Computational Hyperspectral Imaging Based on Dimension-discriminative Low-rank Tensor Recovery

Shipeng Zhang[1] Lizhi Wang[2*] Ying Fu[2] Xiaoming Zhong[3] Hua Huang[2]
[1]Xi'an Jiaotong University [2]Beijing Institute of Technology
[3]Beijing Institute of Space Mechanics and Electricity

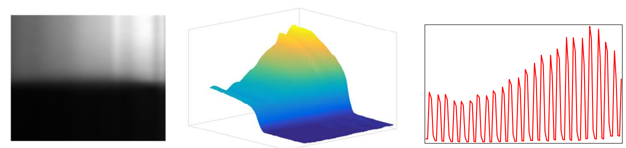
Abstract

Exploiting the prior information is fundamental for the image reconstruction in computational hyperspectral imaging. Existing methods usually unfold the 3D signal as a 1D vector and treat the prior information within different dimensions in an indiscriminative manner, which ignores the high-dimensionality nature of hyperspectral image (HSI) and thus results in poor quality reconstruction. In this paper, we propose to make full use of the high-dimensionality structure of the desired HSI to boost the reconstruction quality. We first build a high-order tensor by exploiting the nonlocal similarity in HSI. Then, we propose a dimension-discriminative low-rank tensor recovery (DLTR) model to characterize the structure prior adaptively in each dimension. By integrating the structure prior in DLTR with the system imaging process, we develop an optimization framework for HSI reconstruction, which is finally solved via the alternating minimization algorithm. Extensive experiments implemented with both synthetic and real data demonstrate that our method outperforms state-of-the-art methods.

1. Introduction

Hyperspectral imaging techniques capture reflectance of a real scene across tens to hundreds of discrete bands. Compared with the traditional RGB image, the hyperspectral image (HSI) can provide more details and features with the additional spectral dimension. Such characteristics have been applied to various computer vision tasks, such as classification [1], recognition [2], and tracking [3].

HSIs are 3D in nature, including one spectral dimension and two spatial dimensions. To obtain a HSI, the conventional spectrometers [4, 5, 6, 7], which are equipped with 1D or 2D detectors, need to scan the scene along the spatial or spectral dimensions. Thus, the scanning-based hyperspectral imaging methods suffer from low temporal resolution and cannot be used to capture dynamic



(a) (b) (c)

Figure 1: Illustration of the spectral-spatial correlation in a local patch. (a) A local spectral-spatial patch of size 31×36. (b) The 3D smooth surface of (a). (c) The vectoriezd 1D signal of (a). It shows that the vectorization process breaks the intrinsic structure information.

scenes. Thanks to the flourish of computational photography, computational hyperspectral imaging has been developed to overcome this problem [8, 9, 10]. Among numerous imaging systems, coded aperture snapshot spectral imaging (CASSI) [11, 12] and its dual camera design (DCD) [13, 14] have attracted increasing attention in recent years due to the advantage of snapshot. With elaborate optical designs, they encode the 3D HSI into the 2D compressive measurement. Consequently, the bottleneck in computational hyperspectral imaging is how to faithfully reconstruct the 3D HSI from the compressive measurement.

Since the reconstruction is severely under-determined, image prior information must be characterized to regularize the reconstruction. So far considerable research efforts in this field have been devoted to exploiting the prior information [14, 15, 16, 17, 18, 19, 20, 21]. However, existing methods usually unfold the 3D signal as a 1D vector and treat the prior information within different dimensions in an indiscriminative manner. The vectorization process ignores the high-dimensionality nature of HSI and breaks the original structure information in HSI. As an intuitive example, Fig. 1 shows a comparison of pixel value distributions of one spatial-spectral patch with and without vectorization. It is clear that the structure information gets lost after vectorization.

Our key observation is that compared with the 1D-vector based signal description, high-order tensors can provide a

*Corresponding author: lzwang@bit.edu.cn

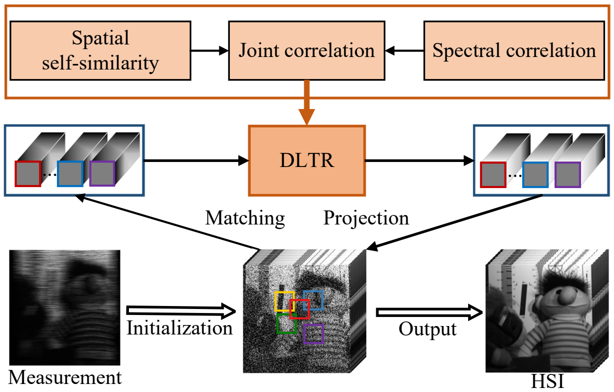


Figure 2: Overview of the proposed method. We first obtain an initialization from the compressive measurement. Then our reconstruction method, including matching, dimension-discriminative low-rank tensor recovery (shortened as DLTR in the figure) and projection, is iteratively performed. Finally, we output the reconstructed HSI.

more accurate representation to figure out the data diversity in each domain and deliver the intrinsic structure of high-dimensionality signals. Such observation motivates us to exploit the tensor representation for HSI reconstruction to address the problem of vectorization.

In this paper, we propose to make full use of the high-dimensionality structure prior in HSI to boost the reconstruction quality for computational hyperspectral imaging. The overall framework of our method is illustrated in Fig. 2. Specifically, we first build a high-order tensor by exploiting the nonlocal similarity in the spatial and spectral domains for each exemplar cubic patch. Then, to characterize the structure prior in each dimension, we propose a dimension-discriminative low-rank tensor recovery (DLTR) model, where the spatial self-similarity, spectral correlation and joint correlation will be fully exploited. By integrating the tensor-based structure prior in DLTR with the system imaging process, we develop an optimization framework for HSI reconstruction, which is finally solved via an alternating minimization algorithm. To the best of our knowledge, it is the first time to characterize the prior in HSI with a dimension-discriminative model for computational hyperspectral imaging. Extensive experiments implemented with both synthetic data and real capture data demonstrate that our method outperforms state-of-the-art methods.

2. Related Works

2.1. Hyperspectral Image Reconstruction

Recovering the 3D HSI from the 2D compressive measurement plays an essential role in computational hyperspectral imaging. So far methods for HSI reconstruction can be generally grouped into two categories: prior knowledge based methods and deep learning based methods.

Traditional methods reconstruct HSI by solving optimization problems with prior knowledge based regularization. Under the sparsity assumption, HSI can be reconstructed by solving an ℓ_0 or an ℓ_1 relaxation optimization problem. With the hypothesis that objects in the scene have piecewise smooth structure, total variation (TV) regularization based methods have been widely used in computational hyperspectral imaging [15, 22, 23]. Various sparse reconstruction methods have been developed [12, 14, 16, 17, 21, 24] by conducting wavelet transform or over-completed dictionary as sparsity basis. Minimization of matrix rank based methods were proposed in [18, 20] to exploit the spatial-spectral correlation. However, existing methods always unfold the 3D signal as a 1D vector and treat the prior information within different dimensions in an indiscriminative way, which ignores the high-dimensionality nature of HSI and results in poor reconstruction quality.

