



**HAL**  
open science

# Nonlocal Coupled Tensor CP Decomposition for Hyperspectral and Multispectral Image Fusion

Yang Xu, Zebin Wu, Jocelyn Chanussot, Pierre Comon, Zhihui Wei

► **To cite this version:**

Yang Xu, Zebin Wu, Jocelyn Chanussot, Pierre Comon, Zhihui Wei. Nonlocal Coupled Tensor CP Decomposition for Hyperspectral and Multispectral Image Fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58 (1), pp.348-362. 10.1109/TGRS.2019.2936486 . hal-02123922

**HAL Id: hal-02123922**

**<https://hal.science/hal-02123922>**

Submitted on 9 May 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Nonlocal Coupled Tensor CP Decomposition for Hyperspectral and Multispectral Image Fusion

Yang Xu, *Member, IEEE*, Zebin Wu, *Senior Member, IEEE*, Jocelyn Chanussot, *Fellow, IEEE*, Pierre Comon, *Fellow, IEEE*, and Zhihui Wei

**Abstract**—Hyperspectral super-resolution, which aims at enhancing the spatial resolution of hyperspectral images (HSIs), has recently attracted considerable attention. A common way of hyperspectral super-resolution is to fuse the HSI with a higher spatial resolution multispectral image (MSI). Various approaches have been proposed to solve this problem by establishing the degradation model of low spatial resolution HSIs and MSIs based on matrix factorization methods, e.g., unmixing and sparse representation. However, this category of approaches cannot well construct the relationship between the high spatial resolution HSI and MSI. In fact, since the HSI and the MSI capture the same scene, these two image sources must have common factors. In this paper, we propose a nonlocal tensor decomposition model for HSI-MSI fusion. Firstly, the nonlocal similar patch tensors of the HSI are constructed according to the MSI, for the purpose of calculating the smooth order of all the patches for clustering. Then, the relationship between the high spatial resolution HSI and the MSI is explored through a coupled tensor canonical polyadic (CP) decomposition. The fundamental idea of the proposed model is that the factor matrices in CP decomposition of the high spatial resolution HSI's nonlocal tensor can be shared with the matrices factorized by the MSI's nonlocal tensor. Alternating direction method of multipliers is used to solve the proposed model. Through this method, the spatial structure of the MSI can be successfully transferred to the HSI. Experimental results on three synthetic datasets and one real dataset suggest that the proposed method substantially outperforms existing state-of-the-art HSI-MSI fusion methods.

**Index Terms**—e Hyperspectral images, multispectral images, data fusion, nonlocal tensor, coupled CP decomposition.

This work was supported in part by the National Natural Science Foundation of China under Grant 61471199, Grant 61772274, Grant 61701238, Grant 91538108, Grant 11431015, Grant 61501241, Grant 61671243, and Grant 61802190, in part by the Jiangsu Provincial Natural Science Foundation of China under Grant BK20170858, Grant BK20180018, in part by the Fundamental Research Funds for the Central Universities under Grant 30919011234, 30917015104, 30919011103, in part by the Jiangsu Provincial Social Developing Project under Grant BE2018727, in part by the China Postdoctoral Science Foundation under Grant 2017M611814, Grant 2015M570450, Grant 2018T110502, and in part by the Jiangsu Province Postdoctoral Science Foundation under Grant 1701148B. (*Corresponding author: Zebin Wu.*)

Y. Xu is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China, and also with the Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France (email: xuyangth90@njust.edu.cn).

Z. Wu is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China, and also with Nanjing Robot Research Institute Co. Ltd, Nanjing, 211135, China (Email: zebin.wu@gmail.com).

J. Chanussot is with the Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France (e-mail: jocelyn.chanussot@gipsalab.grenoble-inp.fr, jocelyn@hi.is).

P. Comon is with CNRS, GIPSA-Lab, University Grenoble Alpes, Grenoble F-38000, France (e-mail: pierre.comon@gipsa-lab.fr).

Z. Wei is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China (Email: gswei@njust.edu.cn).

## I. INTRODUCTION

IN recent years, hyperspectral (HS) imaging has attracted much attention since it can obtain images capturing the same scene in different successive wavelengths at the same time [1–3]. The abundant spectral information is of great importance to improving the performances of many remote sensing applications [4–10]. However, in optical remote sensing systems, the limited amount of incident energy involves a tradeoff between the spatial resolution and the spectral resolution of images [11]. For example, hyperspectral images (HSIs) have a high spectral resolution but a low spatial resolution. To enhance the spatial resolution of HSIs, much effort has been made, under the name of hyperspectral and multispectral image fusion (HSI-MSI fusion), for fusing high spatial resolution (HR) multispectral image (MSI) [12] with low spatial resolution (LR) HSI. The fused image is expected to have both a high spectral resolution and a high spatial resolution.

The archetypal instance of HSI-MSI fusion is pansharpening which fuses a low spatial resolution MSI with a corresponding panchromatic (PAN) image. A variety of pansharpening approaches have been proposed in the last two decades [13–15]. The fusion of LR-HSI and PAN image has been further extensively studied under the name of hyperspectral pansharpening [11]. HSI contains more accurate spectral information than MSI does, and therefore has drawn more attention to the spectral fidelity. HSI-MSI fusion can be regarded as an extension of pansharpening and hyperspectral pansharpening. Therefore, substantial effort has been made to generalize the pansharpening and hyperspectral pansharpening methods for HSI-MSI fusion [13], [15]. In general, the HSI-MSI fusion methods can be classified into four groups, component substitution (CS), multi-resolution analysis (MRA), Bayesian, and sparse representation. The basic idea of CS approaches is to substitute one component of the HSI with the high spatial resolution image. Intensity-hue-saturation (IHS) [16], [17] method changes the intensity channel in the intensity, hue, and saturation domain of the MSI with the PAN image. Principal component analysis (PCA) [18] approaches substitute the first principal component in the LR-HSI with the high spatial resolution image. The spectral structure can be preserved by dividing the spectrum of the HSI into several regions and applying conventional pansharpening techniques in each region separately. In MRA approaches, a high-resolution image for each band of the LR-HSI is synthesized as a linear combination of the HR-MSI. The results obtained in this manner would be significantly better

