Decomposable Nonlocal Tensor Dictionary Learning for Multispectral Image Denoising

Yi Peng[1], Deyu Meng[1], Zongben Xu[1], Chenqiang Gao[2], Yi Yang[3], Biao Zhang[1]
[1]Xi'an Jiaotong University; [2]Carnegie Mellon University; [3]The University of Queensland
{weichiche}@gmail.com, {dymeng, zbxu}@mail.xjtu.edu.cn,
{gaochenqiang, yee.i.yang, megatronbiao}@gmail.com

Abstract

As compared to the conventional RGB or gray-scale images, multispectral images (MSI) can deliver more faithful representation for real scenes, and enhance the performance of many computer vision tasks. In practice, however, an MSI is always corrupted by various noises. In this paper we propose an effective MSI denoising approach by combinatorially considering two intrinsic characteristics underlying an MSI: the nonlocal similarity over space and the global correlation across spectrum. In specific, by explicitly considering spatial self-similarity of an MSI we construct a nonlocal tensor dictionary learning model with a group-block-sparsity constraint, which makes similar full-band patches (FBP) share the same atoms from the spatial and spectral dictionaries. Furthermore, through exploiting spectral correlation of an MSI and assuming over-redundancy of dictionaries, the constrained nonlocal MSI dictionary learning model can be decomposed into a series of unconstrained low-rank tensor approximation problems, which can be readily solved by off-the-shelf higher order statistics. Experimental results show that our method outperforms all state-of-the-art MSI denoising methods under comprehensive quantitative performance measures.

1. Introduction

The radiance of a real scene is distributed across a wide range of spectral bands. A multispectral image (MSI) consists of multiple intensities that represent the integrals of the radiance captured by sensors over various discrete bands. For example, conventional RGB images are achieved by integrating the product of the intensity at three typical band intervals. As compared with the traditional image system, MSI helps to deliver more faithful representation for real scenes, and has been shown to greatly enhance the performance of various computer vision tasks, such as inpainting [8], superresolution [12] and tracking [21].

Figure 1. (a) A collection of similar local patches over the spatial dimensions of the multispectral image (middle). (b) The highly correlated images obtained across the entire spectral dimension of this multispectral image.

In real cases, however, an MSI is always corrupted by some noises that are generally conducted by equipment limitations like sensor sensitivity, photon effects and calibration error [13, 2]. Besides, since the radiance energy is limited and sometimes the band width is fairly narrow, the energy captured by each sensor might be low. The shot noise and thermal noise then happen inevitably. The denoising problem for MSI is thus still of acute and growing importance [14, 23, 10].

In this paper, we propose a novel tensor dictionary learning model for the task of MSI denoising by combinationally considering two characteristics of MSI into a single framework: nonlocal similarity in space and global correlation in spectrum. On one hand, a typical natural scene contains a collection of similar local patches all over the space, composing of homologous aggregation of microstructures. By averaging among these nonlocally similar patches, the spatial noise is expected to be prominently alleviated [22, 18, 15]. On the other hand, an MSI contains a large amount of spectral redundancy [23]. That is, images obtained over different bands are always highly correlated. Through extracting the major components from these globally correlated spectrum information, the spectral MSI noise (the minor components) is expected to be eliminated. Both characteristics can be easily understood by seeing Fig. 1. In our model, we employ a grouped sparsity regularizer to impose similar MSI patches to share the same dictionary atoms in their sparse decomposition to implicitly average out the

noise among these patches. Furthermore, by assuming redundant dictionaries over both the space and spectrum, the proposed tensor dictionary learning model can be readily decomposed into a series of low-rank tensor approximation problems. Each of these problems corresponds to a spectral dimensionality reduction model conducted by the spectral correlation property of MSIs, and can be easily solved by some off-the-shelf higher order statistics. The spectral redundancy problem can thus be alleviated.

Throughout the paper, we denote scalars, vectors, matrices and tensors by the non-bold letters, bold lower case letters, bold upper case letters and calligraphic upper case letters, respectively.

2. Notions and Preliminaries

We first introduce some necessary notions and preliminaries as follows.

A tensor of order N, which corresponds to an N-dimensional data array, is denoted as $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_n \times \cdots I_N}$. Elements of $\mathcal{A}$ are denoted as $a_{i_1 \cdots i_n \cdots i_N}$, where $1 \leq i_n \leq I_n$. The mode-n vectors of an N^{th} order tensor $\mathcal{A}$ are the I_n dimensional vectors obtained from $\mathcal{A}$ by varying index i_n while keeping the other indices fixed. The matrix $\mathbf{A}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \cdots I_{n-1} I_{n+1} \cdots I_N)}$ is composed by taking the mode-n vectors of $\mathcal{A}$ as its columns. This matrix can also be naturally seen as the mode-n flattening of the tensor $\mathcal{A}$. The n-rank of $\mathcal{A}$, denoted as r_n, is the dimension of the vector space spanned by the mode-n vectors of $\mathcal{A}$.

The product of two matrices can be generalized to the product of a tensor and a matrix. The mode-n product of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_n \times \cdots I_N}$ by a matrix $\mathbf{B} \in \mathbb{R}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{B}$, is also an N^{th} order tensor $\mathcal{C} \in \mathbb{R}^{I_1 \times \cdots \times J_n \times \cdots I_N}$, whose entries are computed by

$$c_{i_1 \cdots i_{n-1} j_n i_{n+1} \cdots i_N} = \sum\nolimits_{i_n} a_{i_1 \cdots i_{n-1} i_n i_{n+1} \cdots i_N} b_{j_n i_n}.$$

The mode-n product $\mathcal{C} = \mathcal{A} \times_n \mathbf{B}$ can also be calculated by the matrix multiplication $\mathbf{C}_{(n)} = \mathbf{B}\mathbf{A}_{(n)}$, followed by a re-tensorization of undoing the mode-n flattening. The Frobenius norm of a tensor $\mathcal{A}$ is defined as: $\|\mathcal{A}\|_F = \left(\sum_{i_1, \cdots, i_N} |a_{i_1 \cdots i_N}|^2\right)^{1/2}$. In the following, we shortly write $\|\mathcal{A}\|_F$ as $\|\mathcal{A}\|$.

3. Related Work

There are mainly two approaches for MSI denoising, including the 2D extended approach and the tensor-based approach.

