Exploiting Spectral-Spatial Correlation for Coded Hyperspectral Image Restoration

Ying Fu[1], Yinqiang Zheng[2], Imari Sato[2], Yoichi Sato[1]
[1]The University of Tokyo [2]National Institute of Informatics

Abstract

Conventional scanning and multiplexing techniques for hyperspectral imaging suffer from limited temporal and/or spatial resolution. To resolve this issue, coding techniques are becoming increasingly popular in developing snapshot systems for high-resolution hyperspectral imaging. For such systems, it is a critical task to accurately restore the 3D hyperspectral image from its corresponding coded 2D image. In this paper, we propose an effective method for coded hyperspectral image restoration, which exploits extensive structure sparsity in the hyperspectral image. Specifically, we simultaneously explore spectral and spatial correlation via low-rank regularizations, and formulate the restoration problem into a variational optimization model, which can be solved via an iterative numerical algorithm. Experimental results using both synthetic data and real images show that the proposed method can significantly outperform the state-of-the-art methods on several popular coding-based hyperspectral imaging systems.

1. Introduction

Hyperspectral (HS) imaging captures light from any scene point over tens and hundreds of bands in the spectral domain. Such detailed spectral distribution information has given rise to numerous applications [1] , including diagnostic medicine [2, 3], remote sensing [4, 5], surveillance [6, 7], and more.

To capture a full HS image, traditional HS imaging methods [8, 9, 10, 11, 12] need to scan along either the spatial or the spectral dimension, and they often sacrifice temporal resolution due to the limitations of hardware in perceiving light. To enable hyperspectral acquisition of dynamic scenes, snapshot approaches [13, 14, 15, 16] are developed to capture the full 3D spectral cube in a single image, which multiplex the 3D HS image into a 2D spatial sensor, at the cost of reducing spatial resolution.

Recently, some coding-based HS imaging approaches [17, 18, 19, 20, 21, 22] have been proposed to overcome the tradeoff between temporal and spatial resolution, relying on the compressive sampling (CS) theory. All these imaging approaches are under-determined, and their underlying restoration methods exploit the l_1-norm based sparsity of HS images. Since the number of measurements is far less than that of variables in the desired HS image, the l_1-norm based constraints are still insufficient for accurate hyperspectral image restoration. This inspires us to better exploit the intrinsic properties of a HS image, i.e. the high correlation across spectra [23] and the non-local self-similarity in space [24].

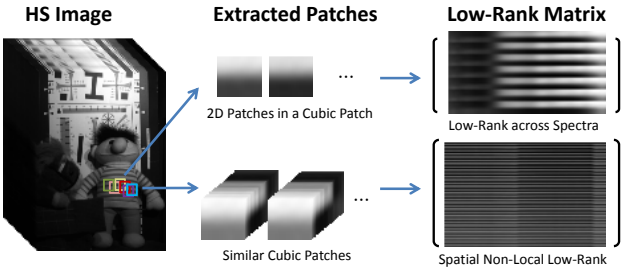


Figure 1. Illustration of the low-rank matrices from a HS image. Each cubic patch is reshaped as a 2D matrix, where each row describes the spectral distribution of each pixel. This low-rank matrix encodes the correlation across spectra. Besides, a set of similar patches for each exemplar patch are grouped into a low-rank matrix, which accounts for the spatial non-local similarities.

In this paper, we propose an effective coded HS image restoration method, by exploiting spectral and spatial correlation via low-rank approximation (Figure 1). Specifically, to utilize the sparsity across spectra, we reshape each exemplar patch as a 2D matrix, where each row describes the spectral distribution of each spatial pixel, and use the spectral low-rank constraint on it. To take into account the non-local self-similarity in space, we group a set of similar patches for each exemplar patch and enforce the spatial non-local low-rank regularization on this set. In addition, we employ the weighted nuclear norm as a smooth surrogate function for the low-rank regularization, which can adaptively adjust the regularization parameters. Later, these two low-rank regularizations are involved into a uni-

fied variational optimization model, which can be efficiently solved via an iterative numerical algorithm. The effectiveness of our method is demonstrated on several coding-based HS imaging systems, which outperforms the state-of-the-art methods designed for these systems on synthetic and real data.

In summary, our main contributions are that we

1. Present an effective and universal method for coded HS image restoration, and demonstrate it on several recent coding-based HS imaging systems;
2. Exploit the intrinsic properties of a HS image—the high correlation across spectra and spatial non-local self-similarity—via proper low-rank regularizations;
3. Develop an iterative numerical algorithm to efficiently solve the proposed model and adaptively adjust the regularization parameters.

The remainder of this paper is organized as follows. Section 2 reviews related works. Low-rank approximation for universal coded HS image restoration is presented in Section 3, while the mathematical representation of several representative coding-based imaging systems is shown in Section 4. We present extensive experimental results in Section 5 and conclude this work in Section 6.

2. Related Works

In the following, we will review the most relevant studies on HS imaging system and low-rank approximation.

2.1. Hyperpectral Imaging System

Conventional HS cameras rely on certain scanning techniques. For example, whiskbroom and pushbroom based [8, 9] acquisition systems scanned the full scene pointwisely or linewisely. In contrast, rotating and tunable filters based systems [10, 11] scanned throughout the spectral dimension. Spatial variant color filters [12] were also used to capture different spectra at different points. All these methods suffer from limited temporal resolution.

To enable dynamic scene acquisition, snapshot approaches are developed to capture the full 3D HS image in a single image, which multiplex the 3D HS image into a 2D spatial sensor by sacrificing the spatial resolution. Examples include computed tomography imaging spectrometer [13], the 4D imaging spectrometer [14], the snapshot image mapping spectrometer [15], and the prism-mask system for multispectral video imaging [16].

Recently, some coding-based snapshot HS imaging approaches have been proposed to overcome the tradeoff between temporal and spatial resolution. The coded aperture snapshot spectral imager (CASSI) employed two dispersers [17] (or one disperser later in [18]) with a coding aperture to uniformly encode the optical signals along space by using CS-based methods. The performance of CASSI could be improved by using multiple shots with changing coded masks [19, 20], or dual camera design (DCD) with another aligned panchromatic camera [21]. Later, a compressive hyperspectral imager (SSCSI) [22] was presented to jointly encode the spatial and spectral dimensions in a single gray image.

All these coding-based HS imaging systems rely on the CS theory and the full HS image is restored by employing the l_1-norm based sparsity of the HS image. In contrast, we investigate more intrinsic properties of a HS image, i.e. high correlation across spectra and non-local self-similarity among space, and develop a unified variational framework for accurate HS image restoration, which jointly exploits the redundancy across spectra and space via low-rank regularizations.