than the results produced by using one high-resolution image for all bands. Representative methods include smoothing filtered-based intensity modulation (SFIM) [19] and generalized Laplacian pyramid (GLP) [20]. In Bayesian approaches, the proposed Bayesian formulation introduces the appropriate prior distribution of the MSI, such as naive Gaussian [21], [22] and sparsity promoting prior [23], [24]. The variational approaches can be regarded as special variants of the Bayesian methods. The fused image is estimated by minimizing the objective function which is built upon the posterior probability density of the fused image. Among these variant methods, HS super-resolution (HySure) [25] makes use of the vector-total-variation-based regularization. The sparse representation-based method [26] estimates the HS dictionary and sparse codes simultaneously by introducing the spatial-spectral sparse regularization. Akhtar *et al.* proposed to learn the dictionary and sparse codes in the Bayesian framework [27]. Huang *et al.* [28] proposed to use sparse matrix factorization to enhance the spatial resolution of low-resolution images of different spatial and spectral resolutions. Coupled nonnegative matrix factorization (CNMF) approach [29] estimates HR-HSI based on HS unmixing. In addition, a coupled matrix factorization approach with nonnegativity and sparsity constraints is proposed for HSI-MSI fusion [30]. Moreover, with the fast development of deep learning in remote sensing and computer vision tasks [31–35], researchers have attempted to employ deep learning techniques in HSI super-resolution. Li *et al.* [36] used a deep spectral difference convolution neural network to map the spectral difference between the LR-HSI and the HR-HSI. Yuan *et al.* [37] designed a multi-scale and multi-depth CNN in to recover the high spatial-spectral resolution image.

In existing HSI-MSI fusion approaches, the spatial and spectral data are always vectorized and processed separately. However, this class of techniques destroy the HSI's intrinsic structure when reconstructing the HR-HSI. In fact, since the original HSIs are organized in data cubes [38–40], representativeness and informativeness play an important role in real applications [41]. For this reason, several tensor-based methods have been proposed and successfully applied. In [38], Dian *et al.* proposed a nonlocal sparse tensor factorization (NLSTF) method to fuse the LR-HSI and the HR-MSI. It regards the HSI as a three-order tensor, and decomposes the tensor into a core tensor multiplication by factor matrices representing the three dimensions. Yi *et al.* [42] proposed a weighted low-rank tensor recovery (WLRTR) model that treated the singular values differently. Similar to the non-local self-similarity across space (NSS) method [43], the nonlocal similarity between spectral-spatial cubic and spectral correlation can be characterized in tensors. Among all these effective tensor representation methods, parallel factor analysis (PARAFAC) [44] model, which is also called canonical polyadic (CP) decomposition [45–47], has been justified as a powerful method to decompose a tensor data onto a few rank-one tensors in a multilinear way. In HS data analysis, the CP decomposition has been widely used because it can describe the low-rank structure in HS data. Veganzones *et al.* used a novel compression-based nonnegative CP decomposition to the blind analysis of hyperspectral big data [39], and

proposed a nonnegative unmixing method to estimate multi-angle bidirectional reflectance (BRF) data [48]. Fang *et al.* [49] proposed an effective CP tensor HS compression method with the observed HSI characterized by a three-order tensor. Xian *et al.* [50] used the CP decomposition to extract the rank-1 component in HSI and combined the signal rank-1 profiles to denoising.

Inspired by the tensor decomposition of HSI, we propose a novel nonlocal coupled tensor CP decomposition (NCTCP) HSI-MSI fusion method. In the proposed method, we assume that the nonlocal tensor is composed of the similar nonlocal patch cubes in HSI that lies in a low dimensional subspace and can be regarded as a low-rank tensor. Thus, we use the CP decomposition to characterize the low-rank structure. According to the linear degradation model of the LR-HSI and the HR-MSI, nonlocal patch tensors in the LR-HSI and the HR-MSI can also be represented using CP decomposition. Besides, the factor matrices representing the spatial mode and nonlocal similarities mode are shared between the HR-HSI and the HR-MSI. The relationship between shared factor matrices can be used to transfer the spatial information of the HR-MSI to the HR-HSI. In addition, the clustering of the nonlocal patch cubes is a key factor to ensure the constructed nonlocal tensor low rank. In this paper, we use the HR-MSI to compute a smooth order of all the nonlocal patch cubes. Then, consecutive nonlocal patch cubes in the smooth order are extracted to form the patch tensor. Experimental results demonstrate that the proposed fusion method substantially outperforms existing state-of-the-art HSI-MSI fusion methods.

The remainder of this paper is organized as follows. Section II formulates the problem of HSI-MSI fusion and introduces the basic notations of tensors. Section III presents the proposed NCTCP method for HSI-MSI fusion. The optimization algorithm of the proposed model is detailed in Section IV. Results on three synthetic datasets and one real dataset are presented in Section V. Finally, Section VI concludes the whole paper with some remarks.

## II. TENSOR NOTATIONS AND PROBLEM FORMULATION

### A. Tensor Notations

A tensor is a multi-dimensional data array denoted by Euler Script letters, e.g.,  $\mathcal{X}$ . A slice is a two-dimensional subtensor where all but two indices in a tensor are fixed. A fiber is a one-dimensional subtensor where every index but one is fixed. For a three-order tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ ,  $\mathcal{X}(k, :, :)$ ,  $\mathcal{X}(:, k, :)$  and  $\mathcal{X}(:, :, k)$  represent the horizontal, lateral and frontal slices separately.  $\mathcal{X}(:, i, j)$ ,  $\mathcal{X}(i, :, j)$  and  $\mathcal{X}(i, j, :)$  represent the  $(i, j)_{th}$  mode-1, mode-2 and mode-3 fiber, respectively. For sake of simplicity, we use  $\mathcal{X}^{(k)}$  to represent  $\mathcal{X}(:, :, k)$ . The inner product of  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{I_1}$  is  $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$ . The inner product of  $\mathbf{X}$  and  $\mathbf{Y} \in \mathbb{R}^{I_1 \times I_2}$  is  $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{Tr}(\mathbf{X}^T \mathbf{Y})$ , where  $\text{Tr}(\cdot)$  represents the matrix trace. The inner product of  $\mathcal{X}$  and  $\mathcal{Y}$  in  $\mathbb{R}^{I_1 \times I_2 \times I_3}$  is defined as  $\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i=1}^{I_3} \langle \mathcal{X}^{(i)}, \mathcal{Y}^{(i)} \rangle$ . Using the inner product of two tensors, the Frobenius norm of a tensor can be represented as  $\|\mathcal{X}\|_F = \langle \mathcal{X}, \mathcal{X} \rangle$ . The mode- $i$  unfolding of  $\mathcal{X}$  is a matrix denoted by  $\mathbf{X}_{(i)}$ , which can be obtained by rearranging