2D extended approach: As one of the classical problems in computer vision, 2D image denoising has been addressed for more than 50 years and a large amount of researches have been proposed on this problem, such as NLM [4], K-SVD [20] and BM3D [9]. These methods can be directly applied to MSI denoising by treating the images located at different bands separately. This extension, however, neglects the intrinsic properties of MSIs and generally cannot attain good performance in real applications. Another more reasonable extension is specifically designed for the patch-based image denoising methods, which takes the small local patches of the image into consideration. By building small 3D cubes of an MSI instead of 2D patches of a traditional image, the corresponding 3D-cube-based MSI denoising algorithm can then be constructed [22]. The state-of-the-art of 3D-cube-based approach is represented by the BM4D method [15, 16], which exploits the 3D non-local similarity of MSI to remove noise in similar MSI 3D cubes collaborately. These methods, however, have not taken into account the high correlation across MSI spectrum, and thus still have much room for improvement.

Tensor-based approach: An MSI is composed by a stack of 2D images, which can be naturally regarded as a 3^{rd}-order tensor. The tensor-based approach implements the MSI denoising by applying the tensor factorization techniques to the MSI tensor. As a special case of multiway filtering, tensor factorization can be seen as an extension of the traditional singular value decomposition (SVD). The state-of-the-art along this line of research is represented by two approaches. Renard et al. [23] presented a low-rank tensor approximation (LRTA) method by employing the Tucker factorization [24] method to obtain the low-rank approximation of the input MSI. Very recently, Liu et al. [14] designed the PARAFAC method by utilizing the parallel factor analysis [7]. The advantage of both methods is that they took the correlation between MSI images over different bands into consideration, and tried to eliminate the spectral redundancy of MSIs. However, they have not utilized the nonlocal similarity property of MSI, and their performance may be sensitive to noise extents and types.

4. Decomposable Nonlocal MSI Dictionary Learning

In this section, we first introduce the tensor dictionary learning (DL) model, and then present the main idea of our decomposable nonlocal MSI DL model and the related algorithm. The parameter setting problems are also discussed thereafter.

4.1. From Image DL to MSI DL

We first briefly introduce the traditional DL model for image restoration. For a set of image patches (ordered lexicographically as column vectors) $\{\mathbf{x}_i\}_{i=1}^n \subset \mathbb{R}^d$, where d is the dimensionality and n is the number of image patches, DL aims to calculate the dictionary $\mathbf{D} = [\mathbf{d}_1, \cdots, \mathbf{d}_m] \in \mathbb{R}^{d \times m}$, composed by a collection of atoms $\mathbf{d}_i$ ($m > d$, implying that the dictionary is redundant), and the coefficient

matrix $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_n] \in \mathbb{R}^{m\times n}$, composed by the representation coefficients $\mathbf{z}_i$ of $\mathbf{x}_i$, by the following optimization model [1]:

$$\min_{\mathbf{D},\mathbf{z}_1,\cdots,\mathbf{z}_n} \sum\nolimits_{i=1}^{n} \|\mathbf{x}_i - \mathbf{D}\mathbf{z}_i\| \quad s.t. \quad \mathcal{P}(\mathbf{z}_i) \le k, \quad (1)$$

where $\mathcal{P}(\cdot)$ denotes certain sparsity controlling operator such as the l_0 or l_1 norm.

The similar dictionary learning model can be easily extended to MSI cases. First we construct MSI patches like the image case as follows. An MSI with $d_W \times d_H$ spatial resolution (d_W, d_H denote the spatial width and height of the MSI, respectively) and d_S spectral bands can be expressed as a 3^{rd} order tensor $\mathcal{H} \in \mathbb{R}^{d_W\times d_H\times d_S}$ with two spatial modes and one spectral mode. By sweeping all across the MSI with overlaps, we can build a group of 3D full-band patches (FBP) $\{\mathcal{P}_{i,j}\}_{1\le i\le d_W-d_w+1, 1\le j\le d_H-d_h+1} \subset \mathbb{R}^{d_w\times d_h\times d_S}$ ($d_w < d_W, d_h < d_H$) from the MSI. For simplicity, we reformulate all FBPs as a group of 3D patches $\{\mathcal{X}_i\}_{i=1}^n$, where $n = (d_W - d_w + 1)(d_H - d_h + 1)$ denotes the patch number. Each FBP so constructed contains local spatial while global spectral dimensionality, which can easily help us to consider the two important properties underlying an MSI: the nonlocal similarity between spatial patches and the global correlation across all bands.

Based on this FBP set $\{\mathcal{X}_i\}_{i=1}^n$, the MSI DL model can then be constructed to calculate the spatial and spectral dictionaries $\{\mathbf{D}^W \in \mathbb{R}^{d_w\times m_W}, \mathbf{D}^H \in \mathbb{R}^{d_h\times m_H}, \mathbf{D}^S \in \mathbb{R}^{d_S\times m_S}\}$ with $m_W > d_w$, $m_H > d_h$ and $m_S > d_S$, implying the redundancy of these dictionaries, as follows:

$$\begin{aligned} \min_{\mathbf{D}^W,\mathbf{D}^H,\mathbf{D}^S,\mathcal{Z}_i} & \sum_{i=1}^{n} \left\|\mathcal{X}_i - \mathcal{Z}_i \times_1 \mathbf{D}^W \times_2 \mathbf{D}^H \times_3 \mathbf{D}^S\right\| \\ s.t., \quad & \mathcal{P}(\mathcal{Z}_i) \le k \end{aligned}, \quad (2)$$

where $\mathcal{Z}_i \in \mathbb{R}^{m_W\times m_H\times m_S}$ corresponds to the coefficient tensor for $\mathcal{X}_i$ which governs the affiliated interaction between the dictionaries, and $\mathcal{P}(\cdot)$ denotes the sparsity regularization term like l_0 or l_1 operator [31].

4.2. From Image Group-Sparsity to MSI Group-Block-Sparsity

DL has been effectively applied to image denoising by considering the nonlocal similarity property of images [17]. The basic idea is to firstly group the similar patches into clusters $\mathbf{X}^{(k)} = \{\mathbf{x}_{i_j^k}\}_{j=1}^{n_k}$, $k = 1, 2, \cdots, K$, where K is the cluster number, n_k is the patch number in the k^{th} cluster and i_j^k denotes the index of the j^{th} patch in the k^{th} cluster, and then to encourage each cluster share similar atoms in the dictionary. Let's denote the coefficient matrix corresponding to the k^{th} cluster $\mathbf{X}^{(k)}$ as $\mathbf{Z}^{(k)} = [\mathbf{z}_{i_1^k}, \mathbf{z}_{i_2^k}, \cdots, \mathbf{z}_{i_{n_k}^k}] \in \mathbb{R}^{m\times n_k}$, and this simultaneous-sparse-coding aim can then be achieved by applying to (1) the following group-sparsity regularizer on each $\mathbf{Z}^{(k)}$ [17]:

$$\|\mathbf{Z}^{(k)}\|_{p,q} = \sum\nolimits_{i=1}^{m} \|\widehat{\mathbf{z}}_i^k\|_q^p, \quad (3)$$

where $\widehat{\mathbf{z}}_i^k$ denotes the i^{th} row vector of $\mathbf{Z}^{(k)}$. The pair (p, q) is usually set as $(1, 2)$ or $(0, \infty)$. Such group-sparsity regularizer helps to impose some all-zero rows of $\mathbf{Z}^{(k)}$, as depicted in Fig. 2.