2.2. Low-Rank Approximation

Low-rank approximation seeks to recover the underlying low-rank matrix from degraded observations, which is shown to be tractable [25] by solving its convex nuclear norm relaxation. Cai *et al.* [26] further developed the singular value thresholding scheme for fast computation. To improve the flexibility of nuclear norm, Fazel *et al.* [27] used the logarithm of the determinant as a smooth approximation of rank, which is a non-convex surrogate of the rank. Gu *et al.* [28] showed that weighted nuclear norm minimization could effectively improve the restoration results by adaptively adjusting the weights for each singular value in the optimization process. Lu *et al.* [29] studied the generalized singular value thresholding to solve the general non-convex surrogate of rank.

Low-rank approximation has been widely used in image restoration [23, 30], image alignment [31], transform invariant texture modeling [32], background modeling [33], reflection separation [34] and more. In this work, we will resort to it for coded HS image restoration.

3. Coded Hyperspectral Image Restoration

In this section, we first explore spectral and spatial correlation, and then show how to incorporate it into HS image restoration via low-rank regularizations. Finally, an iterative numerical algorithm is developed for solving.

3.1. Spectral and Spatial Correlation

It is well known that large sets of spectra can be properly represented by low dimensional linear models [35]. This implies that different spectra of realistic scenes assume rich redundancy. Due to difference in material distribution, the degree of correlation varies across different patches of the HS image. To account for this property, we properly divide the HS image into cubic patches, as illustrated in Figure 1.

Specifically, let $\mathcal{S} \in \mathbb{R}^{M\times N\times B}$ denote the original 3D HS image, and $\mathbf{S} \in \mathbb{R}^{MNB}$ be the vectorized form of $\mathcal{S}$, in

which M, N and B stand for the number of image rows, columns and spectral bands, respectively. The HS image $\mathcal{S}$ is first divided into overlapping cubic patches of size $P \times P \times B$, where $P < M$ and $P < N$. Let vector $\boldsymbol{s}_{i,j} \in \mathbb{R}^{P^2B}$ denote the vectorized form of a cubic patch extracted from $\mathbf{S}$ and centered at the spatial location (i,j). $\boldsymbol{s}_{i,j}$ can be described as

$$\boldsymbol{s}_{i,j} = \mathbf{R}_{i,j}\mathbf{S}, \tag{1}$$

where $\mathbf{R}_{i,j} \in \mathbb{R}^{P^2B \times MNB}$ is the matrix extracting patch $\boldsymbol{s}_{i,j}$ from $\mathbf{S}$. Now, we introduce a linear transform operator $\mathcal{T} : \mathbb{R}^{P^2B} \rightarrow \mathbb{R}^{P^2 \times B}$ that reshapes the vectorized cubic patch $\boldsymbol{s}_{i,j}$ as 2D matrix, such that each row of $\mathcal{T}(\boldsymbol{s}_{i,j})$ denotes the spectral distribution of each pixel. In practice, $\mathcal{T}(\boldsymbol{s}_{i,j})$ may be corrupted by some noise. We thus model the matrix $\mathcal{T}(\boldsymbol{s}_{i,j})$ as $\mathcal{T}(\boldsymbol{s}_{i,j}) = \mathbf{A}_{i,j} + \mathbf{N}_{i,j}$, where $\mathbf{A}_{i,j}$ and $\mathbf{N}_{i,j}$ describe the desired low-rank matrix and the Gaussian noise matrix, respectively. The spectral low-rank matrix $\mathbf{A}_{i,j}$ can be recovered by

$$\hat{\mathbf{A}}_{i,j} = \arg\min_{\mathbf{A}_{i,j}} \text{rank}(\mathbf{A}_{i,j}),\ s.t.\ \|\mathcal{T}(\boldsymbol{s}_{i,j}) - \mathbf{A}_{i,j}\|_F^2 \leq \sigma_A^2, \tag{2}$$

where $\|\cdot\|_F^2$ denotes the Frobenius norm and σ_A^2 is the variance of the additive Gaussian noise. The minimization problem in Equation (2) can be solved by its Lagrangian form,

$$\hat{\mathbf{A}}_{i,j} = \arg\min_{\mathbf{A}_{i,j}} \beta\|\mathcal{T}(\boldsymbol{s}_{i,j}) - \mathbf{A}_{i,j}\|_F^2 + \alpha\text{rank}(\mathbf{A}_{i,j}). \tag{3}$$

Equation (3) is equivalent to Equation (2), when a proper parameter for β/α is chosen.

The cubic patches in the HS image also have rich self-similarity with its neighboring patches [36] in the spatial domain, which implies that the grouped similar patches for each exemplar patch assume low-rank structures. We call it as the spatial non-local low-rankness to distinguish it with the aforementioned spectral low-rankness.

For each exemplar patch, its similar patches are searched by k-nearest neighbor method within a local window centered at (i,j). Let $\tilde{\boldsymbol{R}}_{i,j}\mathbf{S} = [\boldsymbol{R}_{i,j,1}\mathbf{S}, \boldsymbol{R}_{i,j,2}\mathbf{S}, \cdots, \boldsymbol{R}_{i,j,k}\mathbf{S}] = [\boldsymbol{s}_{i,j,1}, \boldsymbol{s}_{i,j,2}, \cdots, \boldsymbol{s}_{i,j,k}]$ denote the formed matrix by the set of the similar patches for the exemplar patch $\boldsymbol{s}_{i,j}$. Each column in $\tilde{\boldsymbol{R}}_{i,j}\mathbf{S}$ represents a vectorized similar patch to $\boldsymbol{s}_{i,j}$. Similar to the description for spectral low-rank approximation, we employ $\mathbf{B}_{i,j}$ to represent the desired non-local low-rank matrix and the spatial non-local low-rank matrix approximation can be described as

$$\hat{\mathbf{B}}_{i,j} = \arg\min \eta\|\tilde{\boldsymbol{R}}_{i,j}\mathbf{S} - \mathbf{B}_{i,j}\|_F^2 + \gamma\text{rank}(\mathbf{B}_{i,j}). \tag{4}$$

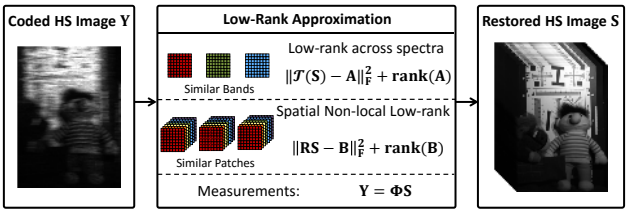


Figure 2. Overview of the proposed method for coded HS image restoration by using spectral low-rank and spatial non-local low-rank regularizations.

3.2. Restoration via Low-Rank Regularization

As will be shown in Section 4, a coding-based HS imaging system can be described in general as $\mathbf{Y} = \mathbf{\Phi S}$, in which $\mathbf{Y}$ and $\mathbf{\Phi}$ denote the image observations and the projection operator, respectively. To exploit the rich redundancy across spectra and non-local self-similarity along space in the HS image, the coded HS image restoration task can be achieved by solving the following regularized optimization problem

$$\begin{aligned}(\hat{\mathbf{S}}, \hat{\mathbf{A}}_{i,j}, \hat{\mathbf{B}}_{i,j}) = \arg\min \|\mathbf{Y} - \mathbf{\Phi S}\|_F^2 + \sum_{i,j} \Big(\\ \beta\|\mathcal{T}(\boldsymbol{s}_{i,j}) - \mathbf{A}_{i,j}\|_F^2 + \alpha\text{rank}(\mathbf{A}_{i,j}) + \\ \gamma\|\tilde{\boldsymbol{R}}_{i,j}\mathbf{S} - \mathbf{B}_{i,j}\|_F^2 + \eta\text{rank}(\mathbf{B}_{i,j}) \Big).\end{aligned} \tag{5}$$

The overall framework of the proposed method is shown in Figure 2.

