Hyperspectral image super-resolution via non-local sparse tensor factorization

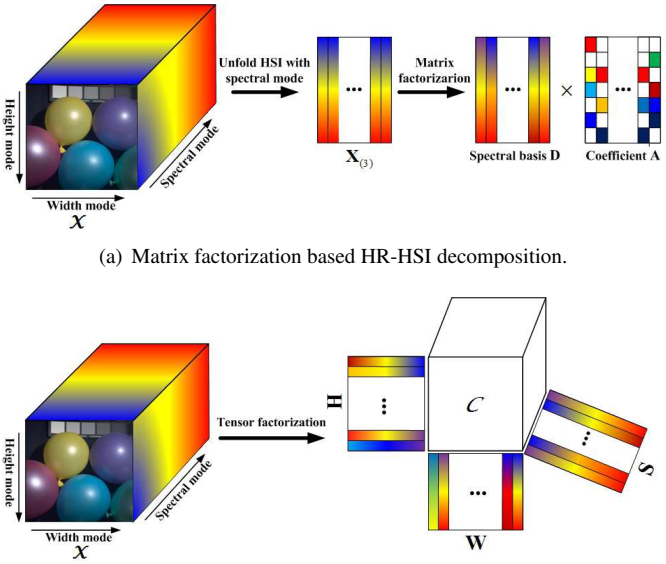
Renwei Dian, Leyuan Fang, Shutao Li
College of Electrical and Information Engineering, Hunan University
drw@hnu.edu.cn, fangleyuan@gmail.com, shutao_li@hnu.edu.cn

Abstract

Hyperspectral image (HSI) super-resolution, which fuses a low-resolution (LR) HSI with a high-resolution (HR) multispectral image (MSI), has recently attracted much attention. Most of the current HSI super-resolution approaches are based on matrix factorization, which unfolds the three-dimensional HSI as a matrix before processing. In general, the matrix data representation obtained after the matrix unfolding operation makes it hard to fully exploit the inherent HSI spatial-spectral structures. In this paper, a novel HSI super-resolution method based on non-local sparse tensor factorization (called as the NLSTF) is proposed. The sparse tensor factorization can directly decompose each cube of the HSI as a sparse core tensor and dictionaries of three modes, which reformulates the HSI super-resolution problem as the estimation of sparse core tensor and dictionaries for each cube. To further exploit the non-local spatial self-similarities of the HSI, similar cubes are grouped together, and they are assumed to share the same dictionaries. The dictionaries are learned from the LR-HSI and HR-MSI for each group, and corresponding sparse core tensors are estimated by spare coding on the learned dictionaries for each cube. Experimental results demonstrate the superiority of the proposed NLSTF approach over several state-of-the-art HSI super-resolution approaches.

1. Introduction

Hyperspectral imaging has been recently applied in many computer vision tasks, including the tracking [19], face recognition[20] and segmentation[28]. However, hyperspectral images (HSIs) usually have abundant spectral information, but limited spatial resolution due to hardware restrictions [13]. On the contrary, the high-resolution (HR) gray images and multispectral images (MSIs) with much less spectral bands can be easily obtained by current imaging sensors. To enhance the spatial resolution of the HSI, the low-resolution (LR) HSIs are generally fused with these HR images. The traditional spatial-spectral image fusion methods focus on combining the LR-HSI with a HR panchromatic images (gray image), which is called as pan-sharpening [5]. Representative methods in pan-sharpening include the Intensity-Hue-Saturation (IHS) transform [22], PCA-based method [23], and compressed sensing based method [16]. Since the single band pan-sharpening has very limited spectral resolution, the reconstructed HR-HSIs by these approaches usually contain spectral distortions.



(a) Matrix factorization based HR-HSI decomposition.

(b) Tensor factorization based HR-HSI decomposition.

Figure 1. Illustration of the traditional matrix decomposition and tensor decomposition of the HR-HSI.

More recently, the HSI super-resolution approaches which fuse a LR-HSI with a HR-MSI (often RGB image) based on matrix factorization have been actively investigated [13, 34, 38, 15, 2, 24, 30, 9, 33, 3]. Assuming that a typical scene of the HSI contains only a small number of pure spectral signatures, these approaches first unfold HSI as a matrix, and then decompose the matrix as spectral basis and corresponding coefficients, as shown in Fig. 1(a). The problem of the HSI super-resolution becomes the estimation of spectral basis and corresponding coefficients from the LR-HSI and the HR-MSI of the same scene. In specific,

Kawakami *et al.* [13] firstly introduce matrix factorization into the spatial-spectral fusion by decomposing the HR-HSI on the learned dictionary with a sparse prior. By incorporating a non-negativity constraint to the spectral basis and the coefficients, Wycoff *et al.* [34] use the framework of alternating direction method of multipliers (ADMM) to acquire spectral basis and corresponding coefficients. Instead of estimating spectral basis in advance and keeping it fixed, non-negative coupled matrix factorization methods [38, 15] are utilized to unmix both the LR-HSI and HR-MSI simultaneously. In addition to the consideration of the spectral information of the HSI, some approaches [24, 30, 9, 33] also use the spatial structures of the HSI to solve the HSI super-resolution problem. For example, Akhtar *et al.* [2] acquire coefficients of the HR-HSI with the simultaneous greedy pursuit algorithm for each local patch, which exploits the prior that nearby pixels are likely to represent the same materials in the HR images. Similarly, Veganzones *et al.* [30] emphasize that the HSIs are often locally low rank. They learn the spectral basis and conduct sparse coding process independently for each local patch. Furthermore, Dong *et al.* [9] propose a clustering-based structure sparse coding method to utilize the non-local spatial self-similarities of the HSI. In addition, Simoes *et al.* [24] use a total variation regularizer to favour spatial smoothness of the solution. By imposing priors on the distribution of the image intensities, Bayesian approaches [33, 3] apply MAP inference to regularize the fusion problem. These matrix factorization based methods start by unfolding the three-dimensional data structures into matrices. Although, the information presented in the two representations is the same, the methods operating with matrices makes it hard to fully exploit the inherent HSI spatial-spectral correlations.

In the past years, tensor factorization has been successfully applied into multiframe data denoising [10, 21], completion [41, 17, 40], compressive sensing [36] and classification [35]. As one of the most effective tensor decomposition methods, Tucker decomposition method [29] decomposes a tensor as a core tensor multiplied by factor matrix along each mode. On the other hand, a typical natural scene usually contains a collection of similar patches from all over the image. These non-local similar patches are often clustered together before processing, which can be exploited to enhance the performance of image denoising [39] and demosaicking [18].