(lexicographically in the indices other than the  $i_{th}$  index) the mode- $i$  fibers as the columns of the matrix. For convenience, we use  $unfold_i(\mathcal{X}) = \mathbf{X}_{(i)}$  and  $\mathcal{X} = fold_i(\mathbf{X}_{(i)})$  to represent *unfolding* and *folding*, respectively. The  $n$ -mode product of a tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  and a matrix  $\mathbf{U} \in \mathbb{R}^{M \times I_n}$  is  $\mathcal{X} \times_n \mathbf{U}$ , which is of size  $I_1 \times \dots \times I_{n-1} \times M \times I_{n+1} \times \dots \times I_N$ . To represent  $n$ -mode product using elements in tensor and matrix, we have

$$\begin{aligned} & \mathcal{X} \times_n \mathbf{U}(i_1, \dots, i_{n-1}, m, i_{n+1}, \dots, i_N) \\ &= \sum_{i_n} \mathcal{X}(i_1, \dots, i_{n-1}, i_n, i_{n+1}, \dots, i_N) \mathbf{U}(m, i_n). \end{aligned} \quad (1)$$

Alternatively, the  $n$ -mode product can also be computed by matrix multiplication

$$\mathcal{Y} = \mathcal{X} \times_n \mathbf{U} \Leftrightarrow \mathbf{Y}_{(n)} = \mathbf{U} \mathbf{X}_{(n)}. \quad (2)$$

The CP decomposition decomposes a tensor into a sum of component rank-one tensors [51, 52]. For a tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ , the CP decomposition is defined as

$$\mathcal{X} = \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \quad (3)$$

where  $\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_R] \in \mathbb{R}^{n_1 \times R}$ ,  $\mathbf{B} = [\mathbf{b}_1 \dots \mathbf{b}_R] \in \mathbb{R}^{n_2 \times R}$ , and  $\mathbf{C} = [\mathbf{c}_1 \dots \mathbf{c}_R] \in \mathbb{R}^{n_3 \times R}$  are called factor matrices.  $R$  is the rank of the tensor  $\mathcal{X}$ , and is defined as the minimal number of rank-one components [51]. The symbol  $\circ$  denotes the usual outer product of two vectors,  $\mathbf{a} \circ \mathbf{b} = \mathbf{ab}^T$ . Accordingly, each entry of the tensor can be determined by

$$\mathcal{X}(i, j, k) = \sum_{r=1}^R \mathbf{A}(i, r) \mathbf{B}(j, r) \mathbf{C}(k, r). \quad (4)$$

Usually the observed noisy version of  $\mathcal{X}$  can be written as

$$\tilde{\mathcal{X}} = \mathcal{X} + \mathcal{E}, \quad (5)$$

where  $\mathcal{E}$  is Gaussian independent and identically distributed (i.i.d.) elements. In order to estimate the factor matrices, we solve the following problem

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \frac{1}{2} \|\tilde{\mathcal{X}} - \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket\|_F^2. \quad (6)$$

Let  $\mathcal{W} = \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket$ . Then, we have

$$\mathbf{W}_{(1)} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T, \quad (7)$$

$$\mathbf{W}_{(2)} = \mathbf{B}(\mathbf{C} \odot \mathbf{A})^T, \quad (8)$$

$$\mathbf{W}_{(3)} = \mathbf{C}(\mathbf{B} \odot \mathbf{A})^T \quad (9)$$

where  $\odot$  is the Khatri-Rao (columnwise Kronecker) product. The objective function can also be written as

$$\begin{aligned} \frac{1}{2} \|\tilde{\mathcal{X}} - \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket\|_F^2 &= \frac{1}{2} \|\tilde{\mathbf{X}}_{(1)} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T\|_F^2 \\ &= \frac{1}{2} \|\tilde{\mathbf{X}}_{(2)} - \mathbf{B}(\mathbf{C} \odot \mathbf{A})^T\|_F^2 \\ &= \frac{1}{2} \|\tilde{\mathbf{X}}_{(3)} - \mathbf{C}(\mathbf{B} \odot \mathbf{A})^T\|_F^2. \end{aligned} \quad (10)$$

These equations are the fundamental formulas for alternative least-squares (ALS) [53] CP optimization. We can update any one matrix as long as the other two matrices are determined, by solving a simple linear least-squares problem.

## B. HSI-MSI Fusion Problem

Denote the real HR-HSI as  $\mathcal{X} \in \mathbb{R}^{W \times H \times B}$ , where  $W$ ,  $H$ , and  $B$  are the size of the image width, image height, and spectral band number, respectively.  $\tilde{\mathcal{X}} \in \mathbb{R}^{w \times h \times B}$  represents the observed LR-HSI with the same spectral band- $B$ .  $\tilde{\mathcal{X}}$  can be regarded as the spatial downsampled version of  $\mathcal{X}$ . Here, we assume  $W > w$  and  $H > h$ .  $\mathcal{Y} \in \mathbb{R}^{W \times H \times l}$  is the HR-MSI capturing the same scene with the the same spatial dimension, and  $l$  is the band number of MSI. The goal of HSI-MSI fusion is to estimate the real HR-HSI  $\mathcal{X}$  by fusing  $\tilde{\mathcal{X}}$  with  $\mathcal{Y}$ . In the conventional way, the LR-HSI acquisition process can be modeled as