This nonlocal method can be easily extended to MSI cases as follows. First, we group the similar FBPs into clusters denoted by $\{\mathcal{X}_{i_j^k}\}_{j=1}^{n_k}$ ($k = 1, 2, \cdots, K$), where K is the cluster number, n_k is the FBP number in the k^{th} cluster and i_j^k denotes the index of the j^{th} patch in the k^{th} cluster. And then we attempt to enforce each cluster share the similar atoms in each of the spatial dictionaries D^W, D^H and spectral dictionary D^S. For convenience we combine the FBP samples in the k^{th} cluster together to formulate a 4^{th} order tensor: $\mathcal{X}^{(k)} \in \mathbb{R}^{d_w\times d_h\times d_S\times n_k}$, whose supplemental 4^{th} mode corresponds to the FBPs located at different spatial positions of the MSI. Analogously, we align all coefficient tensors $\{\mathcal{Z}_{i_j^k}\}_{j=1}^{n_k}$ corresponding to the k^{th} FBP cluster to form $\mathcal{Z}^{(k)} \in \mathbb{R}^{m_W\times m_H\times m_S\times n_k}$. Then the aim of the nonlocal MSI tensor DL can be attained by the following group-block-sparsity regularizer.

Definition 1 *(Group-block-sparsity) For a coefficient tensor $\mathcal{Z} \in \mathbb{R}^{m_W\times m_H\times m_S\times n}$, its group-block-sparsity with respect to the spatial and spectral modes is $\|\mathcal{Z}\|_B = (r^W, r^H, r^S)$ if and only if the smallest index subsets $\boldsymbol{I}^W, \boldsymbol{I}^H, \boldsymbol{I}^S$ satisfying $z_{i_1 i_2 i_3 i_4} = 0$ for all $(i_1, i_2, i_3) \notin \boldsymbol{I}^W \times \boldsymbol{I}^H \times \boldsymbol{I}^S$ contain r^W, r^H, r^S elements, respectively. $Sub(\mathcal{Z}) \in \mathbb{R}^{r^W\times r^H\times r^S\times n}$ denotes the intrinsic sub-tensor of $\mathcal{Z}$ extracted from the entries of the three dimensions of $\mathcal{Z}$ specified by the index sets $\boldsymbol{I}^W, \boldsymbol{I}^H, \boldsymbol{I}^S$, respectively.*

The above definition can be easily understood by seeing Fig. 2. Note that the group-sparsity [17] can be seen as the degenerated case of the group-block-sparsity in 2D images. Furthermore, when we set $n = 1$ (meaning only one FBP in a cluster), the group-block-sparsity so defined exactly corresponds to the concept of block sparsity proposed in [6], which has been substantiated to be capable of enhancing better recovery of the original high order signals since it implicitly incorporates valuable prior information on real signals and facilitates making full use of the dictionary atoms of each mode in signal representation.

Then we can construct the following nonlocal MSI DL model:

$$\begin{aligned} \min_{\mathbf{D}^W,\mathbf{D}^H,\mathbf{D}^S,\mathcal{Z}^{(k)}} & \sum_{k=1}^{K} \left\|\mathcal{X}^{(k)} - \mathcal{Z}^{(k)} \times_1 \mathbf{D}^W \times_2 \mathbf{D}^H \times_3 \mathbf{D}^S\right\| \\ s.t., \quad & \|\mathcal{Z}^{(k)}\|_B \preceq (r_k^W, r_k^H, r_k^S) \end{aligned}, \quad (4)$$

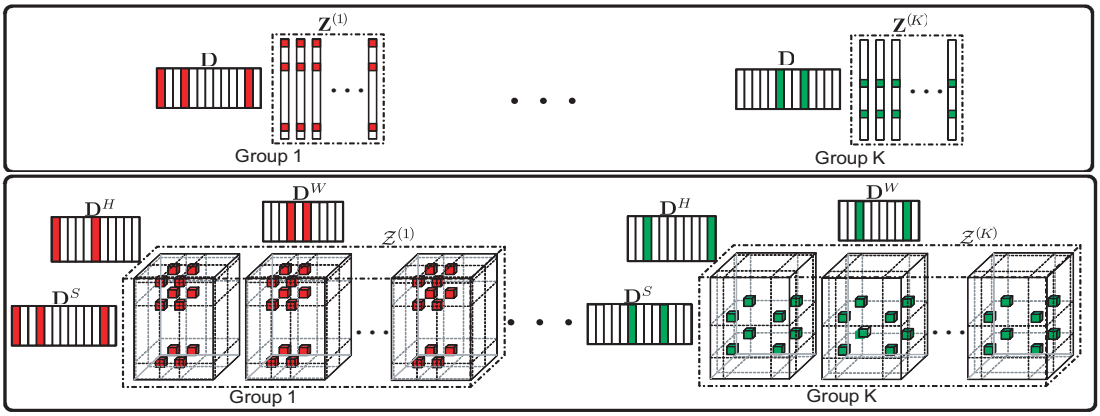


Figure 2. Upper: The image group-sparsity model. In each group the coefficient vectors $\mathbf{Z}^{(k)}$ $(k = 1, \cdots, K)$ share the same atoms of the dictionary $\mathbf{D}$. Lower: The MSI group-block-sparsity model. In each group the coefficient tensors $\mathcal{Z}^{(k)}$ $(k = 1, \cdots, K)$ share the same atoms of the spatial dictionaries $\mathbf{D}^W, \mathbf{D}^H$ and spectral dictionary $\mathbf{D}^S$.

where $\mathbf{v}_1 \preceq \mathbf{v}_2$ denotes that each entry of $\mathbf{v}_1$ is no more than the corresponding entry of $\mathbf{v}_2$. The group-block-sparsity of $\mathcal{Z}^{(k)}$ guarantees that each cluster $\mathcal{X}^{(k)}$ shares r_k^W, r_k^H, r_k^S atoms of the dictionaries $\mathbf{D}^W, \mathbf{D}^H, \mathbf{D}^S$, respectively, and thus the nonlocal similarity among these cluster samples can then be implied.

