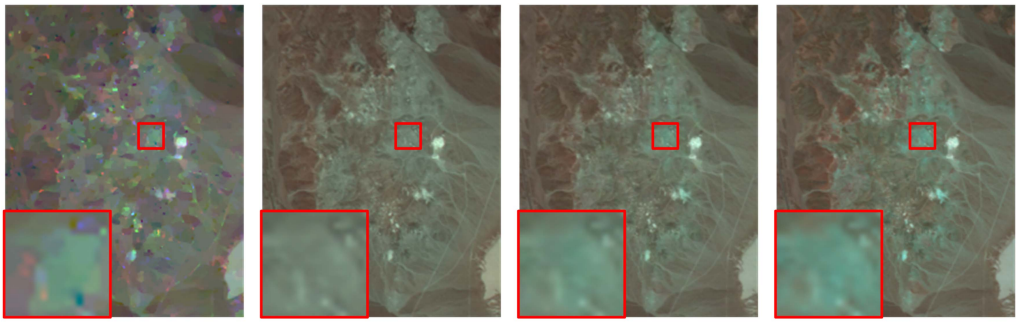
the number of snapshots is 1. Specifically, when the number of snapshots is 2, PDA-M1TNN is 1.74 dB higher than DP-LRTR in MPSNR and 10.44 dB higher than Twist-TV in MPSNR. Furthermore, when the number of snapshots changes from 1 to 2, the MPSNR in our PDA-M1TNN increases by 10.36 dB. This is a rapid increase, and as the snapshot time increases, the reconstruction performance is further improved.

The visual results by different methods on Cuperite are shown in Fig. 7. As observed from Fig. 7, the recovery results of Twist-TV show excessive smoothness, while DP-LRTR shows spectral distortion. The reconstruction result of the PDA-M1TNN has better visual effects. Fig. 8 shows a comparison of the reconstruction of different methods in spectral dimensions along different spatial pixel positions represented by the center of the circle in Fig. 8(a). The reconstruction results are recovered from four snapshots and the selected three spatial pixels represent three different substances. The spectral signatures of the three different spatial pixel locations are, respectively, shown in Fig. 8(b)–(d). In addition, the correlation coefficients are also included. As observed, the PDA-M1TNN recovers the spectral information with high fidelity. Furthermore, the correlation coefficients of the PDA-M1TNN are as high as 0.99. These further demonstrate the effectiveness of the PDA-M1TNN.

C. Experimental Results on the Urban Dataset

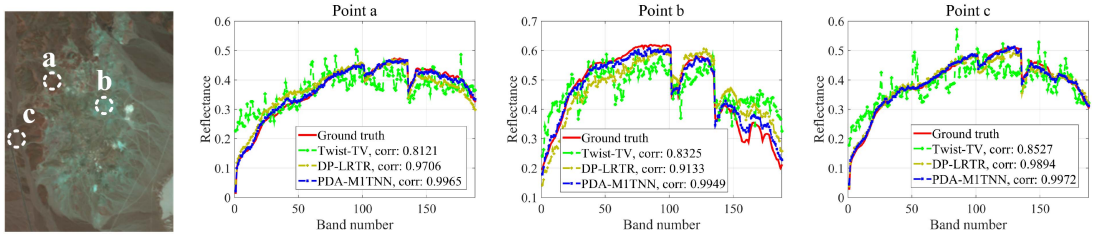
This dataset is obtained by the HYDICE sensor. The spatial size of the Urban dataset is 307×307. The original data have 210 bands. 162 bands are left after removing noise and water absorption bands, and only a subimage of size $256 \times 256 \times 162$ is used for the experiments.

The quantitative results of the competing methods at different snapshot times with $s = 1$, 2, 4, and 6 on the Urban are listed in Table III. As observed, the PDA-M1TNN outperforms the Twist-TV in all evaluation metrics in all the cases. When the number of snapshot is 1, the DP-LRTR achieves the best results. When the snapshot times are 2 and 4, the PDA-M1TNN obtains the best results. As the snapshot times increases, the reconstruction performance is further improved. Besides, when the number of snapshots is changed from 1 to 2, the MPSNR in



(a) (b) (c) (d)

Fig. 7. Reconstruction results on the Cuperite. The red, green, and blue in color image are, respectively, composed of bands 140, 56, 31. (a) Twist-TV. (b) DP-LRTR. (c) PDA-M1TNN. (d) Original.