In recent years, deep learning techniques have been employed for image restoration. Several methods on deep neural network attempted to learn prior knowledge for natural image compressive imaging [25, 26]. However, the heterogeneity of HSI makes those methods difficult to be extended for computational hyperspectral imaging. A convolutional autoencoder (AE) was proposed in [27] to learn a non-linear sparse representation for CASSI. A recent method named HSCNN used a convolutional network to finish the reconstruction task [28]. But these methods all try to learn a single prior information and can not exploit the high-dimensionality intrinsic nature of HSI. Moreover, few experiments are carried out to verify that these networks can be applied to the actual hardware systems.

Based on our observation that high-dimensional tensors are more faithful to deliver intrinsic properties, we propose a DLTR based method to further promote the reconstruction fidelity. Besides, extensive experiments on actual hardware implementations validate the generalization of our method.

2.2. Low-rank Tensor Recovery

Low-rank tensor recovery has been widely used in various computer vision tasks[29, 30, 31, 32, 33] in recent years. Based on the high-order singular value decomposition (HOSVD) [34], low-rank tensor recovery can be regarded as a factorization based problem. However, HOSVD needs to predefine the rank of each mode, which is a hard and unstable task. Sum of ranks minimization based methods [29, 35] have been widely used for tensor completion. However, these methods treat the rank of all modes equally, which ignore the discrepant physical meaning of different modes. To figure out this issue, a weighted tensor nuclear norm regulation based model [36] was presented and its relaxation with non-convex forms [37, 38] were further developed.

In this paper, we take advantage of the trend of this domain to boost the reconstruction quality for computational hyperspectral imaging. To the best of our knowledge, it is the first time that the DLTR model has been utilized for computational hyperspectral imaging.

3. Dimension-differentiated Low-rank Tensor Model

3.1. Notations and Preliminaries

We first introduce notations and preliminaries as follows. Matrices are denoted as boldface capital letters, e.g., $\boldsymbol{X}$, vectors are represented with blodface lowercase letters, e.g. $\boldsymbol{x}$, and scalars are indicated as lowercase letters, e.g., x. We denote a tensor of order N as boldface Euler script $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_N}$ and its *Frobenius* norm is the square root of the sum of the squares of all its elements, i.e., $\|\boldsymbol{\mathcal{X}}\|_F = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x(i_1, i_2, \cdots, i_N)^2}$. By varying index i_n while keeping the others fixed, the mode-n fiber of $\boldsymbol{\mathcal{X}}$ can be obtained. By arranging the mode-n fibers of $\boldsymbol{\mathcal{X}}$ as column vectors, we get the mode-n unfolding matrix of $\boldsymbol{\mathcal{X}}$, which is denoted as $\boldsymbol{X}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \cdots I_{n-1} I_{n+1} \cdots I_N)}$. $\text{fold}_n(\cdot)$ is the operator that converts the matrix back to the tensor format along the mode-n.

3.2. Model Formulation

The fundamental solution for HSI reconstruction is to solve an optimization problem with prior knowledge based regularization. Therefore, it is essential to fully leverage the intrinsic properties behind the desired signal and develop an appropriate regularization. So far two kinds of utilized priors for HSI recovery are the spectral correlation and the spatial self-similarity. The spectral correlation states a fact that HSI contains a small amount of basis materials and thus exhibits rich spectral redundancy. While the spatial self-similarity indicates that, for each exemplar patch, we can find many similar patches sharing the homologous structure. It has been shown that such two prior knowledge are very helpful for HSI processing [29, 30, 31, 32]. Take such prior knowledge into full consideration, we formulate 3D tensors to promote the reconstruction performance.

We first partition the HSI $\boldsymbol{\mathcal{X}}$ into overlapped cubic patches with the spatial block size of $s \times s$ across full spectral bands. After lexicographically reordering the spatial block of each band into a 1D column vector, we search k nearest similar neighbors for each patch and get a corresponding 3D tensor $\boldsymbol{\mathcal{P}}$ with the size of $s^2 \times \Lambda \times (k+1)$, where Λ is the spectral number. Although it is someway reasonable to construct a 4D tensor to preserve the 2D spatial structure, the spatial size $s \times s$ is usually set to be small, which can not extract the intrinsic spatial features. While

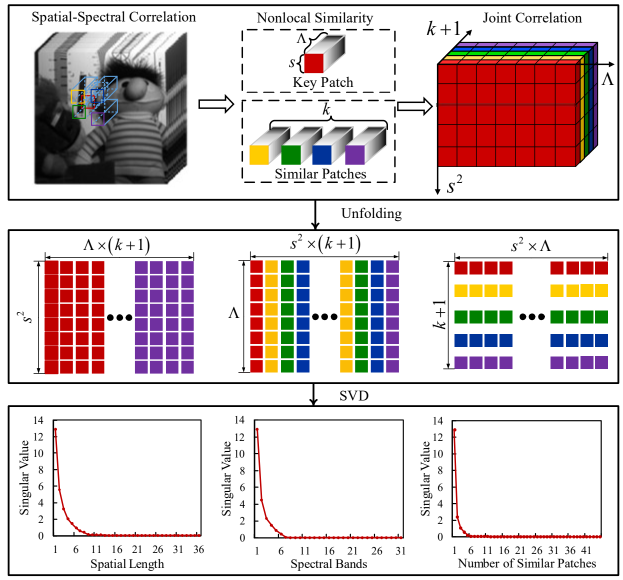


Figure 3: Low-rank property analysis. We exploit the non-local similarity across spatial and spectral domain to reformulate a low-rank tensor in the first row. We unfold the tensor along each mode in the second row and show the corresponding singular value in the third row.

for a large s, unaffordable computation and memory burden will be introduced in practice.

The constructed 3D tensors simultaneously utilize the spatial self-similarity (mode-1), the spectral correlation (mode-2) and the joint correlation (mode-3), which would benefit us for investigating the priors in a unified framework. Consequently, we can obtain a basic representation towards low-rank tensor recovery:

$$\Gamma(\boldsymbol{\mathcal{P}}) = \tau \left\| \mathbf{R}(\boldsymbol{\mathcal{X}}) - \boldsymbol{\mathcal{P}} \right\|_F^2 + \text{rank}(\boldsymbol{\mathcal{P}}), \tag{1}$$

where $\mathbf{R}(\cdot)$ represents the operator extracting the 3D tensor from $\boldsymbol{\mathcal{X}}$ and τ denotes the penalty factor.