There are two remaining problems in the construction of the nonlocal MSI DL model (4): how to generate the clusters for FBPs and how to set the group-block-sparsity threshold r_k^W, r_k^H, r_k^S. For the first problem, we just employ the very efficient k-means++ [3] (with automatically and carefully chosen initial seeds) to obtain clusters of all FBPs. The second problem is to be discussed in the next section.

4.3. Decomposable Nonlocal MSI DL Model

The nonlocal MSI DL problem can be further simplified by assuming that the dictionaries $\mathbf{D}^W, \mathbf{D}^H, \mathbf{D}^S$ are redundant enough such that the dictionary atoms utilized in different clusters have no overlap. That is, We assume that the spatial and spectral dictionaries can be reformulated as $\mathbf{D}^W = [\mathbf{D}_1^W, \cdots, \mathbf{D}_K^W], \mathbf{D}^H = [\mathbf{D}_1^H, \cdots, \mathbf{D}_K^H]$ and $\mathbf{D}^S = [\mathbf{D}_1^S, \cdots, \mathbf{D}_K^S]$, where $\mathbf{D}_k^W \in \mathbb{R}^{d_w \times r_k^W}$, $\mathbf{D}_k^H \in \mathbb{R}^{d_h \times r_k^H}$ and $\mathbf{D}_k^S \in \mathbb{R}^{d_S \times r_k^S}$ with $\sum_{k=1}^K r_k^W = m_W$, $\sum_{k=1}^K r_k^H = m_H$ and $\sum_{k=1}^K r_k^S = m_S$, respectively, such that each MSI cluster $\mathcal{X}^{(k)}$ is only related to the sub-dictionaries: $\mathbf{D}_k^W$, $\mathbf{D}_k^H$ and $\mathbf{D}_k^S$. The rationality of this assumption lies on the redundancy setting of the spatial and spectral dictionaries, and even when we suppose that two clusters share an atom of a dictionary, this assumption still holds by easily duplicating this atom in the dictionary. Under this assumption, each element in the sum of Eq. (4) can be equivalently reformulated as:

$$\begin{aligned} &\left\|\mathcal{X}^{(k)} - \mathcal{Z}^{(k)} \times_1 \mathbf{D}^W \times_2 \mathbf{D}^H \times_3 \mathbf{D}^S\right\| \\ &= \left\|\mathcal{X}^{(k)} - Sub(\mathcal{Z}^{(k)}) \times_1 \mathbf{D}_k^W \times_2 \mathbf{D}_k^H \times_3 \mathbf{D}_k^S\right\|, \end{aligned} \quad (5)$$

where $Sub(\mathcal{Z}^{(k)}) \in \mathbb{R}^{r_k^W \times r_k^H \times r_k^S \times n_k}$ denotes the intrinsic sub-tensor of $\mathcal{Z}^{(k)}$, and the original nonlocal MSI DL problem can then be decomposed into a series of problems imposed on all FBP clusters $(k = 1, \cdots, K)$:

$$\min_{\mathbf{D}_k^W, \mathbf{D}_k^H, \mathbf{D}_k^S, \mathcal{Y}} \left\|\mathcal{X}^{(k)} - \mathcal{Y} \times_1 \mathbf{D}_k^W \times_2 \mathbf{D}_k^H \times_3 \mathbf{D}_k^S\right\|. \quad (6)$$

Note that after such transformation, the original problem (4) with constraints is now reformulated into a series of smaller problems without constraints. This makes the problem much easier to solve.

Then the problems are how to solve Eq. (6) and how to set the group-block-sparsity parameters r_k^W, r_k^H, r_k^S. It should be noted that each MSI cluster tensor $\mathcal{X}^{(k)}$ is of a dimensionality redundancy in its 3-rd spectral mode due to one of its important intrinsic properties: global correlation across spectrum. This implies that $\mathcal{X}^{(k)}$ can be approximated by a low-rank tensor obtained by:

$$\min_{\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \mathbf{U}_4, \mathcal{G}} \|\mathcal{X}^{(k)} - \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3 \times_4 \mathbf{U}_4\|, \quad (7)$$

where $\mathbf{U}_1 \in \mathbb{R}^{d_k^W \times r_k^W}$, $\mathbf{U}_2 \in \mathbb{R}^{d_k^H \times r_k^H}$, $\mathbf{U}_3 \in \mathbb{R}^{d_k^S \times r_k^S}$, $\mathbf{U}_4 \in \mathbb{R}^{d_k^N \times r_k^N}$ correspond the basis vectors in the four modes of $\mathcal{X}^{(k)}$ with $d_k^W \geq r_k^W$, $d_k^H \geq r_k^H$, $d_k^S > r_k^S$ and $d_k^N \geq r_k^N$. Here $\mathcal{G} \in \mathbb{R}^{r_k^W \times r_k^H \times r_k^S \times r_k^N}$ is the so-called core tensor [24] and $r_k^S < d_k^S$ leads to the dimensionality reduction in the spectral mode of $\mathcal{X}^{(k)}$. Eq. (7) can be readily solved by the Tucker decomposition technique [24], and the solution of Eq. (6) can then be easily obtained by letting $\mathbf{D}_k^W = \mathbf{U}_1$, $\mathbf{D}_k^H = \mathbf{U}_2$, $\mathbf{D}_k^S = \mathbf{U}_3$ and $\mathcal{Y} = \mathcal{G} \times_4 \mathbf{U}_4$.

As for the selection of the rank parameters (i.e., the group-block-sparsity thresholds) r_k^W, r_k^H, r_k^S and r_k^N in Eq. (7), we can easily adopt the well known AIC/MDL method [27] on the mode-i $(i = 1, 2, 3, 4)$ flattening $\mathbf{X}_{(i)}^{(k)}$ of each cluster tensor $\mathcal{X}^{(k)}$. Such a simple method is substantiated to be effective throughout all our experiments.