Inspired by the above works, a novel non-local sparse tensor factorization (NLSTF) based HSI super-resolution approach is proposed for the fusion of a LR-HSI and a HR-MSI. In the proposed NLSTF method, the non-local means approach and sparse tensor factorization are unified into one framework, which modifies the HSI super-resolution problem as the estimation of the dictionaries of three modes and corresponding core tensor for each cube of the HR-HSI. Each cube of the HR-HSI contains the local spatial-spectral information. In order to better model the local spatial-spectral information, each cube of the HR-HSI is decomposed as a core tensor and factor matrixes (also called dictionaries) of three modes, as shown in Fig. 1(b). In the decomposition, dictionaries of the width mode and height mode represent spatial information of the HSI, and dictionary of spectral mode represents spectral information. Meanwhile, the core tensor models the relationship of the dictionaries of three modes. In this framework, spatial-spectral correlations of the HSI can be better used since the information of three modes is incorporated into an unified model. In addition, to exploit the non-local self-similarities of the HR-HSI, we group similar cubes of the HR-HSI together. Furthermore, a grouped sparsity regularizer is exploited to impose similar cubes to share the same dictionaries in their sparse tensor decompositions.

The main contributions of this paper include: (1) The tensor factorization is introduced to fuse the LR-HSI with HR-MSI. In this way, the problem of HSI super-resolution is reformulated as the estimation of dictionaries in three modes and corresponding core tensors, which incorporates the spatial-spectral information into an unified framework. (2) Non-local spatial self-similarities of the HSI are incorporated into the tensor factorization.

2. Preliminaries on Tensors

An N-dimensional tensor is denoted as $\mathcal{M} \in R^{I_1 \times I_2, \ldots, \times I_N}$. Elements of $\mathcal{M}$ are denoted as $m_{i_1 i_2, \ldots, i_N}$, where $1 \leq i_n \leq I_n$. The n-mode unfolding vectors of tensor $\mathcal{M}$ are the I_n-dimensional vectors obtained from $\mathcal{M}$ by changing index i_n, while keeping the other indices fixed. The n-mode unfolding matrix $\mathbf{M}_{(n)} \in R^{I_n \times I_1 I_2, \ldots, I_{n-1} I_{n+1}, \ldots, I_N}$ is defined by arranging all the n-mode vectors as the columns of the matrix [14]. The product of two matrices can be generalized to the multiplication of a tensor and a matrix. The n-mode product of the tensor $\mathcal{M} \in R^{I_1 \times I_2 \ldots \times I_N}$ with the matrix $\mathbf{B} \in R^{J_n \times I_n}$, denoted by $\mathcal{M} \times_n \mathbf{B}$, is an N-dimensional tensor $\mathcal{C} \in R^{I_1 \times I_2 \ldots \times J_n \ldots \times I_N}$, whose elements are computed by

$$c_{i_1 \ldots i_{n-1} j_n i_{n+1} \ldots i_N} = \sum_{i_n} m_{i_1 \ldots i_{n-1} i_n i_{n+1} \ldots i_N} b_{j_n i_n}, \quad (1)$$

The n-mode product $\mathcal{M} \times_n \mathbf{B}$ can also be computed by matrix multiplication $\mathbf{C}_{(n)} = \mathbf{B}\mathbf{M}_{(n)}$. For distinct modes in a series of multiplications, the order of the multiplications is irrelevant, which is

$$\mathcal{M} \times_m \mathbf{A} \times_n \mathbf{B} = \mathcal{M} \times_n \mathbf{B} \times_m \mathbf{A} (n \neq m). \quad (2)$$

If the modes of multiplications are the same, the equation (2) is transformed into

$$\mathcal{M} \times_n \mathbf{A} \times_n \mathbf{B} = \mathcal{M} \times_n (\mathbf{B}\mathbf{A}). \quad (3)$$

Besides, the relationship between the Tucker mode and a Kronecker product is specified by Caiafa and Cichocki [7]. Given the n-mode dictionaries $\mathbf{D}_n \in R^{J_n \times I_n} (n = 1, 2, ..., N)$, $\mathbf{c} = vec(\mathcal{C}) \in R^{J \times 1} (J = \prod_{n=1}^{N} J_n)$, and $\mathbf{m} = vec(\mathcal{M}) \in R^{I \times 1} (I = \prod_{n=1}^{N} I_n)$, the following two representations of $\mathcal{C}$ are equivalent:

$$\mathcal{C} = \mathcal{M} \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2 ... \times_N \mathbf{D}_N, \tag{4}$$

$$\mathbf{c} = (\mathbf{D}_N \otimes \mathbf{D}_{N-1} \otimes, ..., \otimes \mathbf{D}_1)\mathbf{m}, \tag{5}$$

where the symbol $\otimes$ denotes Kronecker product. $||\mathcal{M}||_0$ denotes ℓ_0 norm of tensor $\mathcal{M}$, defined as the number of non-zero elements of tensor $\mathcal{M}$, and the Frobenius norm of tensor $\mathcal{M}$ is defined as $||\mathcal{M}||_F = \sqrt{\sum_{i_1,...,i_N} |m_{i_1 ... i_N}|^2}$.

3. Problem Formulation

The desired HR-HSI is denoted by $\mathcal{X} \in R^{W \times H \times S}$, where W, H and S are the dimensions of the width mode, height mode and spectral mode, respectively. $\mathcal{Y} \in R^{w \times h \times S}$ denotes the acquired LR-HSI with invariant spectral bands, which is the spatial downsampled version of $\mathcal{X}$, where $W > w$, $H > h$. $\mathcal{Z} \in R^{W \times H \times s}$ represents the HR-MSI image of the same scene with the invariant spatial dimension, which is the spectral downsampled version of $\mathcal{X}$, where $S > s$. The goal of super-resolution is to estimate the HR-HSI $\mathcal{X}$ by fusing the LR-HSI $\mathcal{Y}$ with HR-MSI $\mathcal{Z}$.

3.1. Matrix factorization based HSI super-resolution

The matrix factorization based super-resolution methods assume each pixel of the target HR-HSI can be written as the linear combination of a small number of distinct spectral signatures [12]. As can be seen from Fig. 1(a), these approaches unfold the HR-HSI with spectral mode as matrix, and then the unfolding matrix can be decomposed as follows:

$$\mathbf{X}_{(3)} = \mathbf{D}\mathbf{A}, \tag{6}$$

where $\mathbf{X}_{(3)} \in R^{S \times WH}$ is the matrix by unfolding the HR-HSI $\mathcal{X}$ with the spectral mode; matrix $\mathbf{D} \in R^{S \times L}$ and $\mathbf{A} \in R^{L \times WH}$ is the spectral basis and corresponding coefficient matrix, respectively. Both the LR-HSI and HR-MSI can be expressed as linear combinations of the desired HR-HSI:

$$\mathbf{Y}_{(3)} = \mathbf{X}_{(3)}\mathbf{M}, \mathbf{Z}_{(3)} = \mathbf{P}_3\mathbf{X}_{(3)}, \tag{7}$$

where $\mathbf{Y}_{(3)} \in R^{S \times wh}$ and $\mathbf{Z}_{(3)} \in R^{s \times WH}$ are the spectral mode unfolding matrixes of $\mathcal{Y}$ and $\mathcal{Z}$, respectively; $\mathbf{M} \in R^{WH \times wh}$ and $\mathbf{P}_3 \in R^{s \times S}$ are the spatial downsampling and spectral downsampling matrixes respectively. In the matrix factorization based super-resolution approaches, the goal is to estimate the spectral basis $\mathbf{D}$ and coefficient matrix $\mathbf{A}$ from $\mathbf{Y}_{(3)}$ and $\mathbf{Z}_{(3)}$.