With the Tucker decomposition [39], the tensor rank is characterized by the sum of ranks of unfolding matrices along each mode, i.e. $\text{rank}(\boldsymbol{\mathcal{P}}) = \sum_{n=1}^{3} \text{rank}(\boldsymbol{\mathcal{P}}_{(n)})$. However, simply summing the ranks lacks a clear investigation on different subspaces. In Fig. 3, we analyze the low-rank property of the constructed 3D tensor. We extract a nonlocal tensor from a clean HSI and implement SVD along each mode. The singular values of each unfolding matrix tend to be dropping to zero fleetly. But discrepancy exists in the descending rates of singular values of different modes. For example, due to the strong redundancy in non-local similar patches, the magnitude of singular values in mode-3 tends to decrease to zero with a relatively faster speed. Thus, a single penalty factor acted on all modes would produce poor tensor recovery results. Modes of lower rank still contain biased redundancy while those of higher rank are overestimated.

Considering such discrepancy among different modes, it is beneficial to estimate rank($\mathcal{P}$) in a dimension-discriminative manner. To address this issue, we propose to utilize the sum of weighted ranks regularization to exploit the multi-dimensionality diversity:

$$\Gamma(\mathcal{P}) = \tau \|\mathbf{R}(\mathcal{X}) - \mathcal{P}\|_F^2 + \sum_{n=1}^{3} w_n \text{rank}(\mathcal{P}_{(n)}), \quad (2)$$

where $w_n > 0$ is the weight factor that can be regarded as the importance measurement of mode-n. In the following, we will illustrate the DLTR based reconstruction method for computational hyperspectral imaging.

4. Dimension-discriminative Low-rank Tensor Recovery Based Reconstruction

4.1. Representative Systems

We give a brief introduction on two representative computational hyperspectral imaging systems, i.e., CASSI and DCD. It is worth noting that our method is also suited for other computational imaging systems, such as multiple snapshot imaging system [40, 41], spatial-spectral encoded imaging system[42] and so on.

The incident light in the CASSI system, as shown in Fig. 4, is first projected onto the plane of a coded aperture through the objective lens. After spatial modulation by a coded aperture, the incident light goes through a relay lens and is spectrally dispersed in the vertical direction by Amici prism. Finally, the modulated and dispersed spectral information is captured by a grayscale camera. Let $\mathcal{X} \in \mathbb{R}^{M\times N\times\Lambda}$ denotes the original HSI and $x(i, j, \lambda)$ is its element, where $1\leq i\leq M$, $1\leq j\leq N$ index the spatial coordinate and $1\leq\lambda\leq\Lambda$ indexes the spectral coordinate. The compressive measurement at position (i, j) on the focal plane of CASSI can be represented as:

$$y^c(i,j) = \sum_{\lambda=1}^{\Lambda} \rho(\lambda)\varphi(i-\phi(\lambda), j)x(i-\phi(\lambda), j, \lambda), \quad (3)$$

where $\varphi(i, j)$ denotes the pattern of the coded aperture, $\phi(\lambda)$ denotes the dispersion introduced by Amici prism and $\rho(\lambda)$ is the spectral response of the detector. For brevity, let $\boldsymbol{Y}^c$ denote the matrix representation of $y^c(i, j)$. Then the matrix form of CASSI imaging can be expressed as:

$$\boldsymbol{Y}^c = \boldsymbol{\Phi}^c(\mathcal{X}), \quad (4)$$

where $\boldsymbol{\Phi}^c$ is the forward imaging function (jointly determined by $\rho(\lambda)$, $\varphi(i, j)$ and $\phi(\lambda)$).

The incident light in the DCD system, as shown in Fig. 4, is first divided into two directions by the beam splitter equivalently. The light in one direction is captured by the

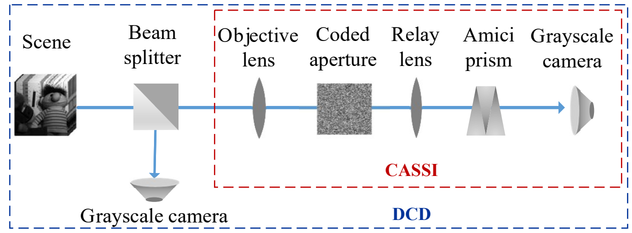


Figure 4: Diagram of two representative computational hyperspectral imaging systems.

CASSI system, while the light on the other direction is captured directly by a grayscale camera. The compressive measurement on the panchromatic detector can be represented as:

$$y^p(i,j) = 0.5 \sum_{\lambda=1}^{\Lambda} \rho(\lambda)x(i,j,\lambda). \quad (5)$$

Similar to the CASSI formulation in Equation (4), Equation (5) can also be rewritten in the matrix form as:

$$\boldsymbol{Y}^p = \boldsymbol{\Phi}^p(\mathcal{X}), \quad (6)$$

where $\boldsymbol{Y}^p$ denotes the uncoded measurement and $\boldsymbol{\Phi}^p$ is the forward imaging function of the grayscale camera.

A general imaging representation of computational hyperspectral imaging can be derived as:

$$\boldsymbol{Y} = \boldsymbol{\Phi}(\mathcal{X}). \quad (7)$$

For CASSI, $\boldsymbol{Y} = \boldsymbol{Y}^c$ and $\boldsymbol{\Phi} = \boldsymbol{\Phi}^c$, For DCD, $\boldsymbol{Y} = [\boldsymbol{Y}^c; \boldsymbol{Y}^p]$ and $\boldsymbol{\Phi} = [\boldsymbol{\Phi}^c; \boldsymbol{\Phi}^p]$. Actually, the goal of computational imaging reconstruction is to estimate $\mathcal{X}$ from the compressive measurement $\boldsymbol{Y}$.

4.2. Reconstruction Formulation

To exploit the multi-dimensional intrinsic property of HSI analyzed in Section 3, a general reconstruction model with DLTR can be derived as follows:

$$\begin{aligned}\min_{\mathcal{X},\mathcal{P}^l} \ \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{\Phi}(\mathcal{X})\|_F^2 + \sum_{l=1}^{L} \Big(\tau \left\|\mathbf{R}^l(\mathcal{X}) - \mathcal{P}^l\right\|_F^2 \\ + \sum_{n=1}^{3} w_n \text{rank}(\boldsymbol{P}^l_{(n)})\Big),\end{aligned} \quad (8)$$

where L denotes the total number of exemplar tensors. Minimizing the rank of matrix in Equation (8) is a NP-hard problem. To make a more explicit and accurate estimation, we further propose to introduce a non-convex relaxation of nuclear norm using a log-sum form:

$$\begin{aligned}\min_{\mathcal{X},\mathcal{P}^l} \ \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{\Phi}(\mathcal{X})\|_F^2 + \sum_{l=1}^{L} \Big(\tau \left\|\mathbf{R}^l(\mathcal{X}) - \mathcal{P}^l\right\|_F^2 \\ + \sum_{n=1}^{3} w_n \Theta(\mathcal{P}^l_{(n)}, \epsilon)\Big),\end{aligned} \quad (9)$$

where $\Theta(\boldsymbol{\mathcal{P}}^l_{(n)}, \epsilon) = \sum_r \log(\sigma_r(\boldsymbol{\mathcal{P}}^l_{(n)}) + \epsilon)$, ϵ is a small positive constant number and $\sigma(\boldsymbol{\mathcal{P}}^l_{(n)})$ defines the singular values of $\boldsymbol{\mathcal{P}}^l_{(n)}$.