The rank function $\text{rank}(\mathbf{A}_{i,j})$ (or $\text{rank}(\mathbf{B}_{i,j})$) in Equation (5) is non-convex, which is proven to be NP-hard and all known algorithms for exactly solving it are doubly exponential [25]. A tractable approach is to optimize its convex envelope, i.e. nuclear norm $\|\cdot\|_*$, and thus solve it via convex optimization [25]. The nuclear norm of $\mathbf{A}_{i,j}$ is defined as the summation of all singular values, i.e. $\|\mathbf{A}_{i,j}\|_* = \sum_{r=1}^{n_A} \sigma_r(\mathbf{A}_{i,j})$. Similarly, $\|\mathbf{B}_{i,j}\|_* = \sum_{r=1}^{n_B} \sigma_r(\mathbf{B}_{i,j})$.

To improve the flexibility of nuclear norm, Gu *et al.* [28] showed that weighted nuclear norm can effectively improve the restoration results by adaptively adjusting the weight for each singular value in the optimization processing. The weighted nuclear norm of matrix $\mathbf{A}_{i,j}$ is formulated as

$$\|\mathbf{A}_{i,j}\|_{w,*} = \sum_{r=1}^{n} w_{Ar}\sigma_r(\mathbf{A}_{i,j}), \tag{6}$$

where $w_{Ar} \geq 0$ is a non-negative weight for $\sigma_r(\mathbf{A}_{i,j})$.

For natural images, we have the general prior knowledge that larger singular values are more important, and should be less shrunk. Therefore, it is reasonable to set the weight

w_{Ar} to be inversely proportional to $\sigma_r(\mathbf{A}_{i,j})$, i.e.

$$w_{Ar} = \frac{1}{\sigma_r(\mathbf{A}_{i,j}) + \epsilon}, \tag{7}$$

where ϵ is a small constant value. Similar definition applies to $\mathbf{B}_{i,j}$.

Therefore, the optimization problem for coded HS image restoration in Equation (5) can be further relaxed as

$$\begin{aligned}(\hat{\mathbf{S}}, \hat{\mathbf{A}}_{i,j}, \hat{\mathbf{B}}_{i,j}) = \arg\min \|\mathbf{Y} - \mathbf{\Phi}\mathbf{S}\|_F^2 + \sum_{i,j} \Big(\\ \beta\|\mathcal{T}(\boldsymbol{s}_{i,j}) - \mathbf{A}_{i,j}\|_F^2 + \alpha\|\mathbf{A}_{i,j}\|_{w,*} + \\ + \gamma\|\tilde{\boldsymbol{R}}_{i,j}\mathbf{S} - \mathbf{B}_{i,j}\|_F^2 + \eta\|\mathbf{B}_{i,j}\|_{w,*} \Big).\end{aligned} \tag{8}$$

3.3. Numerical Algorithm

The proposed model in Equation (8) has three sets of variables, i.e. the full HS image $\mathbf{S}$, the spectral low-rank matrices $\mathbf{A}_{i,j}$ and the spatial non-local low-rank matrices $\mathbf{B}_{i,j}$. To solve Equation (8), we adopt an alternating minimization scheme to split the original problem into three simpler subproblems as follows.

Update $\mathbf{A}_{i,j}$. Given an initial estimate of the latent high resolution HS image $\mathbf{S}$, we first extract patch $\boldsymbol{s}_{i,j}$ and reshape it as 2D matrix $\mathcal{T}(\boldsymbol{s}_{i,j})$, as described in Section 3.1. Each matrix $\mathbf{A}_{i,j}$ can be recovered by

$$\mathbf{A}_{i,j}^{(t)} = \arg\min \beta\|\mathcal{T}(\boldsymbol{s}_{i,j}^{(t-1)}) - \mathbf{A}_{i,j}\|_F^2 + \alpha\|\mathbf{A}_{i,j}\|_{w,*}, \tag{9}$$

where $a^{(t)}$ represents the t-th iteration of any variable a.

Substituting Equation (6) into Equation (9), we can obtain

$$\begin{aligned}\mathbf{A}_{i,j}^{(t)} = \arg\min \frac{1}{2}\|\mathcal{T}(\boldsymbol{s}_{i,j}^{(t-1)}) - \mathbf{A}_{i,j}\|_F^2 \\ + \frac{\alpha}{2\beta}\sum_{r=1}^{n_A} w_{Ar}^{(t-1)}\sigma_r(\mathbf{A}_{i,j}),\end{aligned} \tag{10}$$

where $n_A = \min\{P^2, B\}$. According to [28][30], Equation (10) can be optimized by

$$\mathbf{A}_{i,j}^{(t)} = \boldsymbol{U}_{\boldsymbol{A}}\left(\mathbf{\Sigma}_A - \frac{\alpha}{2\beta}\text{diag}(\boldsymbol{w}_A^{(t-1)})\right)_+ \boldsymbol{V}_A^T, \tag{11}$$

where $\boldsymbol{U}_A\mathbf{\Sigma}_A\boldsymbol{V}_A^T$ is the SVD of $\mathcal{T}(\boldsymbol{s}_{i,j}^{(t-1)})$, $\boldsymbol{w}_A^{(t-1)} = [w_{A_1}^{(t-1)}, w_{A_2}^{(t-1)}, \cdots, w_{A_n}^{(t-1)}]$ is the vectorized representation of the weight in (6) and is calculated by Equation (7), and $(x)_+ = \max\{x, 0\}$.

Update $\mathbf{B}_{i,j}$. With the known latent HS image, we group similar patches for the exemplar patch $\boldsymbol{s}_{i,j}$, as described in Section 3.1. Each matrix $\mathbf{B}_{i,j}$ can be obtained by optimizing

$$\mathbf{B}_{i,j}^{(t)} = \arg\min \gamma\|\tilde{\boldsymbol{R}}_{i,j}\mathbf{S}^{(t-1)} - \mathbf{B}_{i,j}\|_F^2 + \eta\|\mathbf{B}_{i,j}\|_{w,*}. \tag{12}$$

Substituting Equation (6) into Equation (12), we can obtain

$$\begin{aligned}\mathbf{B}_{i,j}^{(t)} = \arg\min \frac{1}{2}\|\tilde{\boldsymbol{R}}_{i,j}\mathbf{S}^{(t-1)} - \mathbf{B}_{i,j}\|_F^2 \\ + \frac{\eta}{2\gamma}\sum_{r=1}^{n_B} w_{Br}^{(t-1)}\sigma_r(\mathbf{B}_{i,j}),\end{aligned} \tag{13}$$

where $n_B = \min\{P^2B, k\}$. Similar to Equation (10), Equation (13) can be optimized by

$$\mathbf{B}_{i,j}^{(t)} = \boldsymbol{U}_B\left(\mathbf{\Sigma}_B - \frac{\eta}{2\gamma}\text{diag}(\boldsymbol{w}_B^{(t-1)})\right)_+ \boldsymbol{V}_B^T, \tag{14}$$

where $\boldsymbol{U}_B\mathbf{\Sigma}_B\boldsymbol{V}_B^T$ is the SVD of $\tilde{\boldsymbol{R}}_{i,j}\mathbf{S}^{(t-1)}$.