4.4. Decomposable Nonlocal MSI DL Algorithm

Based on the aforementioned process, the decomposable nonlocal MSI DL algorithm can be summarized as Algo-

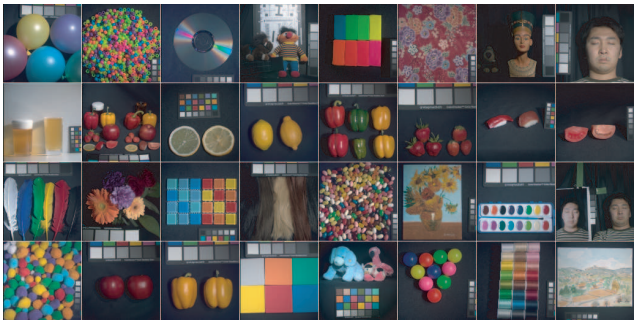


Figure 3. Simulated RGB images using Columbia Multispectral Image Database.

rithm 1. We can then utilize $\mathcal{Z}^{(k)}$, $\mathbf{D}^W$, $\mathbf{D}^H$, $\mathbf{D}^S$ outputted from the proposed algorithm to recover all overlapping F-BPs and average the results to obtain the denoised MSI. It should be noted that all of the utilized k-means++ [3] (step 2), AIC/DIC [27] (step 3) and Tucker factorization [24](step 4) techniques can be fastly implemented, which guarantees the efficiency of our algorithm in practice.

Algorithm 1: Decomposable Nonlocal MSI DL

Input: Input MSI $\mathcal{H} \in \mathbb{R}^{d_W \times d_H \times d_S}$

Output: Spatial dictionaries $\mathbf{D}^W = [\mathbf{D}_1^W, \cdots, \mathbf{D}_K^W]$, $\mathbf{D}^H = [\mathbf{D}_1^H, \cdots, \mathbf{D}_K^H]$, spectral dictionary $\mathbf{D}^S = [\mathbf{D}_1^S, \cdots, \mathbf{D}_K^S]$ and coefficient tensors $\mathcal{Z}^{(k)}$, $k = 1, \cdots, K$

1 Construct the entire FBP set of $\mathcal{H}$ (Section 4.1).

2 Group all FBPs into cluster tensors $\mathcal{X}^{(k)} \in \mathbb{R}^{d_w \times d_h \times d_S \times n_k}$, $k = 1, \cdots, K$ by k-means++ (Section 4.2).

3 Calculate the rank parameters r_k^W, r_k^H, r_k^S and r_k^N by applying the AIC/MDL method to $\mathbf{X}_{(1)}^{(k)}$,$\mathbf{X}_{(2)}^{(k)}$, $\mathbf{X}_{(3)}^{(k)}$ and $\mathbf{X}_{(4)}^{(k)}$, respectively (Section 4.3).

4 Implement the Tensor factorization technique on $\mathcal{X}^{(k)}$ by Eq. (7) to obtain $\mathbf{U}_1$, $\mathbf{U}_2$, $\mathbf{U}_3$, $\mathbf{U}_4$ and $\mathcal{G}$, and let $\mathbf{D}_k^W = \mathbf{U}_1$, $\mathbf{D}_k^H = \mathbf{U}_2$, $\mathbf{D}_k^S = \mathbf{U}_3$ and $Sub(\mathcal{Z}^{(k)}) = \mathcal{G} \times_4 \mathbf{U}_4$. Reformulate the sub-tensor $Sub(\mathcal{Z}^{(k)})$ to obtain $\mathcal{Z}^{(k)}$.

5. Experimental Results

Columbia Datasets: We utilized the Columbia Multispectral Image Database [28][1] to test the proposed method. This dataset contains 32 real-world scenes, each with spatial resolution 512×512 and spectral resolution 31 which includes full spectral resolution reflectance data collected from 400nm to 700nm in 10nm steps. This MSI dataset is of a wide variety of real-world materials and objects, see Fig. 3. Each of these MSIs are scaled into the interval $[0, 1]$ in our experiments.

Noise models: In the experiments we used two types of noises commonly existed in real MSIs. One is the additive white Gaussian noise (AWGN), which comes from many natural sources, such as the spontaneous thermal generation of electrons. And the other is the Poisson noise (also known as shot noise) which is originating from the mechanism of quantized photons and uniform exposure [5]. We parameterized AWGN by its standard deviation σ and Poisson noise by the variance $\mathcal{H}/2^\kappa$ where $\mathcal{H}$ is the noise-free signal. We designed two series of experiments. In the first group, we perturbed each of the 32 Columbia MSI with Gaussian noises of different σ (up to 0.3) and Poisson noise with fixed $\kappa = 5$. In the second case, we used κ from 2 to 6 and fixed $\sigma = 0.1$.

To remove the dependency of the noise variance on the underlying signal before the denoising and compensate the effects of the bias in the produced filtered estimate, in all experiments, the noisy MSI was firstly reformulated by a variance-stabilizing transformation (VST) [19] before implementing a denoising method, and after denoising, a corresponding inverse transformation was used to obtain the final MSI reconstruction.

Implementing details: Like most of the denoising methods based on non-local similarity such as BM3D and BM4D, we employ a preprocessing before the clustering step of our algorithm (Step 2). Our experiments show that a simple band-wise low-pass filtering is capable of greatly improving the accuracy of matching and facilitating the effectiveness of the following steps of our proposed denoising framework. It should be noted that the FBP width d_w and height d_h are the only two parameters needed to be set in our algorithm (all of the other parameters including K, r_k^W, r_k^H, r_k^S and r_k^N can be automatically selected). In all our experiments, we just simply set them as $d_w = d_h = 8$.

Comparison methods: The comparison methods include: band-wise K-SVD [1][2] and band-wise BM3D [9][3], state-of-the-art for the 2D extended band-wise approach; 3D-cube K-SVD [1][2] , ANLM3D [18][4] and BM4D [16][3], state-of-the-art for the 2D extended 3D-cube-based approach; LRTA [23] and PARAFAC [14], state-of-the-art for the tensor-based approach[5]. All parameters involved in the competing algorithms were optimally assigned or automatically chosen as described in the reference papers.

Evaluation measures: To comprehensively assess the performance of all competing algorithms, we employ five quantitative picture quality indices (PQI) for performance evaluation, including peak signal-to-noise ratio (PSNR), structure similarity (SSIM [26]), feature similarity (FSIM [30]), erreur relative globale adimensionnelle de synthèse (ERGAS [25]) and spectral angle map-