Let $f(\sigma, \epsilon) = \log(\sigma + \epsilon)$, $f(\sigma, \epsilon)$ derives a meaningful outcome that the singular values can be disposed adaptively. Specifically, $f(\sigma, \epsilon)$ can be approximated using the first-order Taylor expansion at $\sigma = \sigma^t$, i.e., $f(\sigma, \epsilon) = f(\sigma^t, \epsilon) + < \nabla f(\sigma^t, \epsilon), \sigma - \sigma^t >$, where σ^t denotes the result of the t-th iteration. Actually, the solution of $\min\limits_{\sigma} f(\sigma, \epsilon)$ can be reasonably approximated by $\min\limits_{\sigma}(\sigma/(\sigma^t + \epsilon))$ since the constants in the first Taylor expansion don't influence the minimization problem and can be neglected. Therefore, $\min \Theta(\boldsymbol{\mathcal{P}}^l_{(n)}, \epsilon) = \min \sum_r (\sigma_r(\boldsymbol{\mathcal{P}}^l_{(n)})/(\sigma^t_r(\boldsymbol{\mathcal{P}}^l_{(n)}) + \epsilon))$. It means that those greater singular values of the tth iteration, which deliver more important information, will get a smaller weight and be shrunk by a lower amplitude at $(t+1)$-th iteration. Therefor, structure information can be preserved.

The log-sum based regularization model, which can be effectively solved by singular value thresholding, benefits us for reconstructing HSI with DLTR. An numerical algorithm for solving Equation (9) is then proposed in the following section.

4.3. Optimization Algorithm

To optimize Equation (9), we adopt an alternating minimization scheme to split it into two finer subproblems: updating low-rank tensor $\boldsymbol{\mathcal{P}}^l$ and updating the whole HSI $\boldsymbol{\mathcal{X}}$.

4.3.1 Updating Low-rank Tensor $\boldsymbol{\mathcal{P}}^l$

By fixing HSI $\boldsymbol{\mathcal{X}}$, we can estimate each low-rank tensor $\boldsymbol{\mathcal{P}}^l$ independently by solving the following reformulated equation:

$$\min_{\boldsymbol{\mathcal{P}}^l} \left\| \mathbf{R}^l(\boldsymbol{\mathcal{X}}) - \boldsymbol{\mathcal{P}}^l \right\|_F^2 + \sum\nolimits_{n=1}^{3} w_n \Theta(\boldsymbol{\mathcal{P}}^l_{(n)}, \epsilon). \tag{10}$$

Here, we employ an alternating minimization strategy to facilitate the problem. Specifically, by initializing $\boldsymbol{\mathcal{P}}^l_{(0)} = \mathbf{R}^l(\boldsymbol{\mathcal{X}})$, we estimate low-rank tensor $\boldsymbol{\mathcal{P}}^l$ in a mode-wise way:

$$\min_{\boldsymbol{\mathcal{P}}^l_{(n)}} \frac{1}{2} \left\| \boldsymbol{\mathcal{P}}^l_{(n-1)} - \boldsymbol{\mathcal{P}}^l_{(n)} \right\|_F^2 + \alpha_n \sum\nolimits_r \log(\sigma_r(\boldsymbol{\mathcal{P}}^l_{(n)}) + \epsilon), \tag{11}$$

where $\alpha_n = w_n/2\tau$. For $n = 1, 2$ and 3, Equation (11) admits a closed-form and can be updated by the following:

$$\boldsymbol{\mathcal{P}}^l_{(n)} = \text{fold}_n(\boldsymbol{U}\boldsymbol{\Sigma}_{\alpha_n}\boldsymbol{V}^T), \tag{12}$$

where $\boldsymbol{\Sigma}_{\alpha_n} = \text{diag}(\mathbf{S}_{\alpha_n,\epsilon}(\sigma_1), \mathbf{S}_{\alpha_n,\epsilon}(\sigma_2), ..., \mathbf{S}_{\alpha_n,\epsilon}(\sigma_{D_n}))$, and $\boldsymbol{U}\text{diag}(\sigma_1, \sigma_2, ..., \sigma_{D_n})\boldsymbol{V}^T$ is SVD of $\boldsymbol{\mathcal{P}}^l_{(n-1)}$. The singular value thresholding operation $\mathbf{S}_{\alpha,\epsilon}(\sigma)$ is defined as:

$$\mathbf{S}_{\alpha,\epsilon}(\sigma) = \begin{cases} 0 & \text{if } c_1 \leq 0 \\ \text{sign}(\sigma)(\frac{c_0 + \sqrt{c_1}}{2}) & \text{if } c_1 > 0 \end{cases} \tag{13}$$

with that $c_0 = |\sigma| - \epsilon$, and $c_1 = (c_0)^2 - 4(\alpha - \epsilon|\sigma|)$ [43]. Finally, we update the low-rank tensor with $\boldsymbol{\mathcal{P}}^l = \text{fold}_3(\boldsymbol{\mathcal{P}}^l_{(3)})$.

4.3.2 Updating Whole HSI $\boldsymbol{\mathcal{X}}$

Once we obtain the low-rank tensor $\boldsymbol{\mathcal{P}}^l$, the whole HSI $\boldsymbol{\mathcal{X}}$ can be updated by solving the following problem:

$$\min_{\boldsymbol{\mathcal{X}}} \frac{1}{2} \|\boldsymbol{Y} - \boldsymbol{\Phi}(\boldsymbol{\mathcal{X}})\|_F^2 + \sum\nolimits_{l=1}^{L} \tau \left\| \mathbf{R}^l(\boldsymbol{\mathcal{X}}) - \boldsymbol{\mathcal{P}}^l \right\|_F^2. \tag{14}$$

Actually, Equation (14) is a quadratic optimization problem and admits a straightforward least-square solution:

$$\begin{aligned} \boldsymbol{\mathcal{X}} = \big(\boldsymbol{\Phi}^T\boldsymbol{\Phi} + 2\tau \sum\nolimits_l \mathbf{R}^{l^T}\mathbf{R}^l\big)^{-1} \big(\boldsymbol{\Phi}^T(\boldsymbol{Y}) \\ + 2\tau \sum\nolimits_{l=1}^{L} \mathbf{R}^{l^T}(\boldsymbol{\mathcal{P}}^l)\big), \end{aligned} \tag{15}$$

which can be solved by a conjugate gradient algorithm.