Update S. After solving for each $\mathbf{A}_{i,j}$ and $\mathbf{B}_{i,j}$, the latent HS image can be reconstructed by solving optimization problem

$$\begin{aligned}\mathbf{S}^{(t)} = \arg\min \|\mathbf{Y} - \mathbf{\Phi}\mathbf{S}\|_F^2 + \sum_{i,j}\Big(\beta\|\mathcal{T}(\boldsymbol{s}_{i,j}) - \mathbf{A}_{i,j}^{(t)}\|_F^2 \\ + \gamma\|\tilde{\boldsymbol{R}}_{i,j}\mathbf{S} - \mathbf{B}_{i,j}^{(t)}\|_F^2\Big).\end{aligned} \tag{15}$$

Equation (15) is a quadratic minimization problem and we use a conjugate gradient algorithm to solve it.

In our implementation, the spatial size P of the cubic patch is chosen to be 6. The search region for similar patches is in $[-20, 20] \times [-20, 20]$, and the nearest 45 patches are used. As for the weighting parameters in Equation (8), we have chosen $\beta = \gamma = 10^{-1} \sim 1$ and $\alpha = \eta = 10^{-4} \sim 10^{-3}$.

4. Representative Coding-based Imaging Systems

Here, we show the mathematical representation for three representative coding-based HS imaging systems, including CASSI [18], DCD [21] and SSCSI [22][1].

In the CASSI system, as shown in Figure 3(a), the scene is first projected into the coded aperture, which plays a spatial modulation. Then, the spatially modulated information is spectrally dispersed by the prism and captured by a gray camera. The imaging process for the (i, j)-th pixel can be described by the following integral over the wavelength λ

$$Y^h(i, j) = \int s(i + \psi^h(\lambda), j, \lambda) f(i + \psi^h(\lambda), j) c(\lambda) \mathrm{d}\lambda, \tag{16}$$

where $s(i, j, \lambda)$ denotes the spectral distribution of the (i, j)-th pixel of the latent HS image. $\psi^h(\lambda)$ is the wavelength-dependent dispersion function for the prism [18]. $f(i, j)$ is the transmission function of the coded aperture. $c(\lambda)$ represents the response function of the detector.

[1] Our method can also be used for other coding-based HS imaging systems as well, like the multiple snapshot capture system [19, 20].

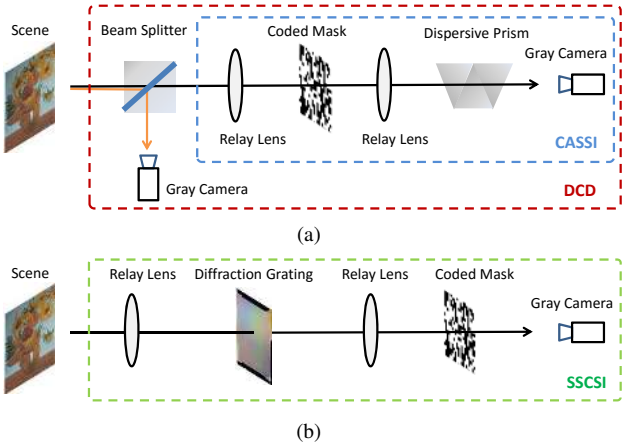


Figure 3. Illustration of the three representative imaging systems.

In the DCD system, as shown in Figure 3(a), the incident light from the scene is firstly split by a beam splitter. The light in one direction is captured by CASSI, while the light in the other direction is captured by a panchromatic camera. The captured image by the panchromatic camera can be described as

$$Y^p(i,j) = \int s(i,j,\lambda)c(\lambda)\mathrm{d}\lambda. \tag{17}$$

In the SSCSI system, as illustrated in Figure 3(b)[2], a diffraction grating is applied to disperse the light into spectrum plane and a coded attenuation mask is inserted between the spectrum plane and the sensor plane to perform spatial-spectral modulation. The coded image can be formulated as

$$Y^s(i,j) = \int s(i,j,\lambda)f(i+\psi^s(i,\lambda),j)c(\lambda)\mathrm{d}\lambda, \tag{18}$$

where $\psi^s(i,\lambda)$ is the spatial location and wavelength dependent dispersion function [22].

Generally, the spectral dimension can be discretized into B bands. Let $\mathbf{Y}^h$, $\mathbf{Y}^p$ and $\mathbf{Y}^s$ be the vectorized representation of the image captured by CASSI $Y^h(i,j)$, image directly captured by the panchromatic camera $Y^p(i,j)$ and image captured by SSCSI $Y^s(i,j)$, respectively. The matrix representation of the three systems can be written as

$$\begin{aligned} \mathbf{Y}^h &= \mathbf{\Phi}^h\mathbf{S} + \boldsymbol{n}^h, \\ \mathbf{Y}^p &= \mathbf{\Phi}^p\mathbf{S} + \boldsymbol{n}^p, \\ \mathbf{Y}^s &= \mathbf{\Phi}^s\mathbf{S} + \boldsymbol{n}^s, \end{aligned} \tag{19}$$

where $\mathbf{\Phi}^h$ is the projection matrix of the CASSI system and jointly determined by $f(i,j)$, $\psi^h(\lambda)$ and $c(\lambda)$. $\mathbf{\Phi}^p$ is the projection matrix of the panchromatic camera and determined by $c(\lambda)$. $\mathbf{\Phi}^s$ is the projection matrix of the SSCSI system and jointly determined by $f(i,j)$, $\psi^s(i,\lambda)$ and $c(\lambda)$. $\boldsymbol{n}^h$, $\boldsymbol{n}^p$ and $\boldsymbol{n}^s$ are the additive noise from CASSI, the panchromatic camera and SSCSI, which are usually modeled as Gaussian noise.