$$\tilde{\mathbf{X}}_{(3)} = \mathbf{X}_{(3)} \mathbf{S} \mathbf{H} + \mathbf{E}_{\mathbf{h}(3)}, \quad (11)$$

where  $\mathbf{H} \in \mathbb{R}^{WH \times wh}$  represents the spatial down-sampling operator,  $\mathbf{S} \in \mathbb{R}^{WH \times WH}$  represents the spatial blurring operator, and  $\mathbf{E}_{\mathbf{h}(3)}$  represents the independent and identically distributed (i.i.d.) noise of the LR-HSI. The acquisition of the HR-MSI is

$$\mathbf{Y}_{(3)} = \mathbf{R} \mathbf{X}_{(3)} + \mathbf{E}_{\mathbf{m}(3)}, \quad (12)$$

where  $\mathbf{R} \in \mathbb{R}^{l \times B}$  is the spectral responses of the multispectral sensor, and  $\mathbf{E}_{\mathbf{m}(3)}$  represents the i.i.d. noise of the HR-MSI. The maximum likelihood estimation of  $\mathbf{X}_{(3)}$  is obtained by

$$\mathbf{X}_{(3)} = \arg \min \{ \|\tilde{\mathbf{X}}_{(3)} - \mathbf{X}_{(3)} \mathbf{S} \mathbf{H}\|_F^2 + \lambda \|\mathbf{Y}_{(3)} - \mathbf{R} \mathbf{X}_{(3)}\|_F^2 \} \quad (13)$$

where  $\lambda$  is a parameter for balancing the relative importance of two terms. However, the problem defined in Eq. (13) is ill-posed. To obtain the HR-HSI, additional prior knowledge about the real solution is required. Many algorithms have been developed to solve this ill-posed problem by introducing additional regularization, such as vectorized total variation (TV) [25], subspace learning [23, 54], and unmixing [30, 55].

## III. PROPOSED NCTCP APPROACH

In this section, we introduce the NCTCP HSI-MSI fusion approach. We first introduce the tensor representation for HSI-MSI fusion. The nonlocal coupled tensor CP decomposition model for HSI-MSI fusion is subsequently proposed.

### A. Tensor Representation for HSI-MSI Fusion

HR-HSI is a 3-order tensor  $\mathcal{X}$ . To express the acquisition process of LR-HSI in the tensors, we introduce the symbol  $\mathcal{H}$  and  $\mathcal{S}$  as the downsampling operator and spatial blurring operator on tensors which have the same effect as  $\mathbf{H}$  and  $\mathbf{S}$  on  $\mathbf{X}_{(3)}$ . Therefore, the acquisition process of LR-HSI can be written as

$$\tilde{\mathcal{X}} = \mathcal{X} \mathcal{S} \mathcal{H} + \mathcal{E}_{\mathbf{h}}. \quad (14)$$

According to Eqs. (2) and (12), the tensor representation for the acquisition of HR-MSI is

$$\mathcal{Y} = \mathcal{X} \times_3 \mathbf{R} + \mathcal{E}_{\mathbf{m}} \quad (15)$$

where  $\mathcal{E}_{\mathbf{h}}$  and  $\mathcal{E}_{\mathbf{m}}$  represent i.i.d. noises.

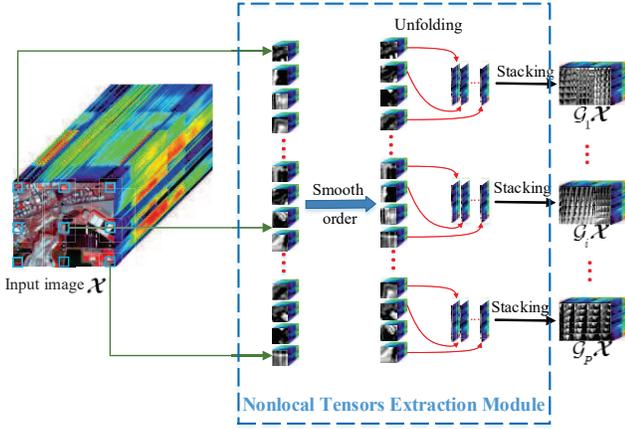


Fig. 1. Procedure of nonlocal tensors extraction module. Here the sequence of smooth ordering and unfolding can be swapped.

### B. Nonlocal Coupled Tensor CP Decomposition Model for HSI-MSI Fusion

The nonlocal tensor based methods have been used in image processing community [56, 57]. Nonlocal similarity suggests that one patch may have many patches with similar structure within the same image. The similarity of the nonlocal patches indicates that these patches are correlative. Recent work has extended this nonlocal similarity to 3D case [43, 58, 59]. For HSI-MSI fusion, the HR-HSI  $\mathcal{X}$  is separated into a group of overlapping local 3D cubes  $\{\mathcal{P}_{ij}\}_{1 \leq i \leq W-d_w, 1 \leq j \leq H-d_h} \subset \mathbb{R}^{d_w \times d_h \times B}$ .  $d_w$  and  $d_h$  represent the width and height of a 3D cube, respectively. We unfold each cube to its matrix format by ordering each band as a column vector lexicographically. Now all the 3D cubes are converted into a set of 2D patches  $\Omega = \{\mathcal{P}_i \in \mathbb{R}^{d_w d_h \times B}\}_1^N$ , where  $N = (W - d_w + 1) \times (H - d_h + 1)$  is the number of patches over the whole HR-HSI. Usually the 2D patch set  $\Omega$  is clustered into  $K$  clusters. We can represent each cluster as a 3-order tensor. Define  $\mathbf{D}_p^m \{\cdot\}$  as the operator extracting the  $m$ -th 2D patch of size  $d_w d_h \times B$  in the  $p$ -th cluster, then denote  $\mathcal{G}_p \mathcal{X}$  as

$$\mathcal{G}_p \mathcal{X} := (\mathbf{D}_p^1 \{\mathcal{X}\}, \mathbf{D}_p^2 \{\mathcal{X}\}, \dots, \mathbf{D}_p^{N_p} \{\mathcal{X}\}) \in \mathbb{R}^{d_w d_h \times N_p \times B}, \quad (16)$$

where  $N_p$  is the total number of nonlocal patches in the  $p$ -th cluster. Here, the sequence of smooth ordering and unfolding 3D cubes to 2D patches can be changed. Fig. 1 illustrates the procedure of nonlocal tensors extraction module from a HSI. Since the patches in the same cluster have similar spatial-spectral structures,  $\mathcal{G}_p \mathcal{X}$  can be seen as a low-rank tensor. In this situation, CP decomposition of  $\mathcal{G}_p \mathcal{X}$  is an appropriate model and can be written as

$$\mathcal{G}_p \mathcal{X} = [\mathbf{A}_p, \mathbf{B}_p, \mathbf{C}_p]. \quad (17)$$