3.2. Tensor Factorization based HSI Super-resolution

It can be seen from Fig. 1(b) that different from matrix factorization based methods, the proposed tensor factorization based method can directly decompose a typical scene of the HR-HSI as a core tensor and dictionaries of the width mode, height mode and spectral mode. The problem can be formulated as follows:

$$\mathcal{X} = \mathcal{C} \times_1 \mathbf{W} \times_2 \mathbf{H} \times_3 \mathbf{S}, \tag{8}$$

where the matrix $\mathbf{W} \in R^{W \times n_w}$, $\mathbf{H} \in R^{H \times n_h}$, $\mathbf{S} \in R^{S \times n_s}$ denote the dictionaries of the width mode with n_w atoms, height mode with n_h atoms and spectral mode with n_s atoms, respectively. The tensor $\mathcal{C} \in R^{n_w \times n_h \times n_s}$ is the coefficient of $\mathcal{X}$ over the three dictionaries. The acquired LR-HSI $\mathcal{Y}$ is the spatially downsampled version of $\mathcal{X}$,

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{P_1} \times_2 \mathbf{P_2}, \tag{9}$$

where $\mathbf{P_1} \in R^{w \times W}$ and $\mathbf{P_2} \in R^{h \times H}$ are the downsampling matrixes along the width mode and height mode, respectively, which describe the spatial response of the imaging sensors.

The HR-MSI $\mathcal{Z}$ is the spectrally downsampled version of $\mathcal{X}$,

$$\mathcal{Z} = \mathcal{X} \times_3 \mathbf{P_3}, \tag{10}$$

where $\mathbf{P}_3 \in R^{s \times S}$ is the downsampling matrix of the spectral mode. Here, the HR-MSI is RGB image.

To reconstruct a typical scene of the HSI, we only need to estimated the dictionaries of the three modes and corresponding core tensor, as shown in Fig. 1(b).

4. Proposed NLSTF Approach

As shown in Fig. 2, the proposed NLSTF algorithm mainly includes three steps: Non-local clustering of the similar cubes, tensor dictionary learning and tensor sparse coding. Instead of estimating the whole HR-HSI directly, we reconstruct the HR-HSI in a cube-by-cube manner, which can reduce the computation cost. According to above tensor based HSI decomposition, the problem of the HSI super-resolution can be changed to estimate dictionaries of three modes and corresponding core tensor for each cube of the HR-HSI. Firstly, to exploit the non-local spatial similarities, we group the similar cubes of the HR-MSI together, and then the cubes of the LR-HSI and unknown HR-HSI are also grouped according to corresponding spatial location. The cubes of HR-HSI in the same group are decomposed on the same dictionaries with a sparse prior. Next, the dictionaries of three modes are learned for every group, and sparse core tensor for every cube is estimated by sparse coding algorithm. Finally, sparse core tensors and dictionaries can be utilized to reconstruct the cubes of the HR-HSI. More details of each step are described in the following.

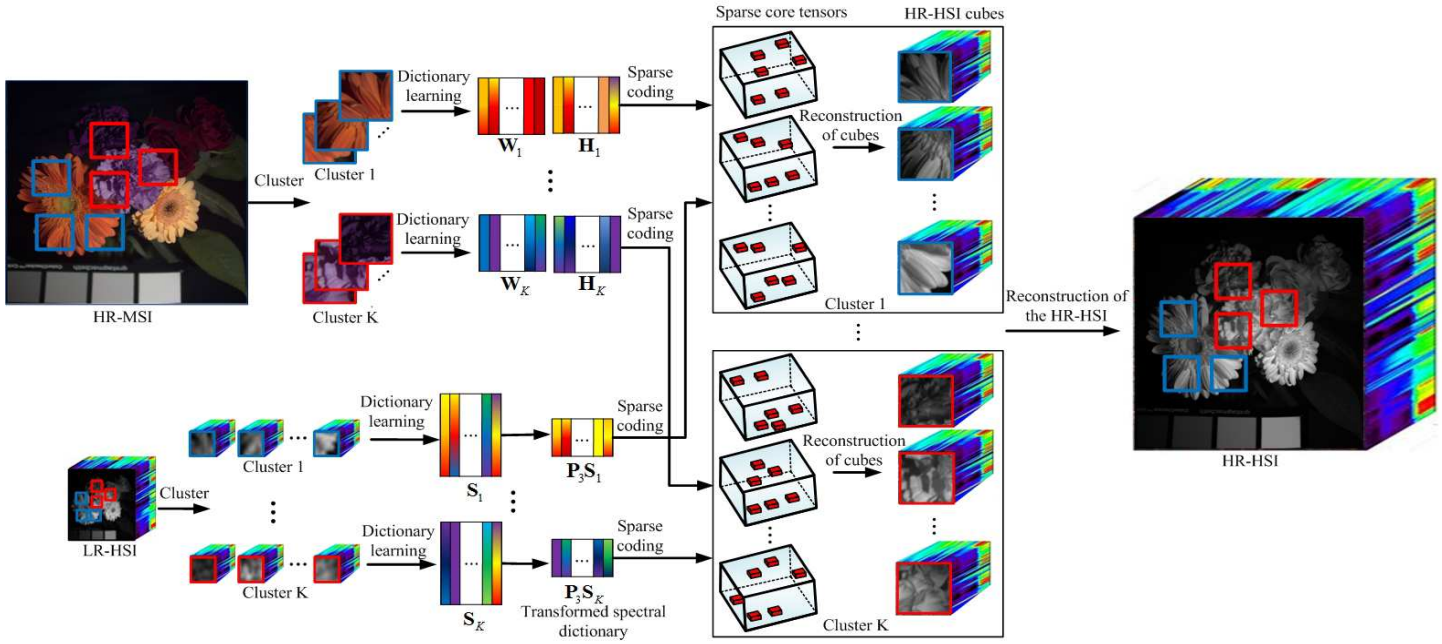


Figure 2. Scheme of the proposed NLSTF method.