(a) (b) (c) (d)

Fig. 8. Spectral signatures comparisons of different methods. (a) Three spatial pixel positions. (b) Point a ($x = 54, y = 80$). (c) Point b ($x = 117, y = 106$). (d) Point c ($x = 15, y = 149$).

TABLE III
QUANTITATIVE RESULTS OF DIFFERENT SNAPSHOTS TIMES WITH $s = 1, 2, 4$, AND 6 ON THE URBAN DATASET

Index	Method	s			
		1	2	4	6
MPSNR	Twist-TV	20.45	21.19	22.34	23.48
	DeSCI	17.53	/	/	/
	DP-LRTR	**23.93**	26.11	**31.19**	33.13
	PDA-M1TNN	22.49	**27.08**	31.03	**33.73**
MSSIM	Twist-TV	0.530	0.563	0.655	0.682
	DeSCI	0.643	/	/	/
	DP-LRTR	**0.758**	**0.813**	**0.914**	0.940
	PDA-M1TNN	0.743	0.803	0.902	**0.943**

The best results in each case are in bold.

our PDA-M1TNN increases by 4.59 dB, which also illustrates the advantage of multishot CASSI over a single snapshot CASSI.

Fig. 9 shows the reconstruction results for the pseudocolor image, which consisted of bands 150, 90, and 30. As can be observed, Twist-TV produces a serious over-smoothing phenomenon and loses a lot of detail information, while the proposed PDA-M1TNN can reconstruct the detail information well. Fig. 10 shows a comparison of the reconstruction of different methods in spectral dimensions along three different spatial pixel represented. The results are recovered from four snapshots compressive measurements and the selected three spatial pixels represent different classes. The spectral signatures of the three different spatial pixel are presented in Fig. 10(b)–(d). In addition, the correlation coefficients between the results of reconstructed methods and original data are also included. As observed, the PDA-M1TNN recovers the spectral information with high fidelity, which verifies the effectiveness of the proposed method.

D. Computational Efficiency

The time complexity of the proposed PDA-M1TNN is the same as that of the mode-1 TNN, whose costs at each iteration is $O(n_1n_2n_3\log n_1 + n_1n_2n_3\,N)$, where $N = \min(n_2, n_3)$. Hence, the time complexity of the proposed PDA-M1TNN at each iteration is $O(n_1n_2n_3\log n_1 + n_1n_2n_3\,N)$.

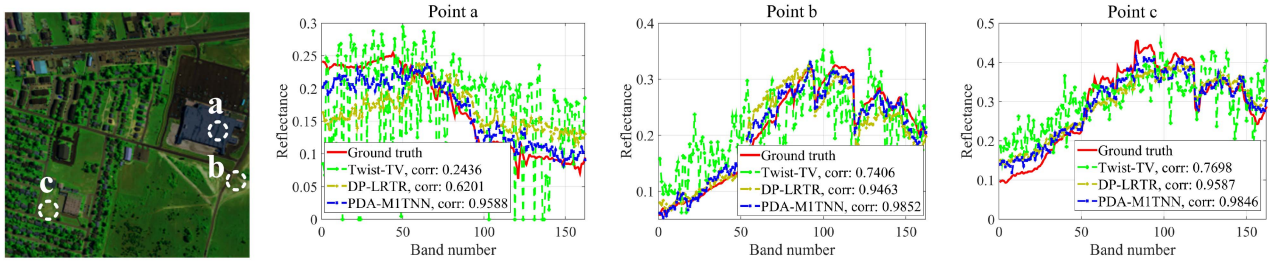
Table IV shows the CPU running time (in hours) of the methods on CAVE, Cuperite, and Urban with different snapshot times. For the running time on the CAVE dataset, it is the average time of the five images selected on this dataset. From a comprehensive view of Table IV, the proposed PDA-M1TNN runs faster than DeSCI and DP-LRTR in each case. Furthermore, compared with Twist-TV, the proposed PDA-M1TNN also has competitive performance. These also verify the effectiveness of the PDA-M1TNN in multishot CASSI reconstruction.