Then, the original  $\mathcal{X}$  can be expressed as

$$\mathcal{X} = \left( \sum_p \mathcal{G}_p^T \mathcal{G}_p \right)^{-1} \sum_p \mathcal{G}_p^T [\mathbf{A}_p, \mathbf{B}_p, \mathbf{C}_p]. \quad (18)$$

This equation denotes that the ideal  $\mathcal{X}$  is calculated by summing over all clusters and averaging the results.

Analogously,  $\mathcal{G}_p \mathcal{Y}$  represents the corresponding patch tensor of the HR-MSI. Since this tensor format does not change the spectral dimension, the acquisition of  $\mathcal{G}_p \mathcal{Y}$  can be written as

$$\begin{aligned} \mathcal{G}_p \mathcal{Y} &= \mathcal{G}_p \mathcal{X} \times_3 \mathbf{R} + \mathcal{G}_p \mathcal{E}_m \\ &= [\mathbf{A}_p, \mathbf{B}_p, \mathbf{C}_p] \times_3 \mathbf{R} + \mathcal{G}_p \mathcal{E}_m \\ &= [\mathbf{A}_p, \mathbf{B}_p, \mathbf{RC}_p] + \mathcal{G}_p \mathcal{E}_m. \end{aligned} \quad (19)$$

Thus, we can use nonlocal tensors and CP decomposition to estimate  $\mathcal{X}$ , by solving the following problem

$$\begin{aligned} \min_{\mathcal{X}, \mathbf{A}_p, \mathbf{B}_p, \mathbf{C}_p} \{ &\|\tilde{\mathcal{X}} - \mathcal{X} \mathcal{S} \mathcal{H}\|_F^2 + \lambda \sum_{p=1}^P \|\mathcal{G}_p \mathcal{Y} - [\mathbf{A}_p, \mathbf{B}_p, \mathbf{RC}_p]\|_F^2 \} \\ \text{s.t. } &\mathcal{X} = \left( \sum_p \mathcal{G}_p^T \mathcal{G}_p \right)^{-1} \sum_p \mathcal{G}_p^T [\mathbf{A}_p, \mathbf{B}_p, \mathbf{C}_p]. \end{aligned} \quad (20)$$

In the above-mentioned model, patch tensor  $\mathcal{G}_p \mathcal{X}$  and  $\mathcal{G}_p \mathcal{Y}$  share the same factor matrices  $\mathbf{A}_p$  and  $\mathbf{B}_p$  that denote the spatial dimension structure and nonlocal patch similarities. With these two shared factor matrices, the spatial information and nonlocal similarities of the HR-MSI are transferred to the HR-HSI. Meanwhile, the spectral structure of the HR-HSI is not broken in nonlocal tensor format, indicating that the degeneration process from the HR-HSI to the HR-MSI still holds in the nonlocal tensor representation. Eq. (20) is a coupled tensor CP decomposition problem that incorporates the nonlocal tensor information. In this sense, we refer to this model as nonlocal coupled tensor CP decomposition (NCTCP).

Another unsolved problem is the clustering problem. Without HR-HSI known a priori, we cannot get exact clustering result directly. However, the HR-MSI has the same spatial resolution as the HR-HSI reflects the correlation between image patches. With that in mind, we use the HR-MSI as the guide image to determine the exact nonlocal similar patches. In conventional nonlocal approaches [43, 60], k-means or k-nn based methods have been popularly used to solve clustering problems. The main obstacle for using these methods is that, the exact number of cluster centers is not pre-specified, and the initialization of the cluster centers has a direct impact upon the clustering results. In this paper, we use an alternative method to obtain the clustering result. This method is originally used in image denoising and inpainting [61]. The core idea is to order all extracted patches such that they form a 'shortest possible path', which is essentially similar to solving the traveling salesman problem. The reordered MSI 3D patches are supposed to induce a smooth or at least piecewise smooth 1-D ordering. Thus, by simply using a limited number of consecutive patch cubes in  $\mathcal{Y}$  according to the new 1-D ordering as one cluster to form  $\mathcal{G}_p \mathcal{Y}$ , the corresponding nonlocal 3D-patches in  $\mathcal{X}$  form the corresponding 3-order nonlocal tensor  $\mathcal{G}_p \mathcal{X}$ . Fig. 2 shows the complete flowchart of the proposed fusion method. By means of this method, the extracted nonlocal information of HR-HSI is more robust and accurate, leading to a more reliable CP decomposition result.