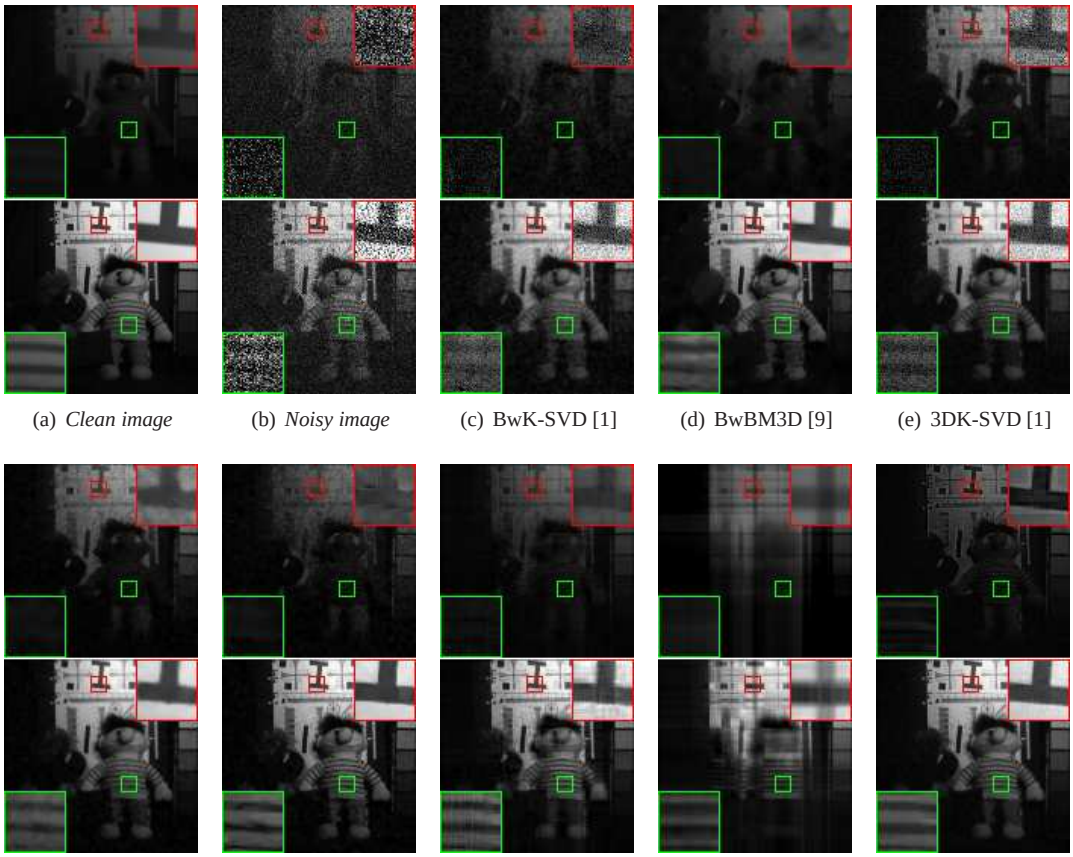
[1] http://www1.cs.columbia.edu/CAVE/databases/multispectral

[2] http://www.cs.technion.ac.il/~elad/software

[3] http://www.cs.tut.fi/ foi/GCF-BM3D

[4] http://personales.upv.es/jmanjon/denoising/arnlm.html

[5] http://www.sandia.gov/ tgkolda/TensorToolbox/index-2.5.html



(a) *Clean image* (b) *Noisy image* (c) BwK-SVD [1] (d) BwBM3D [9] (e) 3DK-SVD [1]

(f) ANLM3D [18] (g) BM4D [16] (h) LRTA [23] (i) PARAFAC [14] (j) *Ours*

Figure 4. (a) The images at two bands ($400\mathrm{nm}$ and $700\mathrm{nm}$) of *chart and stuffed toy*; (b) The corresponding images corrupted by the mixture of $\sigma = 0.2$ Gaussian noise and $\kappa = 5$ Poisson noise; (c)-(j) The restored images obtained by the 8 utilized MSI denoising methods. Two demarcated areas in each image are amplified at a 4 times larger scale for easy observation of details.

per (SAM [29]). PSNR and SSIM are two conventional PQIs in image processing and computer vision. They evaluate the similarity between the target image and the reference image based on MSE and structural consistency, respectively. FSIM emphasizes the perceptual consistency with the reference image. The larger these three measures are, the closer the target MSI is to the reference one. ERGAS and SAM are usually appear corporately in the literature since they extract complementary information from an MSI. ERGAS measures fidelity of the restored image based on the weighted sum of MSE in each band and SAM calculates the average angle between spectrum vectors of the target MSI and the reference one across all spatial positions. Different from the former three measures, the smaller these two measures are, the better does the target MSI estimate the reference one. Note that SAM fully reflects the fidelity of the spectral reflectance of the target MSI.

Performance evaluation: For each noise setting, all of the five PQI values for each competing MSI denoising methods on all 32 scenes have been calculated and recorded. Table 1 lists the average performance (over different scenes and noise settings) of all methods in Poisson/Gaussian mixture noise case. More details are listed in our supplementary material. It can be easily observed that the proposed method outperforms all other competing methods. As a detailed comparison, our method performs unsubstantially worse than BM4D with respect to only two PQI measures (PSNR and ERGAS) at a part of noise levels (say $\sigma \leq 0.2$ with fixed κ and $\kappa \geq 5$ with fixed σ. Please see supplementary material). And in average, our method performs best with respect to all PQIs.

To further depict the denoising performance of our method, we depict in Fig. 4 two bands in *chart and stuffed toy* that centered at $400\mathrm{nm}$ (the darker one) and $700\mathrm{nm}$ (the brighter one), respectively. Two demarcated areas in the scene have been amplified in the figure at a 4 times larger scale for easy observation of details. It is easy to observe that our method obtains a better recovery for both small-scale textures and large-scale structures, especially when the band energy is low (see the dark channel).

Based on the SAM measures in Table 1, our method is substantiated to be able to best recover the spectral reflectance of the MSIs as compared to other competing methods. To further clarify this point, we demonstrate in Fig. 5