E. Discussion

1) Parameter discussion of λ: For the proposed method, the parameter λ is discussed by varying the value of λ in the range of {1000, 3000, 5000, 7000, 9000, 11 000}. When the number of snapshots is 4 in Cuperite dataset, the quantitative results of the PDA-M1TNN in different values of λ are shown in Table V.

(a) (b) (c) (d)

Fig. 9. Reconstruction results on the Urban. The red, green, and blue in color image are, respectively, composed of bands 150, 90, 30. (a) Twist-TV. (b) DP-LRTR. (c) PDA-M1TNN. (d) Original.



(a) (b) (c) (d)

Fig. 10. Spectral signatures comparisons of different methods. (a) Three spatial pixel positions. (b) Point a ($x = 223, y = 123$). (c) Point b ($x = 244, y = 179$). (d) Point c ($x = 46, y = 209$).

TABLE IV
RUNNING TIME OF THE METHODS ON THE DIFFERENT DATASETS IN DIFFERENT SNAPSHOTS TIMES WITH $s = 1, 2, 4,$ AND 6

Times	Method	CAVE	Cuperite	Urban
1	Twist-TV	0.15	0.20	0.21
	DeSCI	1.39	1.13	1.42
	DP-LRTR	3.02	6.35	5.35
	PDA-M1TNN	0.10	0.09	0.09
2	Twist-TV	0.13	0.18	0.20
	DP-LRTR	2.85	6.81	5.35
	PDA-M1TNN	0.13	0.16	0.21
4	Twist-TV	0.10	0.20	0.32
	DP-LRTR	3.13	6.49	5.72
	PDA-M1TNN	0.15	0.21	0.19
6	Twist-TV	0.16	0.26	0.27
	DP-LRTR	3.35	6.60	5.86
	PDA-M1TNN	0.17	0.19	0.19

TABLE V
QUANTITATIVE RESULTS WITH DIFFERENT VALUES OF λ WHEN THE NUMBER OF SNAPSHOTS IS 4

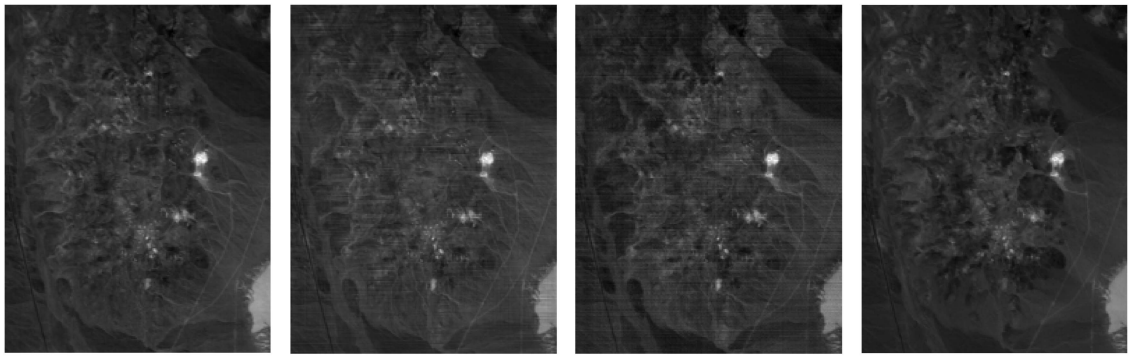
Index	1000	3000	5000	7000	9000	11000
MPSNR	34.16	38.28	38.35	37.68	37.12	36.61
MSSIM	0.924	0.951	0.953	0.949	0.946	0.944

As observed, the quantitative results reach their best when the λ is 5000. Hence, the λ is set to 5000 in the experiments.

2) Discussion of mode-k TNN: It is assumed that the coded scene is spectrally horizontally shifted in this article and the mode-1 TFR is applied to utilize the spatial-spectral correlation of HSI in multishot CASSI reconstruction. In order to illustrate the superiority of mode-1 TFR, we compare it with the mode-2 TFR and mode-3 TFR. Correspondingly, the mode-2 TNN and the mode-3 TNN are, respectively, used as the convex relaxations of mode-2 TRF and mode-3 TRF and the corresponding methods are named as PDA-M2TNN and PDA-TNN. The experiments are conducted on the Cuperite dataset. Fig. 11 shows the reconstruction results of Cuperite by different mode-k TNN by recovering from four snapshots compressive measurements. As observed, there are still a lot of horizontal stripes in the result of PDA-M2TNN. In the result of PDA-TNN, there is not only horizontal stripes, but also other noise, while the proposed PDA-M1TNN has better visual effects. Table VI shows the quantitative results on MPSNR and MSSIM. As observed, the PDA-M1TNN achieves the best performance in all the cases. These further demonstrate that the mode-2 t-SVD and mode-3 t-SVD characterization result in suboptimal CASSI reconstruction performance, while the mode-1 t-SVD can obtain a more accurate HSI structure characterization in multishot CASSI reconstruction.