4.1. Non-local Clustering of the Similar Cubes

Since the spatial information of the HR-HSI mainly exists in the HR-MSI, the HR-MSI $\mathcal{Z} \in R^{W\times H\times s}$ is spatially partitioned into several overlap cubes. The basic idea is that similar cubes of the HR-MSI are grouped into clusters $Z^{(k)} = \{\mathcal{Z}^{(k,j)}\}_{j=1}^{n_k}, k = 1, 2, ..., K$, where K is the number of clusters, and n_k is the number of cubes in the k^{th} cluster. $\mathcal{Z}^{(k,j)} \in R^{d_W\times d_H\times s}$ denotes the j^{th} cube of the k^{th} cluster, where d_W and d_H are the dimensions of the width mode and height mode, respectively. In the cluster process, we employ the efficient K-means++ method [6] (with automatically and carefully chosen initial seeds) to obtain clusters of all HR-MSI cubes. According to the corresponding spatial location, the LR-HSI $\mathcal{Y} \in R^{w\times h\times S}$ and unknown HR-HSI $\mathcal{X} \in R^{W\times H\times S}$ are also grouped into K clusters $Y^{(k)} = \{\mathcal{Y}^{(k,j)}\}_{j=1}^{n_k} \subset R^{d_w\times d_h\times S}$ and $X^{(k)} = \{\mathcal{X}^{(k,j)}\}_{j=1}^{n_k} \subset R^{d_W\times d_H\times S}, k = 1, 2, ..., K$, respectively. Since the spatial size of the LR-HSI is smaller than the HR-MSI, a pixel in the LR-HSI corresponds to a $c\times c$ (downsampling factor) cube in the HR-MSI. Once any pixel of the $c \times c$ cube in the HR-MSI is grouped into one cluster, the pixel in the LR-HSI is also grouped into this group. In this way, a pixel in the LR-HSI may belong to different groups, simultaneously.

4.2. Tensor Dictionary Learning

Since cubes in the same cluster are similar, they are assumed to share the same dictionaries of three modes. The dictionary learning process is the same for all clusters. Without loss of generality, we take the process of dictionary learning in the k^{th} cluster as an example to present our dictionary learning process.

Based on the above mentioned tensor factorization, the cubes $\mathcal{X}^{(k,j)}$ in the k^{th} cluster can be formulated as

$$\mathcal{X}^{(k,j)} = \mathcal{C}^{(k,j)}\times_1\mathbf{W}_k\times_2\mathbf{H}_k\times_3\mathbf{S}_k, j = 1, 2, ..., n_k, \quad (11)$$

where the matrixes $\mathbf{W}_k \in R^{d_W\times l_W}$, $\mathbf{H}_k \in R^{d_H\times l_H}$ and $\mathbf{S}_k \in R^{S\times l_S}$ denote the dictionaries of the width mode with l_W atoms, height mode with l_H atoms and spectral mode with l_S atoms, respectively. The tensor $\mathcal{C}^{(k,j)} \in R^{l_W\times l_H\times l_S}$ is a core tensor which models the relationship of the three dictionaries. Since the HR-MSI mainly contains the spatial information of the HR-HSI, the dictionaries $\mathbf{W}_k$ and $\mathbf{H}_k$ can be learned from $\{\mathcal{Z}^{(k,j)}\}_{j=1}^{n_k}$. According to the equation (10), the cubes $\mathcal{Z}^{(k,j)}$ of the k^{th} cluster in the HR-MSI can be formulated as:

$$\mathcal{Z}^{(k,j)} = \mathcal{X}^{(k,j)} \times_3\mathbf{P}_3, j = 1, 2, ..., n_k, \quad (12)$$

According to equation (11) and (12), the $\mathcal{Z}^{(k,j)}$ can also be formulated as

$$\mathcal{Z}^{(k,j)} = \mathcal{C}^{(k,j)}\times_1\mathbf{W}_k\times_2\mathbf{H}_k\times_3\mathbf{S}_k^*, j = 1, 2, ..., n_k, \quad (13)$$

where $\mathbf{S}_k^* = \mathbf{P}_3\mathbf{S}_k$ is the transformed spectral dictionary. Unfolding the tensor $\mathcal{Z}^{(k,j)}$ and $\mathcal{A}^{(k,j)}$ with the width mode, the equation (13) can be represented as

$$\mathbf{Z}_{(1)}^{(k,j)} = \mathbf{W}_k \times \mathbf{A}_{(1)}^{(k,j)}, j = 1, 2, ..., n_k, \quad (14)$$

where $\mathbf{Z}_{(1)}^{(k,j)}$ and $\mathbf{A}_{(1)}^{(k,j)}$ are 1-mode unfolding matrixes of tensors $\mathcal{Z}^{(k,j)}$ and $\mathcal{A}^{(k,j)} = \mathcal{C}^{(k,j)} \times_2 \mathbf{H}_k \times_3 \mathbf{S}_k^*$, respectively. From equation (14), we can observe each column of $\mathbf{M}_{w_k} = [\mathbf{Z}_{(1)}^{(k,1)}, \mathbf{Z}_{(1)}^{(k,2)}, ..., \mathbf{Z}_{(1)}^{(k,n_k)}]$ can be represented as a linear combination of columns in the matrix $\mathbf{W}_k$. The estimation of $\mathbf{W}_k$ is severely ill-posed problem, because the decomposition of $\mathbf{M}_{w_k}$ is not unique. We use sparsity prior to regularize the problem, which can not only better estimate $\mathbf{W}_k$, but can also promote the sparsity in core tensor. Hence, the estimation of matrix $\mathbf{W}_k$ can be seen as a sparsity-constrained dictionary learning problem. The problem can be formulated as

$$\begin{aligned} &\min_{\mathbf{W}_k, \mathbf{B}_{W_k}} ||\mathbf{M}_{w_k} - \mathbf{W}_k \times \mathbf{B}_{w_k}||_F^2, \\ &s.t. ||\mathbf{B}_{w_k}(:,i)||_0 \leq k_w, 1 \leq i \leq l d_H n_k, \end{aligned} \tag{15}$$

where $||\cdot||_0$ and $||\cdot||_F$ denote the ℓ_0 norm and Forbenius norm, respectively, and k_w represents permissible maximum number of non-zero elements of each column in coefficient matrix $\mathbf{B}_{w_k}$. To solve the problem in (15), we use the dictionary-updates-cycles KSVD (DUC-KSVD) [25] approach which is a modification version of the KSVD algorithm [1]. In the dictionary-update stage, both the dictionary and representations are found while keeping the supports intact. The known representations are leveraged from the previous sparse coding in the quest for the updated representations in the sparse coding stages.