The imaging system can be generally expressed as

$$\mathbf{Y} = \mathbf{\Phi}\mathbf{S} + \boldsymbol{n}. \tag{20}$$

For CASSI, $\mathbf{Y} = \mathbf{Y}^h$, $\mathbf{\Phi} = \mathbf{\Phi}^h$ and $\boldsymbol{n} = \boldsymbol{n}^h$. For DCD, $\mathbf{Y} = [\mathbf{Y}^h; \mathbf{Y}^p]$, $\mathbf{\Phi} = [\mathbf{\Phi}^h; \mathbf{\Phi}^p]$, and $\boldsymbol{n} = [\boldsymbol{n}^h; \boldsymbol{n}^p]$. For SSCSI, $\mathbf{Y} = \mathbf{Y}^s$, $\mathbf{\Phi} = \mathbf{\Phi}^s$ and $\boldsymbol{n} = \boldsymbol{n}^s$. For each system, the projection matrix $\mathbf{\Phi}$ can be calibrated in system construction. Given $\mathbf{\Phi}$, our goal is to recover the full 3D HS image $\mathbf{S}$ from the incomplete measurements $\mathbf{Y}$.

The number of measurements is MN in both CASSI and SSCSI, while $2MN$ for DCD. Obviously, the restoration task is severely under-determined, since the number of measurements from these imaging systems is far less than that of variables in the desired HS image. The coding mechanism for these systems relies on the CS theory, and can be decoded by adding sparsity regularization. CASSI [18] and DCD [21] adopt the total variation regularization on the latent HS image. SSCSI [22] learns the dictionary by using K-SVD method [37] from HS image datasets, and adds the sparse constraint by l_1-norm on the coefficients of the latent HS image. [22] shows that the restoration results are sensitive to the dictionary learning method. In our method, we use low-rank approximation for CS recovery, and exploit the intrinsic correlation properties of HS images along both spectral and spatial dimension. Our method does not need to learn the dictionary, and is thus immune to the drawbacks arising from dictionary learning.

5. Experimental Results

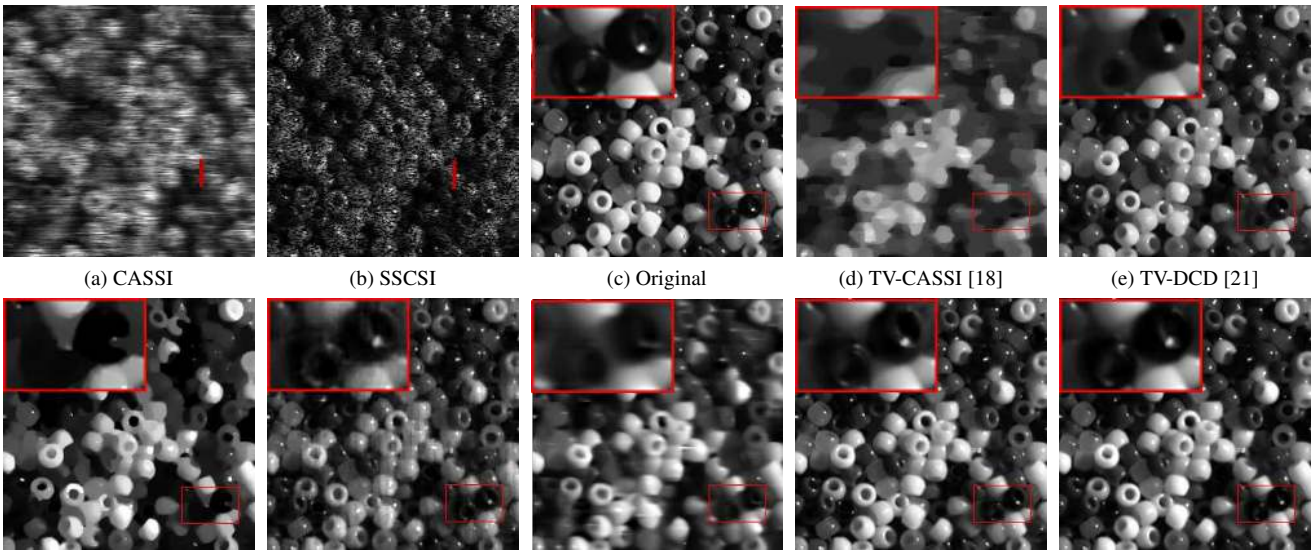
In this section, we evaluate our method for coded HS image restoration on synthetic data and real images.

5.1. Synthetic Data

The HS images in Columbia Multispectral Image Database [38] are used to synthesize data. To show the effectiveness of our proposed method, we use 10 different scenes and compare the restoration results on different imaging systems, including CASSI, DMD and SSCSI.

As for the competing restoration methods, we consider the two-step iterative shrinkage/thresholding along with the total variation (TV) regularization [39], which is used in [18] and [21]. To compare with the restoration method in [22], we generate the results using the basis pursuit denoise optimization [40] with the learned over-complete dictionary by K-SVD, which is named as the dictionary-based reconstruction (DBR). All the parameters involved in the competing algorithms are optimally set or automatically chosen as

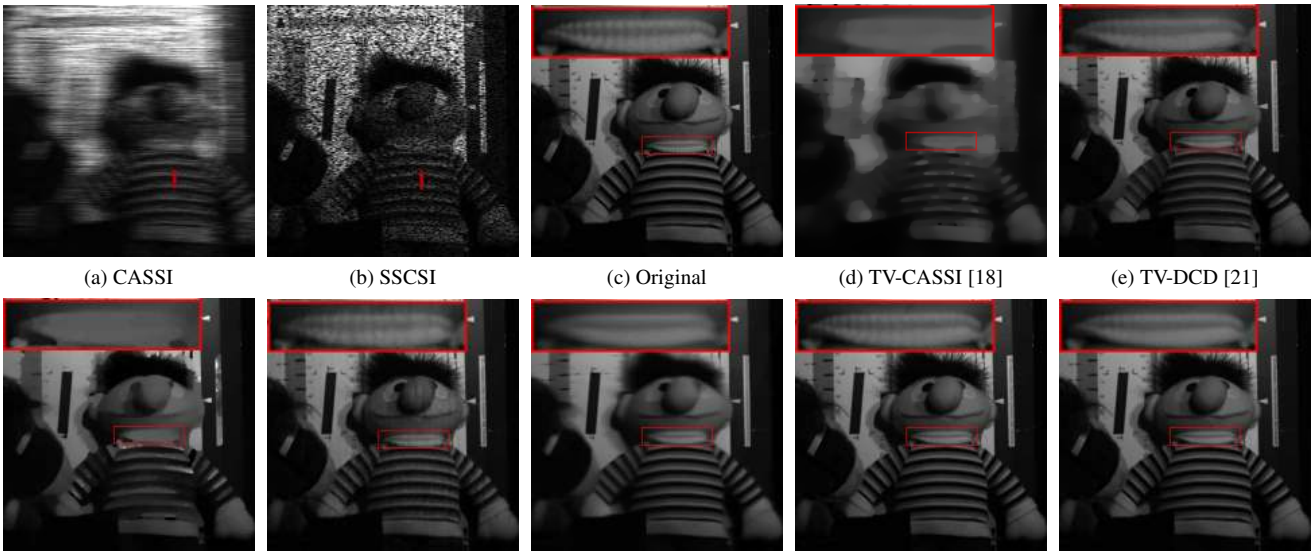
[2]This figure shows the simplified imaging system. The full system can be found in [22].