### IV. OPTIMIZATION ALGORITHM

We apply the alternating direction method of multipliers (ADMM) [62] optimization method to solve the proposed

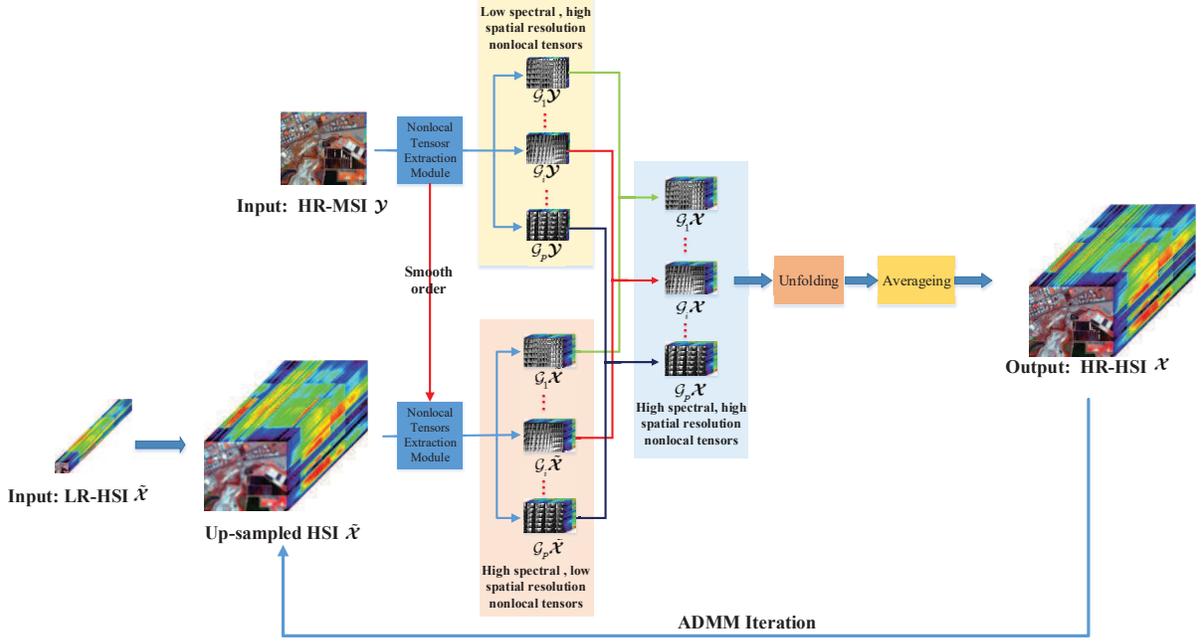


Fig. 2. Flowchart of the proposed method.

NCTCP model. First, new variables are introduced as auxiliary variables  $D_p$  which satisfy  $D_p = \mathbf{RC}_p$ . Then, the original problem can be written as

$$\begin{aligned} \min_{\mathcal{X}, \mathbf{A}_p, \mathbf{B}_p, \mathbf{C}_p} \{ & \|\tilde{\mathcal{X}} - \mathcal{X}\mathcal{S}\mathcal{H}\|_F^2 + \lambda \sum_{p=1}^P \|\mathcal{G}_p \mathcal{Y} - [\mathbf{A}_p, \mathbf{B}_p, \mathbf{D}_p]\|_F^2 \} \\ \text{s.t. } \mathcal{X} = & \left( \sum_p \mathcal{G}_p^T \mathcal{G}_p \right)^{-1} \sum_p \mathcal{G}_p^T [\mathbf{A}_p, \mathbf{B}_p, \mathbf{C}_p], \\ \mathbf{D}_p = & \mathbf{RC}_p, \quad p = 1, 2, \dots, P. \end{aligned} \quad (21)$$

The augmented Lagrangian function of Eq. (21) is

$$\begin{aligned} L_\mu(\mathcal{X}, \mathbf{A}_p, \mathbf{B}_p, \mathbf{C}_p) = & \|\tilde{\mathcal{X}} - \mathcal{X}\mathcal{S}\mathcal{H}\|_F^2 \\ & + \lambda \sum_{p=1}^P \|\mathcal{G}_p \mathcal{Y} - [\mathbf{A}_p, \mathbf{B}_p, \mathbf{D}_p]\|_F^2 \\ & + \frac{\mu}{2} \|\mathcal{X} - \left( \sum_p \mathcal{G}_p^T \mathcal{G}_p \right)^{-1} \sum_p \mathcal{G}_p^T [\mathbf{A}_p, \mathbf{B}_p, \mathbf{C}_p]\|_F^2 \\ & + \left\langle \mathcal{M}, \mathcal{X} - \left( \sum_p \mathcal{G}_p^T \mathcal{G}_p \right)^{-1} \sum_p \mathcal{G}_p^T [\mathbf{A}_p, \mathbf{B}_p, \mathbf{C}_p] \right\rangle \\ & + \frac{\mu}{2} \sum_p \|\mathbf{D}_p - \mathbf{RC}_p\|_F^2 + \sum_p \langle \mathbf{F}_p, \mathbf{D}_p - \mathbf{RC}_p \rangle \end{aligned} \quad (22)$$

where  $\mathcal{M}$  and  $\mathbf{F}_p$  are the Lagrange multipliers,  $\mu$  is the positive penalty scalar. It would be intractable to optimize all these variables simultaneously. Instead of that, we follow the strategy in ALS CP algorithm that optimizes the problem by minimizing one variable with the others fixed. The optimization problem of  $L_\mu$  can be solved by the following sub-problems:

1) Solving  $\mathcal{X}$ : Optimizing  $L_\mu$  with respect to  $\mathcal{X}$  leads to the following linear system:

$$\begin{aligned} \mathcal{X}(2\mathcal{S}\mathcal{H}\mathcal{H}^T \mathcal{S}^T + \mu \mathbf{I}) = & 2\tilde{\mathcal{X}}\mathcal{H}^T \mathcal{S}^T - \mathcal{M} \\ & + \mu \left( \left( \sum_p \mathcal{G}_p^T \mathcal{G}_p \right)^{-1} \sum_p \mathcal{G}_p^T [\mathbf{A}_p, \mathbf{B}_p, \mathbf{C}_p] \right). \end{aligned} \quad (23)$$

In this work, we use the conjugate gradient algorithm to solve this linear system.