$\sigma = 0.1, 0.15, 0.2, 0.25, 0.3$, $\kappa = 5$					
	PSNR	SSIM	FSIM	ERGAS	SAM
Noisy image	14.52 ± 0.04	0.060 ± 0.032	0.469 ± 0.111	1156.9 ± 327.0	1.133 ± 0.194
BwK-SVD [1]	25.77 ± 1.12	0.370 ± 0.027	0.792 ± 0.038	296.5 ± 58.4	0.614 ± 0.163
BwBM3D [9]	33.45 ± 2.58	0.828 ± 0.048	0.910 ± 0.015	124.4 ± 34.0	0.340 ± 0.128
3DK-SVD [1]	28.07 ± 1.29	0.497 ± 0.024	0.859 ± 0.025	232.5 ± 42.5	0.581 ± 0.170
ANLM3D [18]	33.61 ± 2.24	0.836 ± 0.036	0.916 ± 0.020	121.1 ± 27.2	0.365 ± 0.142
BM4D [16]	$\mathbf{36.25 \pm 2.16}$	0.885 ± 0.027	0.938 ± 0.013	90.0 ± 19.0	0.314 ± 0.131
LRTA [23]	33.90 ± 2.74	0.850 ± 0.064	0.925 ± 0.017	118.4 ± 35.8	0.234 ± 0.082
PARAFAC [14]	27.66 ± 2.93	0.747 ± 0.100	0.862 ± 0.053	243.3 ± 70.3	0.388 ± 0.117
Ours	$\mathbf{36.25 \pm 2.56}$	$\mathbf{0.914 \pm 0.031}$	$\mathbf{0.952 \pm 0.008}$	$\mathbf{88.9 \pm 21.6}$	$\mathbf{0.182 \pm 0.070}$
$\sigma = 0.1$, $\kappa = 2, 3, 4, 5, 6$					
	PSNR	SSIM	FSIM	ERGAS	SAM
Noisy image	17.45 ± 1.02	0.088 ± 0.035	0.573 ± 0.094	756.3 ± 129.4	1.055 ± 0.194
BwK-SVD	27.81 ± 1.85	0.528 ± 0.051	0.849 ± 0.031	224.4 ± 32.9	0.514 ± 0.145
BwBM3D	34.14 ± 2.80	0.864 ± 0.046	0.926 ± 0.014	113.5 ± 32.4	0.267 ± 0.098
3DK-SVD	29.95 ± 1.94	0.662 ± 0.048	0.901 ± 0.021	180.3 ± 28.7	0.486 ± 0.151
ANLM3D	31.53 ± 1.53	0.686 ± 0.030	0.874 ± 0.035	155.8 ± 23.5	0.439 ± 0.145
BM4D	36.93 ± 2.36	0.910 ± 0.025	0.951 ± 0.011	81.7 ± 17.1	0.259 ± 0.108
LRTA	33.31 ± 3.20	0.812 ± 0.075	0.916 ± 0.022	140.7 ± 49.6	0.271 ± 0.100
PARAFAC	30.27 ± 3.13	0.802 ± 0.084	0.896 ± 0.038	177.9 ± 60.9	0.323 ± 0.113
Ours	$\mathbf{36.94 \pm 2.75}$	$\mathbf{0.930 \pm 0.029}$	$\mathbf{0.963 \pm 0.007}$	$\mathbf{81.5 \pm 20.5}$	$\mathbf{0.150 \pm 0.052}$

Table 1. Average performance comparison of 8 competing methods with respect to 5 PQIs mixture noise experiments. For both settings, the results are obtained by averaging through the 32 scenes and the varied parameters.

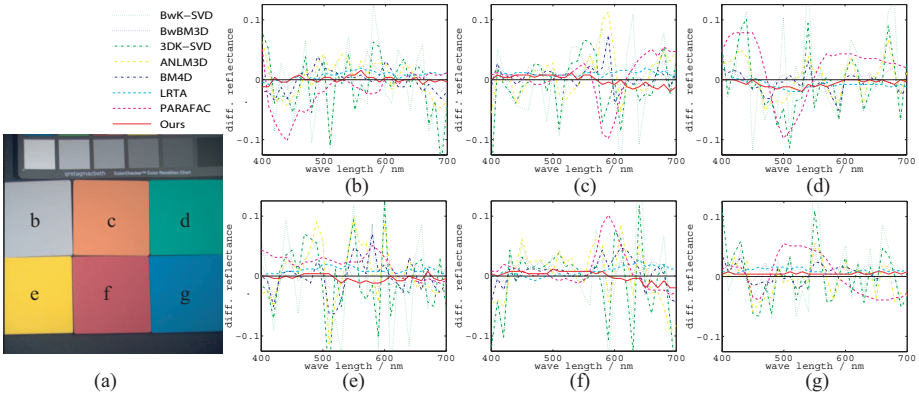


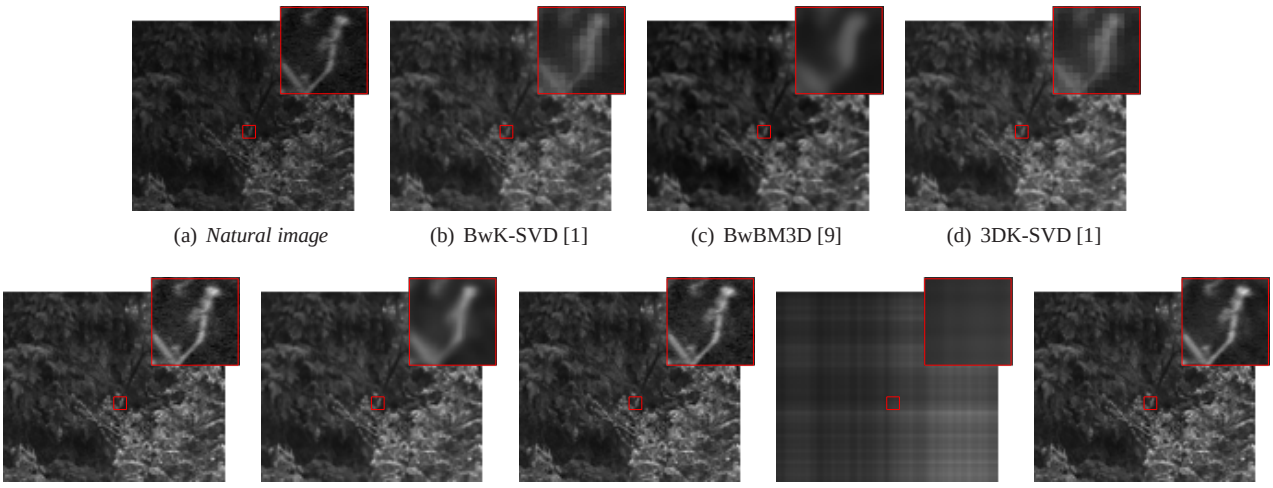
Figure 5. (a) Simulated RGB image of *sponges*. (b)-(g) Spectral reflectance difference curves of 8 competing methods at 6 locations of *sponges*. The noise-free image is corrupted by the mixture of $\sigma = 0.2$ Gaussian noise and $\kappa = 5$ Poisson noise.

the spectral reflectance difference curves of all competing methods at 6 locations in *sponges*. A spectral reflectance difference curve of a MSI denoising method at a spatial location is attained by sequentially interpolating the 31 elements of the deviation between the restored and the clean MSI along their spectral mode. It is easy to see that our method obtains the best approximation of the intrinsic spectral patterns of the original MSI, which fully complies with our quantitative evaluation.