(a) CASSI (b) SSCSI (c) Original (d) TV-CASSI [18] (e) TV-DCD [21]

(f) TV-SSCSI (g) DBR-SSCSI [22] (h) Ours-CASSI (i) Ours-DCD (j) Ours-SSCSI

Figure 4. Visual quality comparison for the HS image *beans* on CASSI, DCD and SSCSI imaging systems. (a) and (b) show the synthesized coded 2D image for CASSI and SSCSI, respectively. (c) shows the original image at 620nm. Restoration results at this band are shown in (d-j).



(a) CASSI (b) SSCSI (c) Original (d) TV-CASSI [18] (e) TV-DCD [21]

(f) TV-SSCSI (g) DBR-SSCSI [22] (h) Ours-CASSI (i) Ours-DCD (j) Ours-SSCSI

Figure 5. Visual quality comparison for the HS image *chart and stuffed toy* on CASSI, DCD and SSCSI imaging systems.

described in the references. We implement our method on these three imaging systems as well. To accelerate convergence, we use the restoration results from the TV method as initialization for our method. We use *method-system* to denote the combination of a restoration method and an imaging system, i.e. TV-CASSI, TV-DCD, TV-SSCSI, DBR-SSCSI, Ours-CASSI, Ours-DCD and Ours-SSCSI.

In Figure 4 and 5, we show the restored results for the *beans* and *chart and stuffed toy* scenes in one spectral band for different methods and imaging systems. As can be seen in Figure 4 and 5, restorations from TV-CASSI and TV-SSCSI suffer from obvious spatial blurring. The restoration quality is improved by DBR-SSCSI, but with more noise. In comparison, TV-DCD produces more details of underlying scene, e.g. the edges of beans and the clothes of the toy (Figure 4 and 5(e)). Compared with all these methods, our method offers much better restoration results. We can see that Ours-CASSI can recover more details than TV-CASSI does, but the restoration results still have some blurring. Ours-DCD improves the results over TV-DCD, and recovers nice texture/edge features with rich details. Besides, our method on SSCIS, i.e. Ours-SSCSI, significantly outper-

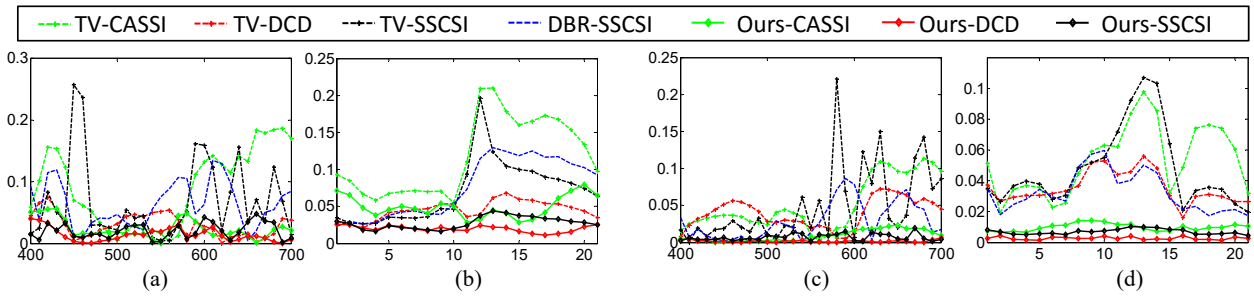


Figure 6. (a) and (c) show the absolute difference from 400nm to 700nm between the ground truth and the restoration results at the center pixel of the labeled lines in Figure 4(a) and 5(a), respectively, for different restoration methods and imaging systems. (b) and (d) show the corresponding RMSE of spectral distribution at all 21 pixels in the labeled lines of Figure 4(a) and 5(b), respectively.

Table 1. Restoration results (PSNR(dB)/SSIM/SAM) of the 10 HS images for different methods and imaging systems.

Methods	Metrics										
TV-CASSI	PSNR	20.62	24.10	34.46	22.92	27.99	32.92	26.64	26.71	28.41	27.32
	SSIM	0.5467	0.7998	0.9247	0.4633	0.8836	0.9012	0.6805	0.7975	0.9203	0.8084
	SAM	0.3165	0.1981	0.2118	0.1713	0.1918	0.2978	0.1845	0.1839	0.0888	0.2181
TV-DCD	PSNR	27.68	34.22	40.87	30.84	40.44	43.10	33.38	34.14	35.22	38.12
	SSIM	0.8569	0.9609	0.9533	0.8781	0.9695	0.9851	0.8797	0.9366	0.9769	0.9591
	SAM	0.2481	0.1425	0.1648	0.1212	0.1299	0.1483	0.1291	0.1174	0.0664	0.1265
TV-SSCSI	PSNR	21.99	27.01	35.10	22.80	35.99	36.92	29.98	33.50	32.86	31.06
	SSIM	0.7507	0.9148	0.9021	0.6350	0.9582	0.9663	0.7804	0.9261	0.9706	0.9128
	SAM	0.3230	0.1642	0.2389	0.2205	0.1737	0.1991	0.1733	0.1253	0.0826	0.1977
DBR-SSCSI	PSNR	24.20	32.28	28.18	28.24	33.21	39.65	32.33	29.49	26.22	34.20
	SSIM	0.7871	0.9503	0.8306	0.8196	0.9074	0.9826	0.9152	0.8885	0.8561	0.9370
	SAM	0.3641	0.1678	0.3954	0.1860	0.2552	0.2558	0.1959	0.2131	0.1914	0.2927
Ours-CASSI	PSNR	24.52	32.12	40.64	26.56	35.74	37.62	30.87	34.61	40.51	34.08
	SSIM	0.7410	0.9286	0.9405	0.7894	0.9307	0.9131	0.7604	0.9034	0.9799	0.8779
	SAM	0.2552	0.1878	0.2129	0.1407	0.1432	0.1920	0.1209	0.0969	**0.0491**	0.1665
Ours-DCD	PSNR	**32.37**	**45.06**	**51.00**	**35.47**	**48.20**	**48.40**	**38.13**	**43.09**	**47.92**	**46.86**
	SSIM	**0.9389**	**0.9935**	**0.9916**	**0.9403**	**0.9948**	**0.9909**	**0.9619**	**0.9817**	**0.9935**	**0.9909**
	SAM	**0.1449**	**0.0757**	**0.1572**	**0.0789**	**0.0714**	**0.1283**	**0.0859**	**0.0860**	**0.0252**	**0.0841**
Ours-SSCSI	PSNR	**30.55**	**39.48**	**41.93**	**32.67**	**39.66**	**41.12**	**35.25**	**38.14**	**41.90**	**41.03**
	SSIM	**0.9252**	**0.9844**	**0.9874**	**0.9573**	**0.9486**	**0.9512**	**0.9161**	**0.9606**	**0.9900**	**0.9739**
	SAM	**0.1812**	**0.1422**	**0.1843**	**0.1163**	**0.1218**	**0.1860**	**0.1198**	**0.0843**	0.0506	**0.1358**