3) Discussion of mask shifting in a vertical direction: If the mask shifts in a vertical direction, the mode-2 t-SVD performing the mode-2 TFR minimization should be used for better multishot CASSI performance.

First, the performance of mode-k ($k = 1,2\ 3$) TNN is compared when the mask shifts in a vertical direction in multishot



(a) (b) (c) (d)

Fig. 11. Reconstruction results of Cuperite by different mode-k TNN. (a) PDA-M1TNN. (b) PDA-M2TNN. (c) PDA-TNN. (d) Original.

TABLE VI
QUANTITATIVE RESULTS OF DIFFERENT MODE-k TNN ON MPSNR AND MSSIM WITH DIFFERENT SNAPSHOT TIMES ON THE CUPERITE DATASET WHEN THE MASK SHIFTS IN A HORIZONTAL DIRECTION

Index	Method	s			
		1	2	4	6
MPSNR	PDA-M1TNN	**24.32**	**34.68**	**38.35**	**40.07**
	PDA-M2TNN	24.14	30.17	36.22	38.36
	PDA-TNN	24.15	24.58	29.15	32.95
MSSIM	PDA-M1TNN	**0.885**	**0.926**	**0.953**	**0.964**
	PDA-M2TNN	0.883	0.912	0.943	0.957
	PDA-TNN	0.884	0.820	0.834	0.892

The best results in each case are in bold.

TABLE VII
QUANTITATIVE RESULTS OF DIFFERENT MODE-k TNN ON MPSNR AND MSSIM WITH DIFFERENT SNAPSHOT TIMES ON THE CUPERITE DATASET WHEN THE MASK SHIFTS IN A VERTICAL DIRECTION

Index	Method	s			
		1	2	4	6
MPSNR	PDA-M1TNN	24.21	30.64	35.94	38.07
	PDA-M2TNN	**24.34**	**35.12**	**38.81**	**40.85**
	PDA-TNN	24.15	26.19	30.72	33.23
MSSIM	PDA-M1TNN	0.882	0.917	0.939	0.953
	PDA-M2TNN	**0.885**	**0.932**	**0.959**	**0.969**
	PDA-TNN	0.884	0.811	0.863	0.898

The best results in each case are in bold.

CASSI. Table VII shows the quantitative results on MPSNR and MSSIM on the Cuperite dataset when the mask shifts in a vertical direction. As observed, the PDA-M2TNN achieves the best performance in all the cases. These demonstrate that the mode-2 TFR minimization obtains the optimal results when the mask shifts in a vertical direction. Second, the quantitative results of the competing methods at different snapshot times with $s = 1, 2,$ 4, and 6 on the Cuperite dataset are shown in Table VIII. As observed, the PDA-M2TNN outperforms Twist-TV and DP-LRTR when the snapshot times are 2, 4, and 6. DP-LRTR gets the best results when the number of snapshots is 1. Specifically, when the number of snapshots is 2, PDA-M2TNN is 2.34 dB higher than DP-LRTR in MPSNR and 11.02 dB higher than Twist-TV in MPSNR. These further demonstrate the effectiveness of the PDA-M2TNN when the mask shifts in a vertical direction in multishot CASSI.

TABLE VIII
QUANTITATIVE RESULTS OF DIFFERENT SNAPSHOTS TIMES WITH $s = 1, 2, 4,$ AND 6 ON THE CUPERITE DATASET WHEN THE MASK SHIFTS IN A VERTICAL DIRECTION

Index	Method	s			
		1	2	4	6
MPSNR	Twist-TV	22.60	24.10	25.27	24.80
	DeSCI	21.54	/	/	/
	DP-LRTR	**31.15**	32.78	33.26	35.69
	PDA-M2TNN	24.34	**35.12**	**38.81**	**40.85**
MSSIM	Twist-TV	0.731	0.737	0.762	0.753
	DeSCI	0.679	/	/	/
	DP-LRTR	**0.924**	**0.933**	0.941	0.945
	PDA-M2TNN	0.885	0.932	**0.959**	**0.969**

The best results in each case are in bold.