5. Experiments on Synthetic Data

In this section, we compare the performance of our method with both prior knowledge based methods and deep learning based methods via experiments on synthetic data.

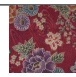
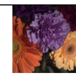
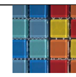
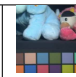
5.1. Implementation Details

Datasets. Three public HSI datasets, including the Columbia dataset [44], the Harvard dataset [45] and the KAIST dataset [27] are used as synthetic data. These datasets contain various real-world objects of different materials. In our experiment, the resolution of all tested images is cropped into 256×256 and 31 spectral bands (400nm to 700nm in Columbia, while 420nm to 720nm in Harvard and KAIST). And all tested HSIs are scaled into the interval [0,1].

Comparison Methods. Our algorithm is compared with 5 prior knowledge based methods, i.e., TV regularization integrated TwIST [46], gradient projection for sparse reconstruction (GPSR) [16], approximate message passing (AMP) [24], nonlocal sparse representation (NSR) [17] and low-rank matrix approximation (LRMA) [18] and 3 deep learning based methods, i.e., AE [27], HSCNN [28] and ISTA [26].

Parameters Setting. For the competitive methods, we make efforts to achieve their best results according to their released codes or publication suggestions. For our method, the penalty factor $\tau = 1$, the adaptive weight $w = [0.1, 0.1, 1]c$, and $c = 10^{-4} \sim 10^{-3}$ is a constant. We

Table 1: Reconstruction results (PSNR(dB)/SSIM/ERGAS) of the 10 HSIs for different methods on CASSI.

Methods	Indexes											Average
TV[12]	PSNR	20.13	18.51	34.58	25.46	20.98	20.19	23.63	20.89	21.42	25.83	23.16
	SSIM	0.675	0.429	0.928	0.845	0.643	0.626	0.571	0.794	0.770	0.846	0.713
	ERGAS	304.93	315.73	160.80	168.38	253.94	341.67	210.29	303.65	251.24	272.62	258.32
GPSR[16]	PSNR	20.74	18.74	31.79	22.37	19.89	19.89	22.32	21.74	21.60	25.36	22.44
	SSIM	0.581	0.431	0.838	0.721	0.556	0.576	0.507	0.732	0.679	0.703	0.632
	ERGAS	295.29	308.48	227.97	238.21	287.67	353.09	246.55	281.43	244.13	291.71	277.46
AMP[24]	PSNR	19.61	18.49	33.68	23.94	21.19	19.86	22.91	20.91	22.96	28.30	23.18
	SSIM	0.570	0.399	0.901	0.770	0.605	0.546	0.524	0.739	0.720	0.827	0.660
	ERGAS	322.56	320.12	178.82	201.53	246.97	354.04	229.17	304.57	207.02	202.82	256.76
NSR[17]	PSNR	23.46	19.12	37.06	29.53	23.41	21.52	24.49	24.25	26.90	**31.60**	26.13
	SSIM	0.756	0.453	**0.959**	0.901	0.705	0.672	0.583	0.827	0.840	0.919	0.761
	ERGAS	207.60	297.59	122.49	105.15	193.31	293.67	191.57	208.17	131.18	**141.20**	189.19
LRMA[18]	PSNR	22.73	19.31	37.38	30.98	24.14	22.04	25.02	23.00	24.93	29.89	25.94
	SSIM	0.806	0.470	0.948	0.944	0.771	0.743	0.608	0.847	0.863	0.933	0.793
	ERGAS	228.00	292.05	117.27	91.65	179.84	279.03	181.52	243.49	167.42	176.00	195.63
Ours	PSNR	**24.27**	**19.70**	**37.41**	**32.50**	**25.58**	**23.20**	**25.62**	**24.43**	**27.16**	31.57	**27.15**
	SSIM	**0.855**	**0.495**	0.948	**0.948**	**0.802**	**0.788**	**0.626**	**0.870**	**0.901**	**0.939**	**0.817**
	ERGAS	**193.35**	**279.83**	**116.57**	**75.29**	**153.06**	**245.12**	**169.83**	**212.34**	**129.62**	146.50	**172.15**

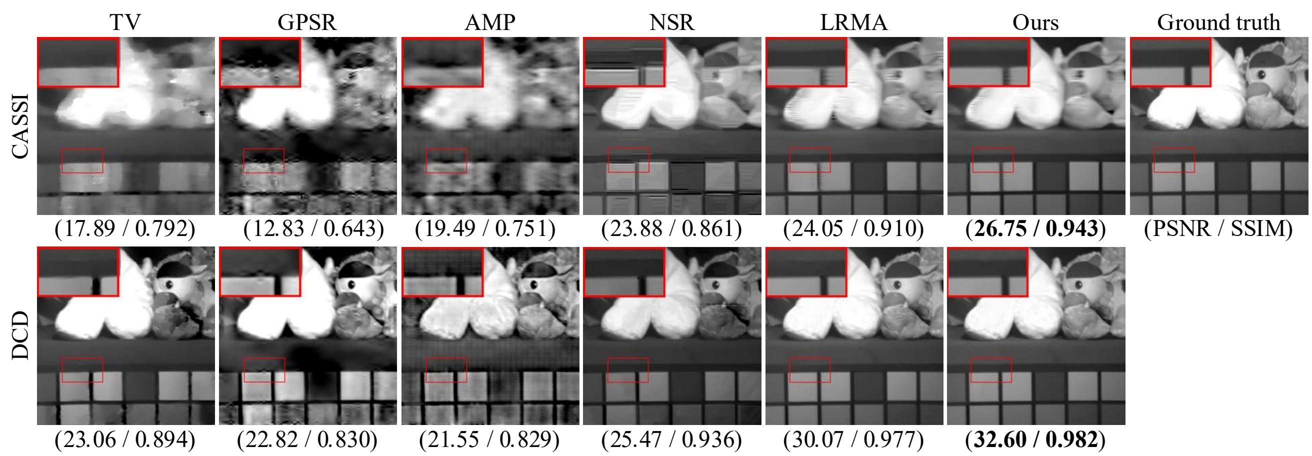


Figure 5: Visual results comparison for *stuffed toys* on CASSI and DCD at 410nm. The average PSNR and SSIM values are presented in parenthesis. Our proposed method obtains better results on both computational imaging systems.

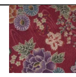
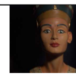
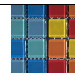
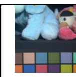
set the spatial patch size $s = 6$ with the overlapping size of 5 empirically. We search 45 nearest similar patches within a $[-20, 20] \times [-20, 20]$ window.