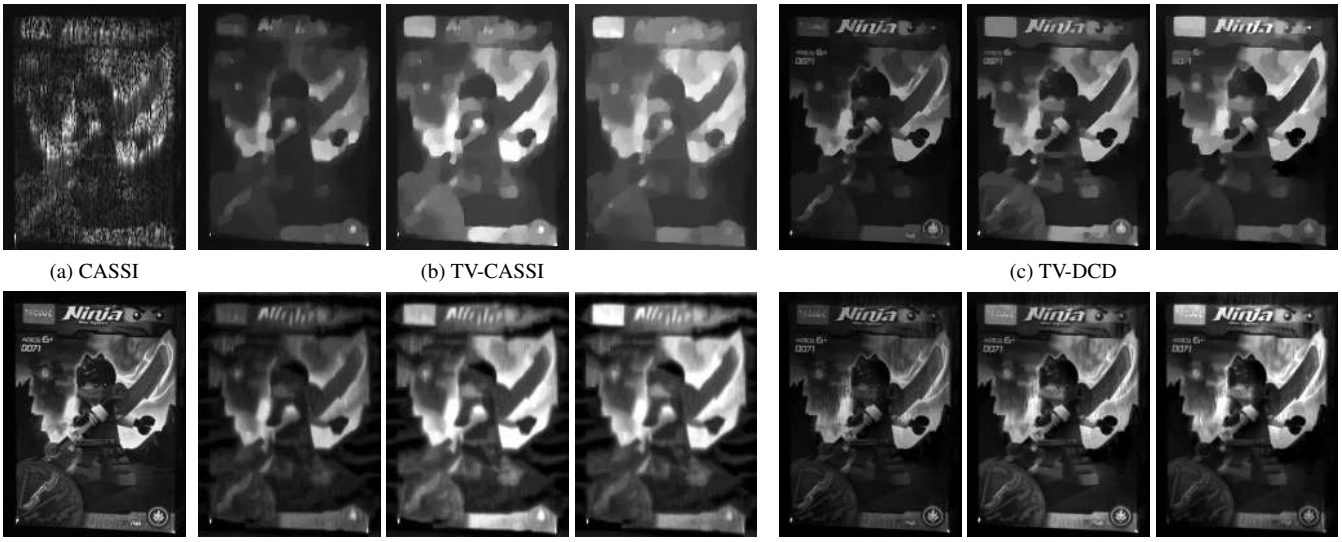
forms TV-SSCSI and DBR-SSCSI.

As for quantitative comparison, we use three image quality metrics to evaluate the performance, including peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [41], and spectral angle mapping (SAM) [42]. PSNR and SSIM are calculated based on each 2D spatial image, which measure the spatial fidelity between the restored HS image and the ground truth. A larger value of these two metrics suggests better restoration. SAM is calculated based on the 1D spectral vector, which measures the spectral fidelity. A smaller value of this metric implies better restoration. All these metrics are averaged across the evaluated dimension.

The quantitative results for these imaging systems and methods for all 10 scenes are shown in Table 1. The best two results for each HS image are highlighted in bold. We first compare the results from different imaging systems under the same restoration method. We can see that SSCSI usually has better performance than CASSI, which demonstrates the advantages of the coding mechanism of SSCSI. The DCD system acquires twice as many measurements as in CASSI and SSCSI, and thus the restoration quality is better.

Compared with the TV and DBR methods, our method provides substantial improvements over all these imaging systems in terms of PSNR, SSIM and SAM. Besides, our method does not need to learn the dictionary, which is obviously affected by different learning methods and datasets [22].

To further analyze the spectral performance of our method, we select a line consisting of 21 consecutive pixels, as labeled in Figures 4(a) and 5(a) (red lines). The absolute differences in the spectral distributions between the ground truth spectrum and the restoration results of all competing methods at the center of the labeled lines are shown in Fig-



(a) CASSI (b) TV-CASSI (c) TV-DCD

(d) Pan (e) Ours-CASSI (f) Ours-DCD

Figure 7. Reconstruction results of 3 spectral bands in 596nm, 619nm and 648nm. (a)CASSI input. (d) Panchromatic input. (b) CASSI recovery [18]. (c) DCD recovery [21]. (e) Ours for CASSI recovery. (f) Ours for DCD recovery.

ure 6 (a) and (c). It is obvious that the recovered spectral distribution from our method is much closer to the ground truth in the same imaging system. In addition, we also show the spectral restoration accuracy of each pixel in the labeled lines by the root mean square error (RMSE) in Figure 6 (b) and (d). It is easy to see that our method obtains the best approximation to the true spectral distributions of the original HS image, which is in accordance with our quantitative evaluation.

5.2. Real Images

We also capture some real images by using the CASSI and DCD systems[3]. A Ninja scene with complex texture is used for test. The coded 2D gray image captured by CASSI is shown in Figure 7(a), while the panchromatic image for DCD is in Figure 7(d). We show the restoration results at 596nm, 619nm and 648nm. By comparing Figure 7(b,e) and (c,f), we can see that DCD performs better than CASSI in terms of the visual quality, which demonstrates again the advantage of the dual setup. As for the restoration methods, our proposed method significantly outperforms TV on both systems. Specifically, the spatial blurring introduced by the TV method in both CASSI and DCD has been dramatically alleviated in our restoration results, e.g., the 'Ninja' letters in the scene. We observe this improvement in the restoration results at other spectral bands, which will be presented in the supplementary material.

[3]Implementation details on these two systems can be found in [18, 21].

6. Conclusion

In this paper, we present an effective method for coded hyperspectral image restoration, which is shown to be generally applicable to several popular coding based imaging systems, without affecting their snapshot advantage. We exploit both sparsity across spectra by a spectral low-rank constraint and structure sparsity in the space via a spatial non-local low-rank regularization. We employ the weighted nuclear norm as a smooth surrogate function for the rank, which can adaptively adjust the regularization parameters. Besides, these two low-rank regularizations are involved into a unified variational framework, which can be efficiently solved by an iterative numerical algorithm.

Our work focuses mainly on exploiting the strong correlation in spectral and spatial domain. [43] showed that temporal correlation can be helpful in accelerating hyperspectral video imaging. To explore correlations among these three factors is left as our future work.

Acknowledgments

The authors would like to thank Lizhi Wang from MSRA for providing the real images of the CASSI and DCD systems.

References

[1] N. Gat, S. Subramanian, J. Barhen, and N. Toomarian, "Spectral imaging applications: remote sensing, environmental monitoring, medicine, military operations, factory automation, and manufacturing," in *Proc. SPIE*, vol. 2962, 1997, pp. 63–77.

[2] R. Ankri, H. Duadi, and D. Fixler, "A new diagnostic

tool based on diffusion reflection measurements of gold nanoparticles," in *Proc. SPIE*, vol. 8225, 2012, pp. 82 250L–82 250L–16.