Unfolding the tensor $\mathcal{Z}^{(k,j)}$ and $\mathcal{A}^{(k,j)}$ with the height mode, the equation (13) can be represented as:

$$\mathbf{Z}_{(2)}^{(k,j)} = \mathbf{H}_k \times \mathbf{B}_{(2)}^{(k,j)}, j = 1, 2, ..., n_k, \tag{16}$$

where $\mathbf{Z}_{(2)}^{(k,j)}$ and $\mathbf{B}_{(2)}^{(k,j)}$ are 2-mode (height mode) unfolding matrixes of tensors $\mathcal{Z}^{(k,j)}$ and $\mathcal{B}^{(k,j)} = \mathcal{C}^{(k,j)} \times_1 \mathbf{W}_k \times_3 \mathbf{S}_k^*$, respectively. From equation (16), we can also find that each column of the matrix $\mathbf{M}_{h_k} = [\mathbf{Z}_{(2)}^{(k,1)}, \mathbf{Z}_{(2)}^{(k,2)}, ..., \mathbf{Z}_{(2)}^{(k,n_k)}]$ can be represented as a linear combination of columns in $\mathbf{H}_k$. Similar with the estimation of $\mathbf{W}_k$, the acquisition of $\mathbf{H}_k$ can also be changed into the sparsity constrained dictionary learning problem:

$$\begin{aligned} &\min_{\mathbf{H}_k, \mathbf{B}_{h_k}} ||\mathbf{M}_{h_k} - \mathbf{H}_k \times \mathbf{B}_{h_k}||_F^2, \\ &s.t. ||\mathbf{B}_{h_k}(:,i)||_0 \leq k_h, 1 \leq i \leq l d_W n_k, \end{aligned} \tag{17}$$

where k_h is the maximum number of non-zero elements of each column in the matrix $\mathbf{B}_{h_k}$. Similarly, the problem in (17) can be solved by the DUC-KSVD algorithm.

Since the LR-HSI is only spatially downsampled, it still has the main spectral information of the HR-HSI. Hence, we can induce the dictionary of the spectral mode $\mathbf{S}_k$ from $\{\mathcal{Y}^{(k,j)}\}_{j=1}^{n_k}$, which are cubes of the k^{th} cluster in the LR-HSI. Assuming each pixel of the HR-HSI cubes of k^{th} cluster can be written as the linear combination of a small number of distinct spectral signatures, the dictionary of the spectral mode $\mathbf{S}_k$ can be estimated by solved the following problem:

$$\begin{aligned} &\min_{\mathbf{S}_k, \mathbf{B}_{s_k}} ||\mathbf{M}_{s_k} - \mathbf{S}_k \times \mathbf{B}_{s_k}||_F^2, \\ &s.t. ||\mathbf{B}_{s_k}(:,i)||_0 \leq k_s, 1 \leq i \leq d_w d_h n_k, \end{aligned} \tag{18}$$

where $\mathbf{M}_{s_k} = [\mathbf{Y}_{(3)}^{(k,1)}, \mathbf{Y}_{(3)}^{(k,2)}, ..., \mathbf{Y}_{(3)}^{(k,n_k)}]$ is the 3-mode (spectral mode) matrix obtained from all the LR-HSI cubes of the k^{th} cluster, and k_s is the maximum number of nonzero elements of each column in the matrix $\mathbf{B}_{s_k}$. Similarly, the problem in (18) can be solved by the DUC-KSVD algorithm.

4.3. Tensor Sparse Coding

Once the dictionaries $\mathbf{W}_k$, $\mathbf{H}_k$ and $\mathbf{S}_k$ of the k^{th} cluster are known, the core tensor $\mathcal{C}^{(k,j)} \in R^{l_W \times l_H \times l_S}$ should be estimated in order to get the HR-HSI cubes of k^{th} cluster. The estimation of $\mathcal{C}^{(k,j)}$ is severely ill-posed problem, and we need to use the prior information to regularize it. The dictionaries of three modes are all estimated with sparse prior, which means the dictionaries of three modes are redundant enough to represent information in each mode. Hence we assume the cubes of the HR-HSI can be sparsely represented by the three dictionaries, which means core tensors $\mathcal{C}^{(k,j)}$ are sparse. In this way, the estimation of the core tensor $\mathcal{C}^{(k,j)}$ can be changed into the following l_0 norm constrained optimization problem:

$$\begin{aligned} &\min_{\mathcal{C}^{(k,j)}} ||\mathcal{Z}^{(k,j)} - \mathcal{C}^{(k,j)} \times_1 \mathbf{W}_k \times_2 \mathbf{H}_k \times_3 \mathbf{S}_k^*||_F^2, \\ &s.t. ||\mathcal{C}^{(k,j)}||_0 \leq m, \end{aligned} \tag{19}$$

where m is permissible maximum sparsity. According to the relationship of the Tucker mode and Kronecker product, the problem in (19) can also be formulated as:

$$\begin{aligned} &\min_{\mathbf{c}^{(k,j)}} ||\mathbf{z}^{(k,j)} - \mathbf{D}_k \times \mathbf{c}^{(k,j)}||_F^2, \\ &s.t. ||\mathbf{c}^{(k,j)}||_0 \leq m, \end{aligned} \tag{20}$$

where $\mathbf{c}^{(k,j)} = vec(\mathcal{C}^{(k,j)}) \in R^{l_W l_H l_K \times 1}$ and $\mathbf{z}^{(k,j)} = vec(\mathcal{Z}^{(k,j)}) \in R^{s d_W d_H \times 1}$ are vectors by stacking all the 1-mode vectors of tensor $\mathcal{C}^{(k,j)}$ and $\mathcal{Z}^{(k,j)}$, respectively, and the matrix $\mathbf{D}_k = \mathbf{S}_k^* \otimes \mathbf{H}_k \otimes \mathbf{W}_k \in R^{s d_W d_H \times l_W l_H l_S}$ is the dictionary. The problem in (20) is a NP-hard problem [11] , which indicates that the problem should be relaxed or solved by greedy strategy. In general, the Kronecker operation will create a very large dictionary $\mathbf{D}_k$, which results in a very heavily computational burden for the sparse coding. To achieve the efficient sparse coding over the large dictionary, a very efficient greedy approach, called as the

Method	**CAVE database** [37]			
	RMSE	SAM	SSIM	ERGAS
SNNMF [34]	4.38	17.85	0.918	0.773
GSOMP [2]	5.44	12.23	0.960	0.781
SSR [24]	4.71	22.00	0.945	0.642
BSR [3]	5.19	12.93	0.955	0.742
NLSTF	**2.60**	**6.83**	**0.980**	**0.372**

Table 1. Quantitative results (on RMSE, SAM, SSIM and ERGAS) of the test methods on the CAVE database [37].

Method	**Harvard database** [8]			
	RMSE	SAM	SSIM	ERGAS
SNNMF [34]	2.46	4.93	0.973	0.381
GSOMP [2]	3.10	4.34	0.971	0.449
SSR [24]	3.08	5.59	0.820	0.459
BSR [3]	2.64	4.48	0.974	0.453
NLSTF	**1.78**	**3.12**	**0.982**	**0.261**

Table 2. Quantitative results (on RMSE, SAM, SSIM and ERGAS) of the test methods on the Harvard database [8].

Matching Pursuit Lasso (MPL) [26, 27], is used. The MPL is based on a novel quadratically constrained linear program formulation, which can greatly reduces the computation cost of sparse coding problem over large dictionary.