2) Solving  $\mathbf{A}_p$ : The sub-problem of  $L_\mu$  with respect to  $\mathbf{A}_p$  ( $p = 1, 2, \dots, P$ ) can be represented as

$$\begin{aligned} \min_{\mathbf{A}_p, p=1,2,\dots,P} \{ & \lambda \sum_{p=1}^P \|\mathcal{G}_p \mathcal{Y} - [\mathbf{A}_p, \mathbf{B}_p, \mathbf{D}_p]\|_F^2 \\ & + \frac{\mu}{2} \|\mathcal{X} - \left( \sum_p \mathcal{G}_p^T \mathcal{G}_p \right)^{-1} \sum_p \mathcal{G}_p^T [\mathbf{A}_p, \mathbf{B}_p, \mathbf{C}_p] + \mathcal{M}/\mu\|_F^2 \}. \end{aligned} \quad (24)$$

Then, we convert it into the following problem:

$$\begin{aligned} \min_{\mathbf{A}_p, p=1,2,\dots,P} \{ & \lambda \sum_{p=1}^P \|\mathcal{G}_p \mathcal{Y} - [\mathbf{A}_p, \mathbf{B}_p, \mathbf{D}_p]\|_F^2 \\ & + \frac{\mu}{2} \sum_{p=1}^P \|\mathcal{G}_p \mathcal{X} - [\mathbf{A}_p, \mathbf{B}_p, \mathbf{C}_p] + \mathcal{G}_p \mathcal{M}/\mu\|_F^2 \} \end{aligned} \quad (25)$$

where  $\mathcal{G}_p \mathcal{M}$  is the  $p$ th 3-order nonlocal tensor produced by  $\mathcal{M}$ . Each  $\mathbf{A}_p$  can be solved independently. According to Eq. (7), we have

$$\begin{aligned} \min_{\mathbf{A}_p} \lambda & \|\mathcal{G}_p \mathbf{Y}_{(1)} - \mathbf{A}_p (\mathbf{D}_p \odot \mathbf{B}_p)^T\|_F^2 \\ & + \frac{\mu}{2} \|\mathcal{G}_p \mathbf{X}_{(1)} + \mathcal{G}_p \mathbf{M}_{(1)}/\mu - \mathbf{A}_p (\mathbf{C}_p \odot \mathbf{B}_p)^T\|_F^2 \end{aligned} \quad (26)$$

where  $\mathcal{G}_p \mathbf{Y}_{(1)}$ ,  $\mathcal{G}_p \mathbf{X}_{(1)}$  and  $\mathcal{G}_p \mathbf{M}_{(1)}$  represent the mode- $i$  unfolding of tensors  $\mathcal{G}_p \mathcal{Y}$ ,  $\mathcal{G}_p \mathcal{X}$  and  $\mathcal{G}_p \mathcal{M}$  respectively. Thus, we can obtain the solution by Eq. (27).

$$(\mu\mathcal{G}_p\mathbf{X}_{(1)}(\mathbf{C}_p \odot \mathbf{B}_p) + 2\lambda\mathcal{G}_p\mathbf{Y}_{(1)}(\mathbf{D}_p \odot \mathbf{B}_p) + \mathcal{G}_p\mathbf{M}_{(1)}(\mathbf{C}_p \odot \mathbf{B}_p)) \cdot (2\lambda(\mathbf{D}_p \odot \mathbf{B}_p)^T(\mathbf{D}_p \odot \mathbf{B}_p) + \mu(\mathbf{C}_p \odot \mathbf{B}_p)^T(\mathbf{C}_p \odot \mathbf{B}_p))^{-1}. \quad (27)$$

$$(\mu\mathcal{G}_p\mathbf{X}_{(2)}(\mathbf{C}_p \odot \mathbf{A}_p) + 2\lambda\mathcal{G}_p\mathbf{Y}_{(2)}(\mathbf{D}_p \odot \mathbf{A}_p) + \mathcal{G}_p\mathbf{M}_{(2)}(\mathbf{C}_p \odot \mathbf{A}_p)) \cdot (2\lambda(\mathbf{D}_p \odot \mathbf{A}_p)^T(\mathbf{D}_p \odot \mathbf{A}_p) + \mu(\mathbf{C}_p \odot \mathbf{A}_p)^T(\mathbf{C}_p \odot \mathbf{A}_p))^{-1}. \quad (28)$$

3) Solving  $\mathbf{B}_p$ : Similar to  $\mathbf{A}_p$ ,  $\mathbf{B}_p$  can be computed by Eq. (28).

4) Solving  $\mathbf{C}_p$ : The optimization problems of  $L_\mu$  with respect to  $\mathbf{C}_p$  is

$$\min_{\mathbf{C}_p} \left\{ \frac{\mu}{2} \|\mathcal{G}_p\mathbf{X}_{(3)} + \mathcal{G}_p\mathbf{M}_{(3)}/\mu - \mathbf{C}_p(\mathbf{B}_p \odot \mathbf{A}_p)^T\|_F^2 + \frac{\mu}{2} \|\mathbf{D}_p - \mathbf{R}\mathbf{C}_p + \mathbf{F}_p/\mu\|_F^2 \right\}. \quad (29)$$

Setting the derivative of the objective function to zero with respect to  $\mathbf{C}_p$ , we have the following matrix equation

$$\begin{aligned} & \mathbf{C}_p(\mathbf{B}_p \odot \mathbf{A}_p)^T(\mathbf{B}_p \odot \mathbf{A}_p) + \mathbf{R}^T\mathbf{R}\mathbf{C}_p \\ &= \mathcal{G}_p\mathbf{X}_{(3)}(\mathbf{B}_p \odot \mathbf{A}_p) + \mathcal{G}_p\mathbf{M}_{(3)}(\mathbf{B}_p \odot \mathbf{A}_p)/\mu \\ & \quad + \mathbf{R}^T\mathbf{D}_p + \mathbf{R}^T\mathbf{F}_p/\mu. \end{aligned} \quad (30)$$

This is a Sylvester matrix equation [63]. Qi *et al.* has proposed many methods to solve the Sylvester matrix equation using block permutation matrix [64] and Woodbury formula [65]. However, as the matrix is very large in their model, specific algorithms are required to cope with large-scale matrices. Thanks to the nonlocal tensor format, the Sylvester equation in the proposed model is of small scale. We can use the Bartels-Stewart algorithm [63] to obtain the solution. This algorithm decomposes  $(\mathbf{B}_p \odot \mathbf{A}_p)^T(\mathbf{B}_p \odot \mathbf{A}_p)$  and  $\mathbf{R}^T\mathbf{R}$  into Schur forms using a QR algorithm. Thus, back-substitution can be used to solve the triangular system.