V. CONCLUSION

In this article, we consider the TFR minimization problem for multishot CASSI and solve it using the PDA. The TFR is generalized from the mode-k t-SVD. Since the mask is shifted along the horizontal direction, we use mode-1 t-SVD to exploit the joint spatial-spectral correlation. The TFR minimization is reduced to a mode-1 TNN minimization problem to achieve a more accurate HSI characterization in multishot CASSI reconstruction. The proposed method is solved by the PDA. Note that if the mask shifts in a vertical direction, the mode-2 t-SVD performing the mode-2 TFR minimization should be used for better

multishot CASSI performance. The proposed PDA-M1TNN is mainly aimed at DD-CASSI. Therefore, the reconstruction of SD-CASSI is worth studying. In addition, the mode-1 TNN in this article can be replaced with better regularization/prior terms or powerful representation learning models [45], [46] for better CASSI performance.

REFERENCES

[1] L. Gao, D. Wang, L. Zhuang, X. Sun, M. Huang, and A. Plaza, "BS ^{3} LNet: A. new blind-spot self-supervised learning network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Feb. 2023, Art. no. 5504218.

[2] W. Rao, Y. Qu, L. Gao, X. Sun, Y. Wu, and B. Zhang, "Transferable network with siamese architecture for anomaly detection in hyperspectral images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 106, Feb. 2022, Art. no. 102669.

[3] W. Rao, L. Gao, Y. Qu, X. Sun, B. Zhang, and J. Chanussot, "Siamese transformer network for hyperspectral image target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2022, Art. no. 5526419.

[4] L. Gao, X. Sun, X. Sun, L. Zhuang, Q. Du, and B. Zhang, "Hyperspectral anomaly detection based on chessboard topology," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Feb. 2023, Art. no. 5505016.

[5] X. Sun, L. Zhuang, L. Gao, H. Gao, X. Sun, and B. Zhang, "Information retrieval with chessboard-shaped topology for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Jun. 2023, Art. no. 5514515.

[6] L. Zhuang, L. Gao, B. Zhang, X. Fu, and J. M. Bioucas-Dias, "Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 2020, Art. no. 5500117.

[7] C.-I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*. Hoboken, NJ, USA: Wiley, 2007.

[8] Q. Tong, Y. Xue, and L. Zhang, "Progress in hyperspectral remote sensing science and technology in China over the past three decades," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 70–91, Jan. 2014.

[9] G. Lu and B. Fei, "Medical hyperspectral imaging: A review," *J. Biomed. Opt.*, vol. 19, no. 1, 2014, Art. no. 0 10901.

[10] A. Wagadarikar, R. John, R. Willett, and D. J. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Appl. Opt.*, vol. 47, no. 10, pp. B44–B51, Feb. 2008.

[11] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Exp.*, vol. 15, no. 21, pp. 14013–14027, Oct. 2007.

[12] X. Lin, Y. Liu, J. Wu, and Q. Dai, "Spatial-spectral encoded compressive hyperspectral imaging," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 233:1–233:11, Nov. 2014.

[13] M. Zhang, L. Wang, and H. Huang, "Sub-pixel dispersion model for coded aperture snapshot spectral imaging," *IEEE Trans. Comput. Imag.*, vol. 9, pp. 1188–1200, Dec. 2023.

[14] T. Zhang, S. Zhao, X. Ma, A. Ramirez-Jaime, Q. Zhao, and G. R. Arce, "Compressive spectral imaging via misalignment induced equivalent grayscale coded aperture," *IEEE Geosci. Remote. Sens. Lett.*, vol. 20, Feb. 2023, Art. no. 5503005.

[15] W. He, N. Yokoya, and X. Yuan, "Fast hyperspectral image recovery of dual-camera compressive hyperspectral imaging via non-iterative subspace-based fusion," *IEEE Trans. Image Process.*, vol. 30, pp. 7170–7183, Aug. 2021.