Once the dictionaries $\mathbf{W}_k$, $\mathbf{H}_k$, $\mathbf{S}_k$, and core tensors $\{\mathcal{C}^{(k,j)}\}_{j=1}^{n_k}$ are known, the HR-HSI cubes $\{\mathcal{X}^{(k,j)}\}_{j=1}^{n_k}$ of the k^{th} cluster can be estimated by equation (11). Finally, the estimated cube sets can be returned to the original place to reconstruct the HR-HSI $\mathcal{X}$. In addition, the performance of the proposed approach can be further improved via back-projection operation [13].

5. Experiments

5.1. Experimental Database

In this section, experiments are conducted on two public databases (CAVE database [37][1] and Harvard database [8][2]) to evaluate the effectiveness of the proposed NLSTF method. The CAVE database [37] consists of 32 indoor HSIs captured under controlled illumination. The images have 31 spectral bands, and each band has a size of 512×512. The images of the scenes are acquired at a wavelength interval 10nm in the range of 400-700nm. The Harvard database [8] has 50 indoor and outdoor images recorded under daylight illumination, and 27 images under artificial or mixed illumination. The spatial resolution of the images is 1392×1040, with 31 spectral bands. The images of the scenes are acquired at a wavelength interval 10nm in the range of 420-720nm. We use only the top left 1024×1024 pixels for the convenience of the spatial downsampling process. The HSIs from two databases are used as ground truth images. We downsample the HR-HSIs by averaging the 32×32 disjoint spatial blocks, to get the LR-HSIs $\mathcal{Y}$. The HR-MSI (RGB image) $\mathcal{Z}$ of the same scene can be stimulated by downsampling $\mathcal{X}$ with spectral model using spectral dowmsampling matrix $\mathbf{P}_3$ derived from the response of a Nikon D700 camera[3].

5.2. Compared Methods

We have compared the proposed method with several state-of-the-art HSI super-resolution methods, including the Sparse Non-negative Matrix Factorization (SNNMF) [34], Generalization of Simultaneous Orthogonal Matching Pursuit (GSOMP) method [2], Subspace Regularization (SSR) method [24] and Bayesian Sparse Representation (BSR) method [3].

5.3. Quantitative Metrics

To evaluate the quality of the reconstructed HSIs, four indexes are used in our study. The first index is root mean square error (RMSE), and the images are on a 8-bit intensity range. We also use the the spectral angle mapper (SAM), which is given in degrees. The third index is SSIM [32] which is defined as the mean SSIM of all bands between the estimated HSI and the ground truth. The fourth index is the relative dimensionless global error in synthesis (ERGAS), proposed in [31].

5.4. Parameters Discussion

The maximum number of non-zero elements sparsity m has an important influence on the accuracy and efficiency of sparse coding problem. Since the size of dictionary $\mathbf{D}_k$ in different databases may be different, it is not convenient to discuss the effects of m directly. Therefore, we test the effects of sparsity scaling parameter, defined as

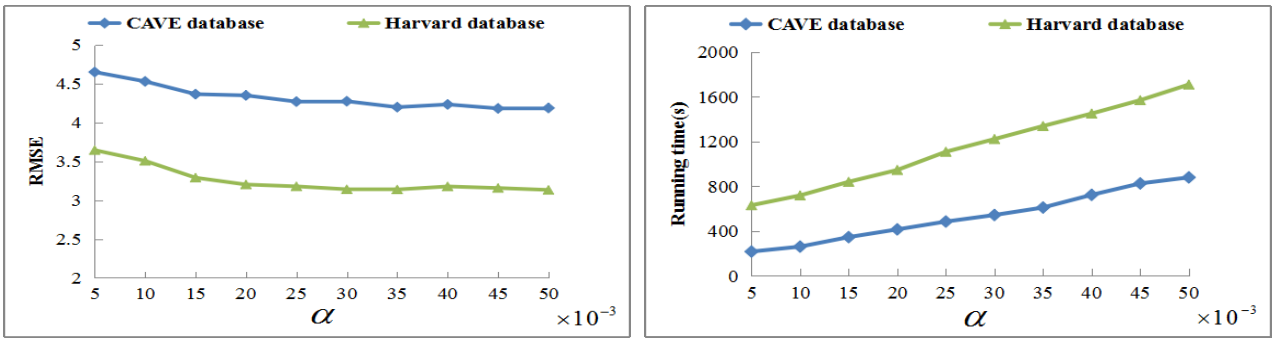
$$\alpha = \frac{m}{l_W l_H l_S}, \tag{21}$$

which is proportional to the sparsity of the solution. Fig. 3 plots the curves of the RMSE and time of reconstructed HSIs *Cloth* (image in CAVE database), and *b4* (image in the Harvard database) under various parameters α. From Fig. 3 (a), we can see that the accuracy of the NLSTF method is affected obviously when tuning α from 0.005 to 0.03. The parameter α also has an important effect on the sparse coding time which occupies the majority of the running time. It can be seen from Fig. 3 (b) that the time has an approximate linear increase with the growth of α in both the CAVE database and Harvard database. This is because the bigger

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

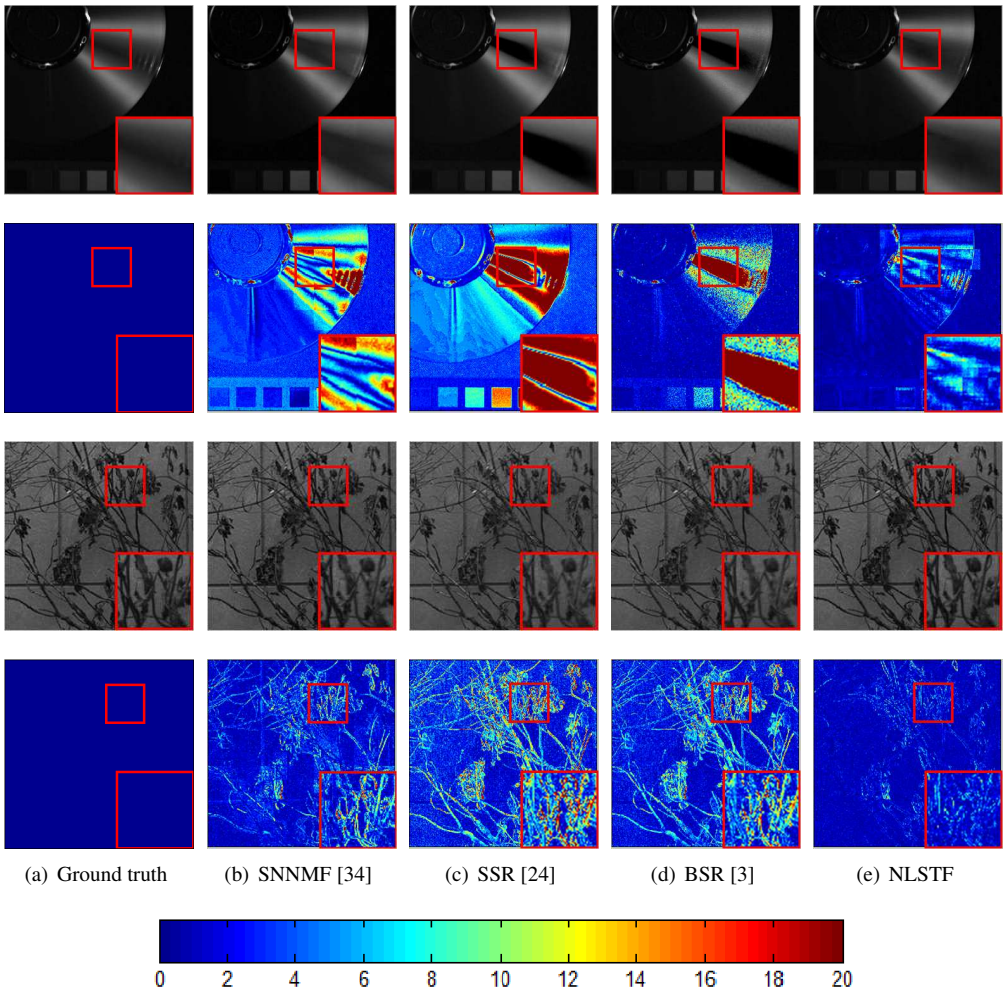
[2] http://vision.seas.harvard.edu/hyperspec/