Evaluation Measures. For quantitative evaluation, three image quality indexes are adopted: peek signal-to-noise ratio (PSNR), structure similarity (SSIM) [47] and erreur relative globale adimensionnelle de synthèse (ERGAS)[48]. PSNR measures the visual quality, SSIM measures the structure similarity and ERGAS measures the spectral fidelity. Generally, a bigger PSNR and SSIM and a smaller ERGAS suggest a better reconstruction accuracy.

5.2. Performance Evaluation

We first compare our method with prior knowledge based methods on both CASSI and DCD to verify its performance. The quantitative results on the Columbia dataset are shown in Table 1 and Table 2, respectively. The best results for each image are highlighted in bold. Comparing the results within the same system, we can see that our method obtains a noticeable promotion in PSNR, SSIM and ERGAS on both systems. We show the reconstructed images of *stuffed toys* on both systems in Fig. 5, which can also demonstrate the superiorty of our method. Specifi-

Table 2: Reconstruction results (PSNR(dB)/SSIM/ERGAS) of the 10 HSIs for different methods on DCD.

Methods	Indexes											Average
TV[12]	PSNR	29.25	21.17	47.19	26.84	22.73	25.54	27.15	30.08	26.87	28.28	28.51
	SSIM	0.921	0.793	0.994	0.912	0.822	0.880	0.873	0.854	0.956	0.934	0.894
	ERGAS	156.67	250.06	38.57	155.41	217.18	188.39	150.89	126.06	142.41	245.79	167.14
GPSR[16]	PSNR	27.62	25.99	36.97	25.86	24.69	25.13	27.88	29.48	26.47	27.92	27.80
	SSIM	0.846	0.866	0.937	0.855	0.833	0.851	0.779	0.851	0.929	0.892	0.864
	ERGAS	181.06	133.41	140.06	164.91	172.74	194.91	138.77	144.54	151.69	241.40	166.35
AMP[24]	PSNR	27.66	23.52	38.77	28.92	26.53	24.97	27.38	26.44	28.38	32.62	28.52
	SSIM	0.829	0.793	0.955	0.880	0.841	0.819	0.882	0.760	0.869	0.897	0.853
	ERGAS	137.55	174.13	102.23	119.43	137.27	197.58	139.14	163.35	111.78	125.03	140.75
NSR[17]	PSNR	31.01	22.27	48.49	36.77	30.87	27.21	28.02	35.45	30.11	35.58	32.58
	SSIM	0.944	0.831	**0.996**	0.976	0.943	0.904	0.966	0.884	0.973	0.959	0.938
	ERGAS	135.46	223.83	33.61	47.36	85.87	156.51	137.69	59.04	98.78	98.90	107.71
LRMA[18]	PSNR	40.65	28.87	48.20	40.11	34.31	33.85	34.71	41.22	37.57	35.03	37.45
	SSIM	0.989	0.927	0.991	0.989	0.971	0.970	0.973	0.945	0.991	0.986	0.973
	ERGAS	34.49	99.71	35.34	32.24	57.01	72.43	59.72	30.35	40.41	111.30	57.30
Ours	PSNR	**42.02**	**29.79**	**48.91**	**40.67**	**35.37**	**34.80**	**35.27**	**41.94**	**38.30**	**35.88**	**38.29**
	SSIM	**0.991**	**0.938**	0.991	**0.990**	**0.976**	**0.976**	**0.977**	**0.950**	**0.992**	**0.988**	**0.977**
	ERGAS	**25.98**	**89.58**	**33.08**	**30.13**	**50.36**	**64.68**	**55.79**	**27.90**	**37.06**	**103.58**	**51.81**

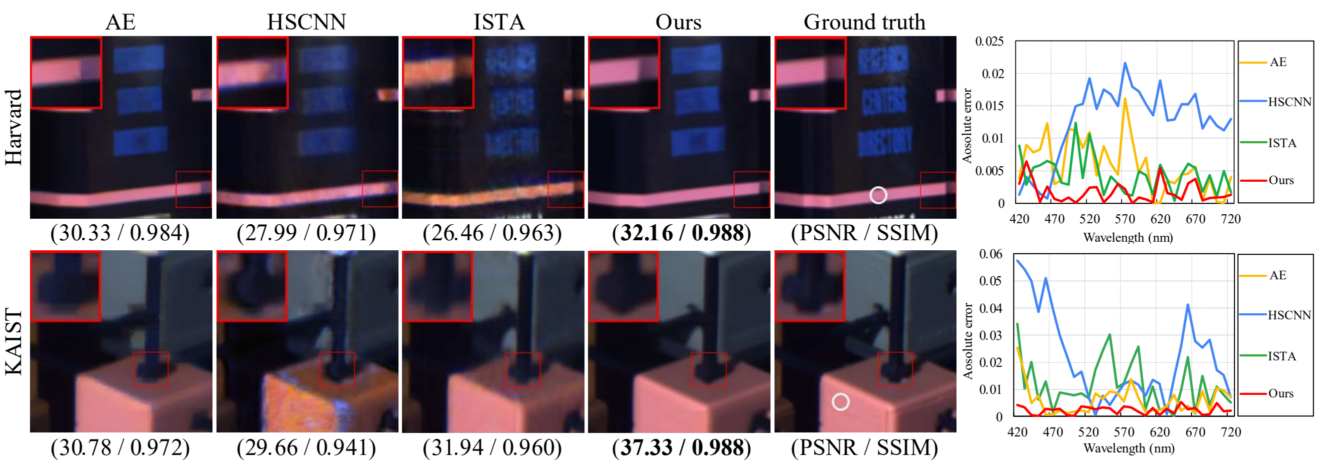
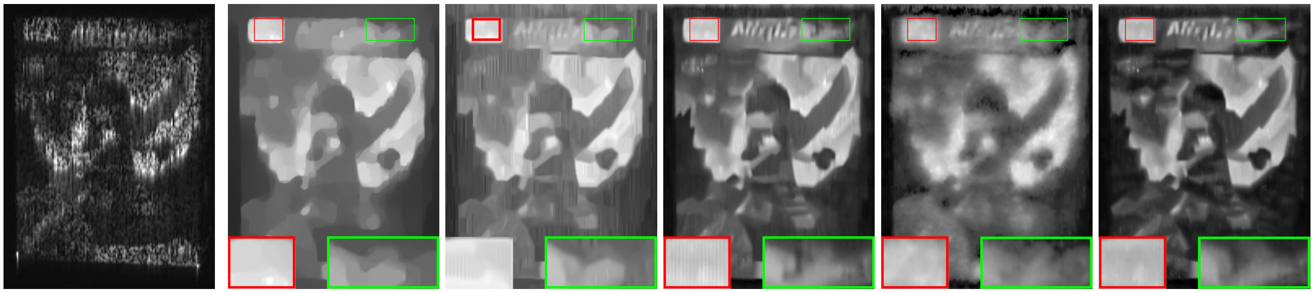