[16] Y. Liu, X. Yuan, J. Suo, D. J. Brady, and Q. Dai, "Rank minimization for snapshot compressive imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 12, pp. 2990–3006, Dec. 2019.

[17] J. Bacca, Y. Fonseca, and H. Arguello, "Compressive spectral image reconstruction using deep prior and low-rank tensor representation," *Appl. Opt.*, vol. 60, no. 14, pp. 4197–4207, May 2021.

[18] L. Galvis, E. Mojica, H. Arguello, and G. R. Arce, "Shifting colored coded aperture design for spectral imaging," *Appl. Opt.*, vol. 58, no. 7, pp. B28–B38, 2019.

[19] H. Arguello and G. R. Arce, "Rank minimization code aperture design for spectrally selective compressive imaging," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 941–954, Mar. 2013.

[20] H. Arguello, H. Rueda, Y. Wu, D. W. Prather, and G. R. Arce, "Higher-order computational model for coded aperture spectral imaging," *Appl. Opt.*, vol. 52, no. 10, pp. D12–D21, 2013.

[21] D. Kittle, K. Choi, A. Wagadarikar, and D. J. Brady, "Multiframe image estimation for coded aperture snapshot spectral imagers," *Appl. Opt.*, vol. 49, no. 36, pp. 6824–6833, 2010.

[22] L. V. G. Carreño and H. A. Fuentes, "Optical codification design in compressive spectral imaging: From mathematical to deep learning optimization," *Opt. Pura Appl.*, vol. 55, no. 1, pp. 1–15, 2022.

[23] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging: An introduction," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 105–115, Jan. 2014.

[24] Y. Wu, I. O. Mirza, G. R. Arce, and D. W. Prather, "Development of a digital-micromirror-device-based multishot snapshot spectral imaging system," *Opt. Lett.*, vol. 36, no. 14, pp. 2692–2694, Jul. 2011.

[25] D. Reddy, A. Veeraraghavan, and R. Chellappa, "P2C2: Programmable pixel compressive camera for high speed imaging," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2011, pp. 329–336.

[26] P. Llull et al., "Coded aperture compressive temporal imaging," *Opt. Exp.*, vol. 21, no. 9, pp. 10526–10545, 2013.

[27] L. Wang, Z. Xiong, G. Shi, F. Wu, and W. Zeng, "Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 10, pp. 2104–2111, Oct. 2017.

[28] Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, and S. K. Nayar, "Video from a single coded exposure photograph using a learned over-complete dictionary," in *Proc. IEEE Int. Conf. Comput. Vis*., 2011, pp. 287–294.

[29] J. Yang et al., "Video compressive sensing using gaussian mixture models," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4863–4878, Nov. 2014.

[30] J. Yang et al., "Compressive sensing by learning a gaussian mixture model from measurements," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 106–119, Jan. 2015.

[31] L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu, "Dual-camera design for coded aperture snapshot spectral imaging," *Appl. Opt.*, vol. 54, no. 4, pp. 848–858, 2015.

[32] X. Yuan, "Generalized alternating projection based total variation minimization for compressive sensing," in *Proc. IEEE Int. Conf. Image Process*., 2016, pp. 2539–2543.

[33] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[34] J. M. Bioucas-Dias and M. A. Figueiredo, "A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2992–3004, Dec. 2007.

[35] H. Li, X.-L. Zhao, J. Lin, and Y. Chen, "Low-rank tensor optimization with nonlocal plug-and-play regularizers for snapshot compressive imaging," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 581–593, 2021.

[36] L. Wang, T. Zhang, Y. Fu, and H. Huang, "HyperReconNet: Joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2257–2270, May 2019.

[37] G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica*, vol. 6, no. 8, pp. 921–943, Aug. 2019.

[38] Y. Fu, T. Zhang, L. Wang, and H. Huang, "Coded hyperspectral image reconstruction using deep external and internal learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3404–3420, Jul. 2022.

[39] Z. Meng, J. Ma, and X. Yuan, "End-to-end low cost compressive spectral imaging with spatial-spectral self-attention," in *Proc. IEEE Eur. Conf. Comput. Vis*., 2020, pp. 187–204.