[3] https://www.maxmax.com/spectral_response.htm



(a) (b)

Figure 3. The RMSE and running time in seconds curves under various sparsity scaling parameters for the proposed NLSTF method. (a) RMSE; (b) Running time.



(a) Ground truth (b) SNNMF [34] (c) SSR [24] (d) BSR [3] (e) NLSTF

Figure 4. The first row: the reconstructed images of *CD* in the CAVE database at 670nm. The second row: the corresponding error images of the competing approaches for the image *CD*. The third row: the reconstructed images of *b4* in the Harvard database at 550nm. The fourth row: the corresponding error images of the competing approaches for the image *b4*.

the value of α is , the more atoms will be chosen in the process of sparse coding , which can add computational cost. We set $\alpha = 0.03$ for both two databases.

The remaining parameters of the proposed NLSTF

method are set as follows: the number of clusters $K = \frac{N_c}{100}$, where N_c is the total number of cubes, the spatial size of HR-HSI cubes is 8×8 ($d_W = 8, d_H = 8$) with overlap 4×4, the number of atoms is $l_W = 10$, $l_H = 10$, $l_S = 32$. The target sparsity is $k_w = 2$, $k_h = 2$, $k_s = 2$ in the process of dictionary learning.

5.5. Experimental Results

All the methods are run with the same spectral downsampling matrix $\mathbf{P}_3$, and the parameters are set as the default values. For the SSR [24] method, we directly use the spectral downsampling matrix $\mathbf{P}_3$ already known in our experiments, instead of estimating it as in [24].

Table 1 shows the average objective results of the CAVE database in terms of RMSE, SAM, SSIM and ERGAS. The best results are marked in bold for clarity. As can be seen from Table 1, the proposed NLSTF method performs consistently better than the other compared methods. Specifically, the NLSFT advantage is considerable in the cases of RMSE, SAM, SSIM, and ERGAS. The significantly lower SAM indicates that our approach performs the best in reconstructing the spectral distribution of the intensities. Our approach has the largest SSIM among all the testing methods, which means the proposed method can better preserve the spatial structures of the HSI. In Fig. 4, we show the reconstructed HR-HSI at 670mm by the competing method for the test image *Cloth* of the CAVE database. For better visual comparison, one meaningful region for each of the resulting image is magnified. It can be seen from Fig. 4 that the proposed NLSTF approach in reconstructing the very detailed structures.

The average RMSE, SAM, SSIM and ERGAS of the recovered HSIs of the Harvard database are reported in Table 2. The proposed NLSTF method also outperforms other competing methods as it has the lowest RMSE, SAM, ERGAS and the biggest SSIM. In Fig. 4, the reconstruction results of testing approaches for image *b4* of the Harvard database is shown. As can be observed, the proposed NLSTF approach can also better recover the HR structures of HSI among the testing methods.

5.6. The Effectiveness of Non-local Part and Sparse Tensor Factorization Part

The proposed NLSTF method mainly has the non-local clustering step and sparse tensor factorization part. In this subsection, we clarify the effectiveness of two steps, respectively.

To the best of our knowledge, no sparse tensor factorization method has been used for the hyperspectral image super-resolution. Hence, we just remove the non-local clustering part of the NLSTF method to get the sparse tensor factorization (STF) method. In the STF method, all HR-HSI cubes are assumed share the same dictionaries $\mathbf{W}$, $\mathbf{H}$, and $\mathbf{S}$. Also, the sparse matrix factorization (SMF) method [13] is included into the comparisons. Both of the SMF and STF methods only exploit a sparse prior without other prior information. The main difference of them is the SMF is based on matrix factorization, and the STF is based on tensor factorization. Since the code of SMF is not available, we directly use the results of SMF on five images of the CAVE database from the reference [4]. Besides, we also compare with the hierarchical beta process (HBP) method [4]. To ensure fair comparison, the experimental settings of NLSTF and STF are the same as that of [4].

Method	CAVE database				
	Ballons	Beads	Cloth	Pompoms	CD
SMF [13]	2.3	8.2	6.0	4.3	7.9
HBP [4]	1.9	5.8	**3.7**	3.9	**5.3**
STF	1.5	6.9	4.6	**2.5**	6.4
NLSTF	**1.3**	**5.5**	**3.7**	**2.5**	**5.3**

Table 3. Quantitative results (on RMSE) of the test methods.

The results of SMF, STF, NLSTF, and HBP are reported in Table 3. The STF method performs consistently better than the SMF method, which can indicate the advantages of tensor factorization over matrix factorization. Furthermore, the NLSTF method outperforms the STF method, which proves that the non-local strategy indeed improves the performance.

6. Conclusions

In this paper, we present a novel non-local sparse tensor factorization based framework to acquire the HR-HSI, by fusing a LR-HSI with a HR-MSI. Unlike recent matrix factorization based HSI super-resolution methods, the proposed NLSTF method considers the HSI as a tensor with three modes, and factorizes the tensor as a sparse core tensor multiplication by dictionaries of the three modes. In addition, non-local spatial self-similarity is incorporated into the sparse tensor factorization. With the proposed framework, the spatial-spectral information of the HSI can be better exploited. Our approach is tested on two public databases, which demonstrates the superiority of the proposed method over several state-of-the-art HSI super-resolution methods.

Acknowledgement

This work was supported by the National Natural Science Fund of China for Distinguished Young Scholars under Grant 61325007, and by the National Natural Science Fund of China for International Cooperation and Exchanges under Grant 61520106001.

References

[1] M. Aharon, M. Elad, and A. M. Bruckstein. K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representations. *IEEE Trans. Signal Process.*, 54(11):4311–4322, 2006.