[3] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of Biomedical Optics*, vol. 19, no. 1, Jan. 2014.

[4] M. Borengasser, W. S. Hungate, and R. Watkins, *Hyperspectral Remote Sensing: Principles and Applications*, ser. Remote Sensing Applications Series. CRC Press, Dec. 2007.

[5] L. Ojha, M. B. Wilhelm, S. L. Murchie, A. S. McEwen, J. J. Wray, J. Hanley, M. Mass, and M. Chojnacki, "Spectral evidence for hydrated salts in recurring slope lineae on Mars," *Nature Geoscience*, vol. advance online publication, Sep. 2015.

[6] A. Banerjee, P. Burlina, and J. Broadwater, "Hyperspectral video for illumination-invariant tracking," in *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Aug. 2009, pp. 1–4.

[7] H. V. Nguyen, A. Banerjee, and R. Chellappa, "Tracking via object reflectance using a hyperspectral video camera," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2010, pp. 44–51.

[8] W. M. Porter and H. T. Enmark, "A System Overview Of The Airborne Visible/Infrared Imaging Spectrometer (Aviris)," in *Proc. of SPIE*, vol. 0834, 1987, pp. 22–31.

[9] R. W. Basedow, D. C. Carmer, and M. E. Anderson, "HYDICE system: implementation and performance," in *Proc. of SPIE*, 1995, pp. 258–267.

[10] M. Yamaguchi, H. Haneishi, H. Fukuda, J. Kishimoto, H. Kanazawa, M. Tsuchida, R. Iwama, and N. Ohyama, "High-fidelity video and still-image communication based on spectral information: natural vision system and its applications," in *Proc. of SPIE*, Jan. 2006, pp. 60 620G–60 620G–12.

[11] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2011, pp. 193–200.

[12] Y. Schechner and S. Nayar, "Generalized mosaicing: wide field of view multispectral imaging," *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, vol. 24, no. 10, pp. 1334–1348, Oct. 2002.

[13] B. Ford, M. Descour, and R. Lynch, "Large-image-format computed tomography imaging spectrometer for fluorescence microscopy," *Optics Express*, vol. 9, no. 9, p. 444, Oct. 2001.

[14] N. Gat, G. Scriven, J. Garman, M. D. Li, and J. Zhang, "Development of four-dimensional imaging spectrometers (4d-IS)," in *Proc. of SPIE Optics + Photonics*, vol. 6302, 2006, pp. 63 020M–63 020M–11.

[15] L. Gao, R. T. Kester, N. Hagen, and T. S. Tkaczyk, "Snapshot Image Mapping Spectrometer (IMS) with high sampling density for hyperspectral microscopy," *Optics Express*, vol. 18, no. 14, p. 14330, Jul. 2010.

[16] X. Cao, H. Du, X. Tong, Q. Dai, and S. Lin, "A Prism-Mask System for Multispectral Video Acquisition," *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, vol. 33, no. 12, pp. 2423–2435, 2011.

[17] 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," *Optics Express*, vol. 15, no. 21, p. 14013, 2007.

[18] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Applied Optics*, vol. 47, no. 10, p. B44, Apr. 2008.

[19] D. Kittle, K. Choi, A. Wagadarikar, and D. J. Brady, "Multiframe image estimation for coded aperture snapshot spectral imagers," *Applied Optics*, vol. 49, no. 36, p. 6824, Dec. 2010.

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

[21] L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu, "Dual-camera design for coded aperture snapshot spectral imaging," *Applied Optics*, vol. 54, no. 4, p. 848, Feb. 2015.

[22] X. Lin, Y. Liu, J. Wu, and Q. Dai, "Spatial-spectral Encoded Compressive Hyperspectral Imaging," *ACM Trans. on Graph. (Proc. of SIGGRAPH Asia)*, vol. 33, no. 6, pp. 233:1–233:11, Nov. 2014.

[23] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral Image Restoration Using Low-Rank Matrix Recovery," *IEEE Trans. Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.

[24] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, Jun. 2005, pp. 60–65.

[25] E. J. Cands and B. Recht, "Exact Matrix Completion via Convex Optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, Apr. 2009.

[26] J. Cai, E. Candes, and Z. Shen, "A Singular Value Thresholding Algorithm for Matrix Completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, Jan. 2010.

[27] M. Fazel, H. Hindi, and S. Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," in *Proc. of the American Control Conference*, vol. 3, Jun. 2003, pp. 2156–2162.

[28] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted Nuclear Norm Minimization with Application to Image Denoising," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2014, pp. 2862–2869.

[29] C. Lu, C. Zhu, C. Xu, S. Yan, and Z. Lin, "Generalized Singular Value Thresholding," in *Proc. of Association for the Advancement of Artificial Intelligence (AAAI)*, Feb. 2015.

[30] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, "Compressive Sensing via Nonlocal Low-Rank Regularization," *IEEE Trans. Image Processing*, vol. 23, no. 8, pp. 3618–3632, Aug. 2014.

[31] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: Robust Alignment by Sparse and Low-Rank Decomposition for Linearly Correlated Images," *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, vol. 34, no. 11, pp. 2233–2246, Nov. 2012.

[32] Z. Zhang, A. Ganesh, X. Liang, and Y. Ma, "TILT: Transform Invariant Low-Rank Textures," *International Journal of Computer Vision (IJCV)*, vol. 99, no. 1, pp. 1–24, Jan. 2012.

[33] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust Principal Component Analysis?" *Journal of the ACM*, vol. 58, no. 3, pp. 11:1–11:37, Jun. 2011.

[34] X. Guo, X. Cao, and Y. Ma, "Robust Separation of Reflection from Multiple Images," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2014, pp. 2195–2202.

[35] D. H. Marimont and B. A. Wandell, "Linear models of surface and illuminant spectra," *Journal of the Optical Society of America A*, vol. 9, no. 11, p. 1905, Nov. 1992.

[36] Y. Qian, Y. Shen, M. Ye, and Q. Wang, "3-d nonlocal means filter with noise estimation for hyperspectral imagery denoising," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2012, pp. 1345–1348.

[37] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[38] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar, "Generalized assorted pixel camera: Post-capture control of resolution, dynamic range and spectrum," Tech. Rep., Nov 2008.

[39] J. Bioucas-Dias and M. Figueiredo, "A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration," *IEEE Trans. Image Processing*, vol. 16, no. 12, pp. 2992–3004, Dec. 2007.

[40] E. van den Berg and M. Friedlander, "Probing the Pareto Frontier for Basis Pursuit Solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, Nov. 2008.

[41] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[42] F. A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz, "The spectral image processing system (SIPS)–interactive visualization and analysis of imaging spectrometer data," *Remote Sensing of Environment*, vol. 44, no. 2-3, pp. 145–163, May 1993.

[43] L. Wang, Z. Xiong, D. Gao, G. Shi, W. Zeng, and F. Wu, "High-speed hyperspectral video acquisition with a dual-camera architecture," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015.