[40] H. Fan, C. Li, Y. Guo, G. Kuang, and J. Ma, "Spatial-spectral total variation regularized low-rank tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6196–6213, Oct. 2018.

[41] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, T.-H. Ma, and T.-Y. Ji, "Mixed noise removal in hyperspectral image via low-fibered-rank regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 734–749, Jan. 2020.

[42] C. Lu, J. Feng, Y. Chen, L. Wei, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2016, pp. 5249–5257.

[43] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging. Vis.*, vol. 40, no. 1, pp. 120–145, 2011.

[44] T. Xie, S. Li, and J. Lai, "Adaptive rank and structured sparsity corrections for hyperspectral image restoration," *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 8729–8740, Sep. 2022.
[45] L. Liu, L. Cai, T. Xie, and Y. Wang, "Self-paced broad learning system," *IEEE Trans. Cybern.*, vol. 53, no. 6, pp. 4029–4042, Jun. 2023.
[46] T. Xie, L. Liu, and L. Zhuang, "Plug-and-play priors for multi-shot compressive hyperspectral imaging," *IEEE Trans. Image Process.*, vol. 32, pp. 5326–5339, 2023.

Ting Xie (Member, IEEE) received the Ph.D. degree in control science and engineering from the College of Electrical and Information Engineering, Hunan University, Changsha, China, in 2020.

From 2018 to 2019, she was a visiting Ph.D. student with the University of Lisbon, Lisbon, Portugal, supported by the China Scholarship Council. She is currently an Associate Professor with the College of Engineering and Design, Hunan Normal University. Her research interests include machine learning, image quality improvement, and remote sensing image processing.

Xudong Kang (Senior Member, IEEE) received the B.Sc. degree in automation from Northeast University, Shenyang, China, in 2007, and the Ph.D. degree in control science and engineering from Hunan University, Changsha, China, in 2015.

In 2015, he joined the College of Electrical Engineering, Hunan University. His research interests include hyperspectral feature extraction, image classification, image fusion, and anomaly detection.

Dr. Kang received the National Nature Science Award of China (Second class and ranked third) and the Second Prize in the Student Paper Competition in IGARSS 2014. He was selected as the Best Reviewer for *IEEE Geoscience and Remote Sensing Letters* and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was an Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2018 to 2019. He serves as an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and *IEEE Journal on Miniaturization for Air and Space Systems*.

Renwei Dian (Member, IEEE) received the B.S. degree in automation from the Wuhan University of Science and Technology, Wuhan, China, in 2015, and the Ph.D. degree in control science and engineering from Hunan University, Changsha, China, in 2020.

He is currently an Associate Professor with the School of Robotics, Hunan University. From 2017 to 2018, he was a visiting Ph.D. student with the University of Lisbon, Lisbon, Portugal, supported by the China Scholarship Council. From 2020 to 2022, he was a Postdoctor with the College of Electrical and Information Engineering, Hunan University, where he was awarded the distinguished postdoctoral fellow. He was a finalist for the Best Student Paper Award at the International Geoscience and Remote Sensing Symposium (IGARSS) 2018. He was awarded the Wu Wenjun Award for Outstanding Youth in Artificial Intelligence, the Fellowship of China National Postdoctoral Program for Innovative Talents, and an Excellent Doctoral Dissertation by the China Society of Image and Graphics. His research interests include hyperspectral image superresolution, image fusion, tensor decomposition, and deep learning. More information can be found on his homepage https://sites.google.com/view/renweidian/.

Tonghan Wang received the Ph.D. degree in computer science from the School of Computer Science and Engineering, Southeast University, Nanjing, China, in 2016.

He is currently an Associate Professor with the School of Information Engineering, East China University of Technology, Fuzhou, China. His main research interests include image processing, pattern recognition, machine learning, and image quality assessment.

Licheng Liu (Member, IEEE) received the B.S. degree in information and computational science from the China University of Geosciences, Wuhan, China, in 2010, the M.S. degree in applied mathematics from Hunan University, Changsha, China, in 2012, and the Ph.D. degree in software engineering from the University of Macau, Macau, China, in 2016.

He is currently an Associate Professor with the College of Electrical and Information Engineering, Hunan University. His research interests include image processing, computer vision, pattern recognition, and machine learning.