[2] N. Akhtar, F. Shafait, and A. Mian. Sparse spatio-spectral representation for hyperspectral image super-resolution. *ECCV*, pages 63–78, 2014.

[3] N. Akhtar, F. Shafait, and A. Mian. Bayesian sparse representation for hyperspectral image super resolution. *IEEE CVPR*, pages 3631–3640, 2015.

[4] N. Akhtar, F. Shafait, and A. Mian. Hierarchical beta process with gaussian process prior for hyperspectral image super resolution. *ECCV*, pages 103–120, 2016.

[5] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce. Comparison of pan-sharpening algorithms: Outcome of the 2006 GRSS data-fusion contest. *IEEE Trans. Geosci. Remote Sens.*, 45(10):3012–3021, 2007.

[6] D. Arthur and S. Vassilvitskii. K-means++: the advantages of careful seeding. *Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1027–1035, 2007.

[7] C. F. Caiafa and A. Cichocki. Block sparse representations of tensors using kronecker bases. *IEEE ICASSP*, pages 2709–2712, 2012.

[8] A. Chakrabarti and T. Zickler. Statistics of real-world hyperspectral images. *IEEE CVPR*, pages 193–200, 2011.

[9] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li. Hyperspectral image super-resolution via non-negative structured sparse representation. *IEEE Trans. Image Process.*, 25(5):2337–2352, 2016.

[10] W. Dong, G. Li, G. Shi, X. Li, and Y. Ma. Low-rank tensor approximation with laplacian scale mixture modeling for multiframe image denoising. *IEEE ICCV*, pages 442–449, 2015.

[11] M. Elad, M. A. T. Figueriredo, and Y. Ma. On the role of sparse andredundant representations in image processing. *Proc. IEEE*, 98(6):972–982, 2010.

[12] M. D. Iordache, J. Bioucas-Dias, and A. Plaza. Sparse unmixing of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.*, 49(6):2014–2039, 2011.

[13] R. Kawakami, J. Wright, Y.-W. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi. High-resolution hyperspectral imaging via matrix factorization. *IEEE CVPR*, pages 2329–2336, 2011.

[14] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Rev.*, 51(3):455–500, 2009.

[15] C. Lanaras, E. Baltsavias, and K. Schindler. Hyperspectral super-resolution by coupled spectral unmixing. *IEEE ICCV*, pages 3586–3594, 2015.

[16] S. Li and B. Yang. A new pan-sharpening method using a compressed sensing technique. *IEEE Trans. Geosci. Remote Sens.*, 49(2):738–746, 2011.

[17] J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(1):208–220, 2013.

[18] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. *IEEE ICCV*, pages 2272–2279, 2009.

[19] H. V. Nguyen, A. Banerjee, and R. Chellappa. Tracking via object reflectance using a hyperspectral video camera. *IEEE CVPRW*, pages 44–51, 2010.

[20] Z. Pan, G. Healey, M. Prasad, and B. Tromberg. Face recognition in hyperspectral images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(12):1552–1560, 2003.

[21] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang. Decomposable nonlocal tensor dictionary learning for multispectral image denoising. *IEEE CVPR*, pages 2949–2956, 2014.

[22] S. Rahmani, M.Strait, D.Merkurjev, M. Moeller, and T. Wittman. An adaptive IHS pan-sharpening method. *IEEE Geosci. Remote Sens. Let.*, 7(4):746–750, 2010.

[23] V. Shah, N. Younan, and R. King. An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets. *IEEE Trans. Geosci. Remote Sens.*, 46(5):1323–1335, 2008.

[24] M. Simoes, J. Bioucas-Dias, L. Almeida, and J. Chanussot. A convex formulation for hyperspectral image super-resolution via subspace-based regularization. *IEEE Trans. Geosci. Remote Sens.*, 53(6):3373–3388, 2015.

[25] L. Smith and M. Elad. Improving dictionary learning:multiple dictionary updates and coefficient reuse. *IEEE Signal Process. Let.*, 20(1):79–82, 2013.

[26] M. Tan, I. W. Tsang, and L. Wang. Matching pursuit LASSO part I: Sparse recovery over big dictionary. *IEEE Trans. Signal Process.*, 63(3):727–741, 2015.

[27] M. Tan, I. W. Tsang, and L. Wang. Matching pursuit LASSO part II: Applications and sparse recovery over batch signals. *IEEE Trans. Signal Process.*, 63(3):742–753, 2015.

[28] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson. Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recog.*, 43(7):2367–2379, 2010.

[29] L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 23(8):3336–3351, 1996.

[30] M. A. Veganzones, M. Simoes, G. Licciardi, N. Yokoya, J. M. Bioucas-Dias, and J. Chanussot. Hyperspectral super-resolution of locally low rank images from complementary multisource data. *IEEE Trans. Image Process.*, 25(1):274–288, 2016.

[31] L. Wald. Quality of high resolution synthesised images: Is there a simplecriterion? *Int. Conf. Fusion Earth Data*, pages 99–103, 2000.

[32] Z. Wang, A. Bovik, and H. Sheikh. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004.

[33] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret. Hyperspectral and multispectral image fusion based on a sparse representation. *IEEE Trans. Geosci. Remote Sens.*, 53(7):3658–3668, 2015.

[34] E. Wycoff, T. H. Chan, K. Jia, W. K. Ma, and Y. Ma. A non-negative sparse promoting algorithm for high resolution hyperspecral imaging. *IEEE ICASSP*, pages 1409–1413, 2013.

[35] X.Guo, X. Huang, L. Zhang, L. Zhang, A. Plaza, and J. A. Benediktsson. Support tensor machines for classification of hyperspectral remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.*, 54(6):3248–3264, 2016.

[36] S. Yang, M. Wang, P. Li, L. Jin, B. Wu, and L. Jiao. Bayesian CP factorization of incomplete tensors with automatic rank determination. *IEEE Trans. Pattern Anal. Mach. Intell.*, 53(11):5943–5957, 2015.

[37] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar. Generalized assorted pixel camera: Post-capture control of resolution,dynamic range and spectrum. *Technical Report*, 2008.

[38] N. Yokoya, T. Yairi, and A. Iwasaki. Coupled non-negative matrix factorization unmixing for hyperspectraland multispectral data fusion. *IEEE Trans. Geosci. Remote Sens.*, 50(2):528–537, 2012.

[39] J. Zhang, D. Zhao, and W. Gao. Group-based sparse representation for image restoration. *IEEE Trans. Image Process.*, 23(8):3336–3351, 2014.

[40] Q. Zhao, A. Cichocki, and L. Zhang. Bayesian CP factorization of incomplete tensors with automatic rank determination. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(9):1751–1763, 2015.

[41] Q. Zhao, G. Zhou, L. Zhang, A. Cichocki, and S. I. Amari. Bayesian robust tensor factorizationfor incomplete multiway data. *IEEE Trans. Neural Netw. Learn. Syst.*, 27(4):736–748, 2016.