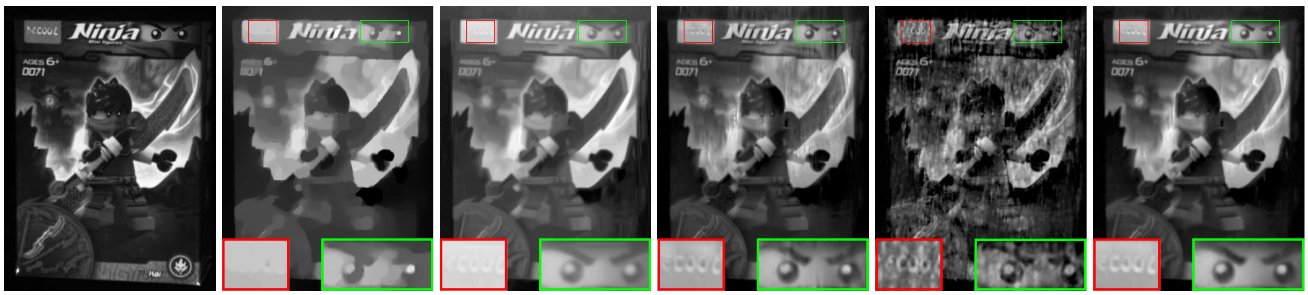
Figure 6: Reconstructed quality comparison. The average PSNR and SSIM values are presented in parenthesis. The offside curves show the absolute spectral error of the white labels on ground truth. Our proposed method outperforms all three deep learning based methods in terms of spatial vision and spectral accuracy.

cally, our method produces remarkable spatial quality promotion compared with TV, AMP and NSR. It indicates that low rank prior can recover more structure information than sparse prior. Further, the promotion upon LRMA indicates that the high-dimensional tensor is more powerful than matrix to exploit the intrinsic characteristic of HSI.

Then, we carry out a thorough comparison with three deep learning based methods on both the Harvard dataset and the KAIST dataset. Specifically, for the Harvard dataset, we use 35 HSIs for training and 9 HSIs for testing, for the KAIST dataset, we use 29 HSIs for training and 3 HSIs for testing. The average numerical results of CASSI are presented in Table 3. It can be seen that our method outperforms all three deep learning based methods as well. Extensive comparison, including the visual results and the absolute error in the spectral distribution, are shown in Fig. 6. For better vision, we convert the reconstructed HSI into RGB using the CIE color mapping function. It shows that our method can exhibit not only clearer spatial details but also higher spectral fidelity.



(a) CASSI image (b) TV-CASSI (c) NSR-CASSI (d) LRMA-CASSI (e) AE-CASSI (f) Ours-CASSI



(g) DCD image (h) TV-DCD (i) NSR-DCD (j) LRMA-DCD (k) AE-DCD (l) Ours-DCD

Figure 7: Reconstruction performance comparison on real data. Compared with other methods, our method can produce better visual details.

Table 3: Comparison with 3 deep learning based methods on CASSI. The results are obtained by averaging the testing images on 2 datasets. Our method produces a notable promotion upon deep learning based methods.

Datasets	Indexes	AE	HSCNN	ISTA	Ours
Harvard	PSNR	29.20	27.60	29.87	**31.14**
	SSIM	0.912	0.900	0.913	**0.932**
	ERGAS	91.62	105.90	85.21	**74.92**
KAIST	PSNR	26.03	21.06	28.57	**34.87**
	SSIM	0.920	0.814	0.909	**0.962**
	ERGAS	172.32	273.37	151.49	**37.40**

5.3. Real Data

We further evaluate the performance of the proposed method using the real data. The captured scene is a cartoon cover under the laboratory ambient light condition. The compressive measurement of CASSI is shown in Fig. 7(a) and the panchromatic image of DCD is shown in Fig. 7(g). It can be seen that the proposed method can obtain better results with clearer contents and abundant textures compared with the other methods on both systems. By contrast, TV produces over-smoothing results, both NSR and LRMA introduce noise and lost spatial details, while AE suffers from severe artifacts. Therefore, our method is also potential to be applied into the real nature scene.

6. Conclusion

In this paper, we have proposed a general reconstruction model for computational hyperspectral imaging based on the dimension-discriminative low-rank tensor recovery. We utilized 3D tensors to exploit the intrinsic properties of hyperspectral image, including spatial self-similarity, spectral correlation and joint correlation to boost the reconstruction performance. To clearly figure out the multi-dimension diversity, a dimensional-discriminative low-rank tensor recovery model based on a weighted nuclear norm regularization is developed. Then the low-rank regularization and the imaging process are unified in a general reconstruction model, which is effectively solved by an iterative numerical algorithm. The executive experiments on CASSI and DCD have verified the superior performance and the application perspective of our method. In our future work, we will investigate the dimension-discriminative low-rank tensor recovery to more actual computational imaging systems.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grant 61425013, Grant 61701025, and Grant 61672096, and in part by the Beijing Municipal Science and Technology Commission under Grant Z181100003018003.

References

[1] Min H. Kim, Todd Alan Harvey, David S. Kittle, Holly Rushmeier, Julie Dorsey, Richard O. Prum, and David J. Brady. 3d imaging spectroscopy for measuring hyperspectral patterns on solid objects. *ACM Transactions on Graphics*, 31(4):1–11, 2012.

[2] Zhihong Pan, Glenn Healey, Manish Prasad, and Bruce Tromberg. Face recognition in hyperspectral images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1552–1560, 2003.

[3] Hien Van Nguyen, Amit Banerjee, and Rama Chellappa. Tracking via object reflectance using a hyperspectral video camera. In *IEEE Conference on Computer Vision and Pattern Recognition - Workshops*, pages 44–51, 2010.

[4] Wallace M. Porter and Harry T. Enmark. A system overview of the airborne visible/infrared imaging spectrometer (aviris). In *Imaging Spectroscopy II*, volume 834, pages 22–32. International Society for Optics and Photonics, 1987.

[5] Robert W. Basedow, Dwayne C. Carmer, and Mark E. Anderson. Hydice system: Implementation and performance. In *Imaging Spectrometry*, volume 2480, pages 258–268. International Society for Optics and Photonics, 1995.

[6] Masahiro Yamaguchi, Hideaki Haneishi, Hiroyuki Fukuda, Junko Kishimoto, Hiroshi Kanazawa, Masaru Tsuchida, Ryo Iwama, and Nagaaki Ohyama. High-fidelity video and still-image communication based on spectral information: Natural vision system and its applications. In *Spectral Imaging: Eighth International Symposium on Multispectral Color Science*, volume 6062, page 60620G. International Society for Optics and Photonics, 2006.

[7] Yoav Y. Schechner and Shree K. Nayar. Generalized mosaicing: Wide field of view multispectral imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(10):1334–1348, 2002.

[8] Takayuki Okamoto and Ichirou Yamaguchi. Simultaneous acquisition of spectral image information. *Optics Letters*, 16(16):1277–1279, 1991.