MSI Denoising performance on natural scenes: We also used some MSIs from real-world scenes [11][6] to test our denoising method. This dataset comprises 15 rural scenes (containing rocks, trees, leaves, grass, earth, *etc*.) and 15 urban scenes (containing walls, roofs, windows, plants, indoor, *etc*.). All of them are illuminated by the direct sunlight between mid-morning to mid-afternoon. As the images are taken from a fairly far distance and the energy is spread over all bands, these MSIs contain certain degree of noises. We employed the similar implementation strategies and parameter settings for our method and compared the similar competing methods as the first series of experiments.

Our experimental results show that our method can generally ameliorate the image quality contained in these MSIs. For easy observation we illustrates an example image located at a band of a rural MSI in Fig. 6. An area of interest is amplified in the restored image obtained by all competing methods. It can be easily observed that the image restored by our method properly removes the noise while finely preserves the structure underlying the image, while the results obtained by most of other competing methods contain evident significant blurry area as compared to the original image. Among these methods, ANLM3D and LRTA perform comparatively better in structure preserving. However, the images recovered by them remain more unexpected sharp noises than that obtained by our method.

[6] http://personalpages.manchester.ac.uk/staff/david.foster/Hyperspectral_images_of_natural_scenes_02



(a) *Natural image* (b) BwK-SVD [1] (c) BwBM3D [9] (d) 3DK-SVD [1]

(e) ANLM3D [18] (f) BM4D [16] (g) LRTA [23] (h) PARAFAC [14] (i) *Ours*

Figure 6. (a) The image located at the 12^{th} band in scene 2 of the natural scene dataset. (b)-(i) The restored images obtained by the 8 utilized MSI denoising methods. The demarcated area in each image is amplified for easy observation of details. The figure is better seen by zooming on a computer screen.

Acknowledgement

This research was supported by 973 Program of China with No. 3202013CB329404 and the NSFC projects with No. 61373114, 11131006, 6107505.

References

[1] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Processing*, 54(11):4311–4323, 2006.

[2] B. Aiazzi, L. Alparone, A. Barducci, S. Baronti, and I. Pippi. Information theoretic assessment of sampled hyperspectral imagers. *IEEE Trans. Geoscience and Remote Sensing*, 39(7):1447–1458, 2001.

[3] D. Arthur and S. Vassilvitskii. K-means++: the advantages of careful seeding. *Society for Industrial and Applied Mathematics*.

[4] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. In *ICVPR*, 2005.

[5] P. D. Burns. *Analysis of Image Noise in Multispectral Color Aquisition.* PhD thesis, Center for Imaging Science, Rochester Institute of Technology, 1997.

[6] C. F. Caiafa and A. Cichocki. Computing sparse representations of multidimensional signals using kronecker bases. *Neural Computation*, 25(1):186–220, 2013.

[7] J. D. Carroll and J. J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of eckart-young decomposition. *Psychometrika*, 35(3):283–319, 1970.

[8] A. Chen. The inpainting of hyperspectral images: a survey and adaptation to hyperspectral data. In *SPIE*, 2012.

[9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3d transform-domain collaborative filtering. *IEEE Trans. Image Processing*, 16(8):2080–2094, 2007.

[10] L. De Lathauwer, B. De Moor, and J. Vandewalle. On the best rank-1 and rank-$(r_1, r_2, \cdots, r_n)$ approximation of higher-order tensors. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1324–1342, 2000.

[11] D. H. Foster, K. Amano, S. M. C. Nascimento, and M. J. Foster. Frequency of metamerism in natural scenes. *Journal of the Optical Society of America A*, 23(10):2359–2372, 2006.

[12] R. Kawakami, J. Wright, Y. W. Tai, Y. Matsushita, and M. K. Ikeuchi. High-resolution hyperspectral imaging via matrix factorization. In *CVPR*, 2011.

[13] J. Kerekes and J. Baum. Full-spectrum spectral imaging system analytical model. *IEEE Trans. Geoscience and Remote Sensing*, 5(2):571–580, 2005.

[14] X. F. Liu, S. Bourennane, and C. Fossati. Denoising of hyperspectral images using the parafac model and statistical performance analysis. *IEEE Trans. Geoscience and Remote Sensing*, 50(10):3717–3724, 2012.

[15] M. Maggioni and A. Foi. Nonlocal transform-domain denoising of volumetric data with groupwise adaptive variance estimation. In *SPIE*, 2012.

[16] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi. A nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Trans. Image Process.*, 22(1):119–133, 2013.

[17] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *ICCV*, 2009.

[18] J. V. Manjón, P. Coupé, L. Martí-Bonmatí, L. Collins, and M. Robles. Adaptive non-local means denoising of mr images with spatially varying noise levels. *Journal of Magnetic Resonance Imaging*, 31(1):192–203, 2010.

[19] M. Mäkitalo and A. Foi. Optimal inversion of the generalized anscombe transformation for poisson-gaussian noise. *IEEE Trans. Image Process.*, 22(1):91–103, 2013.

[20] E. Michael and A. Michal. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Processing*, 15(12):3736–3745, 2006.

[21] H. V. Nguyen, A. Banerjee, and R. Chellappa. Tracking via object reflectance using a hyperspectral video camera. In *CVPR Workshop*, 2010.

[22] Y. T. Qian, Y. H. Shen, M. C. Ye, and Q. Wang. 3-d nonlocal means filter with noise estimation for hyperspectral imagery denoising. In *IGARSS*, 2012.

[23] N. Renard, S. Bourennane, and J. Blanc-Talon. Denoising and dimensionality reduction using multilinear tools for hyperspectral images. *IEEE Trans. Geoscience and Remote Sensing*, 5(2):138–142, 2008.

[24] L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1996.

[25] L. Wald. *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*. Presses de l'Ecole des Mines, 2002.

[26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[27] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Transaction on Acoustics, Speech and Signal Processing*, 33(2):387–392, 1985.

[28] F. Yasuma, T. Mitsunaga, and D. Iso. Generalized assorted pixel camera: post-capture control of resolution, dynamic range and spectrum. *IEEE Transaction on Image Processing*, 19(9):2241–2253, 2010.

[29] R. H. Yuhas, J. W. Boardman, and A. F. H. Goetz. Determination of semi-arid landscape endmembers and seasonal trends using convex geometry spectral unmixing techniques. In *Summaries of the* 4^{th} *Annual JPL Airborne Geoscience Workshop*, 1993.

[30] L. Zhang, L. Zhang, X. Q. Mou, and D. Zhang. Fsim: a feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, 2011.

[31] S. Zubair, W. Dai, and W. Wang. Sparseness constrained tensor factorization algorithm for dictionary learning over high-dimensional space. In *IMA*, 2012.