[9] Michael Descour and Eustace Dereniak. Computed-tomography imaging spectrometer: experimental calibration and reconstruction results. *Applied Optics*, 34(22):4817–4826, 1995.

[10] Xun Cao, Hao Du, Xin Tong, Qionghai Dai, and Stephen Lin. A prism-mask system for multispectral video acquisition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2423–2435, 2011.

[11] Gonzalo R. Arce, David J. Brady, Lawrence Carin, Henry Arguello, and David S. Kittle. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1):105–115, 2014.

[12] Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied Optics*, 47(10):B44–B51, 2008.

[13] Lizhi Wang, Zhiwei Xiong, Dahua Gao, Guangming Shi, and Feng Wu. Dual-camera design for coded aperture snapshot spectral imaging. *Applied Optics*, 54(4):848–858, 2015.

[14] Lizhi Wang, Zhiwei Xiong, Dahua Gao, Guangming Shi, Wenjun Zeng, and Feng Wu. High-speed hyperspectral video acquisition with a dual-camera architecture. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4942–4950, 2015.

[15] Ashwin A. Wagadarikar, Nikos P. Pitsianis, Xiaobai Sun, and David J. Brady. Spectral image estimation for coded aperture snapshot spectral imagers. In *Image Reconstruction from Incomplete Data V*, volume 7076, page 707602. International Society for Optics and Photonics, 2008.

[16] Mário A.T. Figueiredo, Robert D. Nowak, and Stephen J. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. In *IEEE Journal of Selected Topics in Signal Processing*, pages 586 – 597, 2007.

[17] Lizhi Wang, Zhiwei Xiong, Guangming Shi, Feng Wu, and Wenjun Zeng. Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(10):2104–2111, 2017.

[18] Ying Fu, Yinqiang Zheng, Imari Sato, and Yoichi Sato. Exploiting spectral-spatial correlation for coded hyperspectral image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3727–3736, 2016.

[19] Lizhi Wang, Chen Sun, Ying Fu, Min H Kim, and Hua Huang. Hyperspectral image reconstruction using a deep spatial-spectral prior. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 8032–8041, 2019.

[20] Yang Liu, Xin Yuan, Jinli Suo, David Brady, and Qionghai Dai. Rank minimization for snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[21] Lizhi Wang, Zhiwei Xiong, Hua Huang, Guangming Shi, Feng Wu, and Wenjun Zeng. High-speed hyperspectral video acquisition by combining nyquist and compressive sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(4):857–870, 2018.

[22] Shipeng Zhang, Huang Hua, and Ying Fu. Fast parallel implementation of dual-camera compressive hyperspectral imaging system. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018.

[23] Shipeng Zhang, Lizhi Wang, Ying Fu, and Hua Huang. Gpu assisted towards real-time reconstruction for dual-camera compressive hyperspectral imaging. In *Pacific Rim Conference on Multimedia*, pages 711–720. Springer, 2018.

[24] Jin Tan, Yanting Ma, Hoover Rueda, Dror Baron, and Gonzalo R. Arce. Compressive hyperspectral imaging via approximate message passing. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):389–401, 2016.

[25] Jian Sun, Huibin Li, and Zongben Xu. Deep admm-net for compressive sensing mri. In *Advances in Neural Information Processing Systems*, pages 10–18, 2016.

[26] Jian Zhang and Bernard Ghanem. Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1828–1837, 2018.

[27] Inchang Choi, Daniel S. Jeon, Giljoo Nam, Diego Gutierrez, and Min H. Kim. High-quality hyperspectral reconstruction using a spectral prior. *ACM Transactions on Graphics*, 36(6):218, 2017.

[28] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, Dong Liu, and Feng Wu. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. In *IEEE International Conference on Computer Vision*, pages 518–525, 2017.

[29] Chang Li, Yong Ma, Jun Huang, Xiaoguang Mei, and Jiayi Ma. Hyperspectral image denoising using the robust low-rank tensor recovery. *JOSA A*, 32(9):1604–1612, 2015.

[30] Yi Peng, Deyu Meng, Zongben Xu, Chenqiang Gao, Yi Yang, and Biao Zhang. Decomposable nonlocal tensor dictionary learning for multispectral image denoising. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2949–2956, 2014.

[31] Qi Xie, Qian Zhao, Deyu Meng, and Zongben Xu. Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(8):1888–1902, 2018.

[32] Yi Chang, Luxin Yan, and Sheng Zhong. Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5901–5909, 2017.

[33] Na Qi, Yunhui Shi, Xiaoyan Sun, Jingdong Wang, Baocai Yin, and Junbin Gao. Multi-dimensional sparse models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(1):163–178, 2018.

[34] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[35] Bernardino Romera-Paredes and Massimiliano Pontil. A new convex relaxation for tensor completion. In *Advances in Neural Information Processing Systems*, pages 2967–2975, 2013.

[36] Ji Liu, Przemyslaw Musialski, Peter Wonka, and Jieping Ye. Tensor completion for estimating missing values in visual data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):208–220, 2013.

[37] Ting Xie, Shutao Li, Leyuan Fang, and Licheng Liu. Tensor completion via nonlocal low-rank regularization. *IEEE Transactions on Cybernetics*, pages 1–11, 2018.

[38] Wenfei Cao, Yao Wang, Can Yang, Xiangyu Chang, Zhi Han, and Zongben Xu. Folded-concave penalization approaches to tensor completion. *Neurocomputing*, 152:261–273, 2015.

[39] Ledyard R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

[40] David Kittle, Kerkil Choi, Ashwin Wagadarikar, and David J. Brady. Multiframe image estimation for coded aperture snapshot spectral imagers. *Applied Optics*, 49(36):6824, 2010.

[41] Yuehao Wu, Iftekhar O. Mirza, Gonzalo R. Arce, and Dennis W. Prather. Development of a digital-micromirror-device-based multishot snapshot spectral imaging system. *Optics Letters*, 36(14):2692–2694, 2011.

[42] Xing Lin, Yebin Liu, Jiamin Wu, and Qionghai Dai. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Transactions on Graphics*, 33(6):1–11, 2014.

[43] Pinghua Gong, Changshui Zhang, Zhaosong Lu, Jianhua Huang, and Jieping Ye. A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In *International Conference on Machine Learning*, pages 37–45, 2013.

[44] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K. Nayar. Generalized assorted pixel camera: post-capture control of resolution, dynamic range, and spectrum. *IEEE Transactions on Image Processing*, 19(9):2241–2253, 2010.

[45] Ayan Chakrabarti and Todd Zickler. Statistics of real-world hyperspectral images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 193–200, 2011.

[46] José M Bioucas-Dias and Mário AT Figueiredo. A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image Processing*, 16(12):2992–3004, 2007.

[47] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[48] Lucien Wald. *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*. Presses des MINES, 2002.