5) Solving  $\mathbf{D}_p$ : The optimization problem of  $L_\mu$  with respect to  $\mathbf{D}_p$  is

$$\min_{\mathbf{D}_p} \left\{ \lambda \|\mathcal{G}_p\mathbf{Y}_{(3)} - \mathbf{D}_p(\mathbf{B}_p \odot \mathbf{A}_p)^T\|_F^2 + \frac{\mu}{2} \|\mathbf{D}_p - \mathbf{R}\mathbf{C}_p + \mathbf{F}_p/\mu\|_F^2 \right\}. \quad (31)$$

So  $\mathbf{D}_p$  is computed by

$$\begin{aligned} & (2\lambda\mathcal{G}_p\mathbf{Y}_{(3)}(\mathbf{B}_p \odot \mathbf{A}_p) + \mu\mathbf{R}\mathbf{C}_p - \mathbf{F}_p) \cdot \\ & (2\lambda(\mathbf{B}_p \odot \mathbf{A}_p)^T(\mathbf{B}_p \odot \mathbf{A}_p) + \mu\mathbf{I})^{-1} \end{aligned} \quad (32)$$

6) Multipliers update:

$$\mathcal{M} \leftarrow \mathcal{M} + \mu \left( \mathcal{X} - \left( \left( \sum_p \mathcal{G}_p^T \mathcal{G}_p \right)^{-1} \sum_p \mathcal{G}_p^T [\mathbf{A}_p, \mathbf{B}_p, \mathbf{C}_p] \right) \right). \quad (33)$$

$$\mathbf{F}_p \leftarrow \mathbf{F}_p + \mu(\mathbf{D}_p - \mathbf{R}\mathbf{C}_p), \quad p = 1, 2, \dots, P. \quad (34)$$

$\mu$  is updated by

$$\mu \leftarrow \min(\rho\mu, \mu_{\max}), \quad (35)$$

where  $\rho \geq 1$ . The proposed NCTCP algorithm for HSI-MSI fusion can be summarized in Algorithm 1.

---

#### Algorithm 1 NCTCP for HSI-MSI fusion

---

**Input:** LR- $\mathbf{X}$ , HR- $\mathbf{Y}$ ,  $\mathcal{H}$ ,  $\mathbf{S}$ ,  $\mathbf{R}$ ,  $\lambda$ ,  $d_w$ ,  $d_h$

- 1: Initial  $\mathcal{X}^{(0)} = \tilde{\mathcal{X}}\mathcal{H}^T\mathbf{S}^T$ ,  $\mathcal{M}^{(0)} = \mathbf{0}$ ,  $\mathbf{F}_p^{(0)} = \mathbf{0}$ ,  $\mathbf{A}_p^{(0)}$ ,  $\mathbf{B}_p^{(0)}$  and  $\mathbf{C}_p^{(0)}$ ,  $\mathbf{D}_p^{(0)}$ , ( $p = 1, 2, \dots, P$ ) are randomly initialized,  $k = 1$ ,  $\mu = 10^{-4}$ ,  $\rho = 1.01$ ,  $maxIter = 10$ , and  $\epsilon = 10^{-3}$
  - 2: **while** not convergence and  $k < maxIter$  **do**
  - 3:   **for**  $p = 1 : P$  **do**
  - 4:     Update  $\mathbf{A}_p^{(k)}$  by (27)
  - 5:     Update  $\mathbf{B}_p^{(k)}$  by (28)
  - 6:     Update  $\mathbf{C}_p^{(k)}$  by solving Problem (30)
  - 7:     Update  $\mathbf{D}_p^{(k)}$  by (32)
  - 8:   **end for**
  - 9:   Update  $\mathcal{X}^{(k)}$  by (23)
  - 10:   Update the multipliers  $\mathcal{M}^{(k)}$ ,  $\mathbf{F}_p^{(k)}$ ,  $p = 1, 2, \dots, P$  and let  $\mu = \rho\mu$
  - 11:   Check the convergence conditions
 
$$\|\mathcal{X}^{(k)} - (\sum_p \mathcal{G}_p^T \mathcal{G}_p)^{-1} \sum_p \mathcal{G}_p^T [\mathbf{A}_p^{(k)}, \mathbf{B}_p^{(k)}, \mathbf{C}_p^{(k)}]\|_F^2 < \epsilon,$$

$$\|\mathcal{X}^{(k)} - \mathcal{X}^{(k-1)}\|_F^2 < \epsilon,$$

$$\|(\sum_p \mathcal{G}_p^T \mathcal{G}_p)^{-1} \sum_p \mathcal{G}_p^T [\mathbf{A}_p^{(k)}, \mathbf{B}_p^{(k)}, \mathbf{C}_p^{(k)}] - (\sum_p \mathcal{G}_p^T \mathcal{G}_p)^{-1} \sum_p \mathcal{G}_p^T [\mathbf{A}_p^{(k-1)}, \mathbf{B}_p^{(k-1)}, \mathbf{C}_p^{(k-1)}]\|_F^2 < \epsilon$$
  - 12:    $k \leftarrow k + 1$
  - 13: **end while**
- Output:** HR- $\mathbf{X}$   $\mathcal{X}^{(k)}$
- 

## V. EXPERIMENTS

In this section, we describe the datasets that were used in the experimental study as well as the quality measures used for evaluating the similarity between fused and reference images. The experimental results by the proposed method are later presented and compared with other competing fusion methods.

### A. Datasets

Similar to [11], we generate the datasets following Wald's protocol. Three real-world HSIs are used as reference images. In what follows, detailed information of the datasets are presented.

1) Moffett field dataset<sup>1</sup>: This image presents a mixed urban/rural scene. It is a high spectral resolution image of size  $395 \times 185 \times 176$  acquired over Moffett field, CA, in 1994 by the JPL/NASA airborne visible/infrared imaging spectrometer (AVIRIS). After removing the water vapor absorption bands, ranging from  $0.4\mu m$  to  $2.5\mu m$ , from the original 224 bands, 176 bands were preserved in the final dataset.

<sup>1</sup><http://openremotesensing.net/knowledgebase/hyperspectral-pansharpening-a-review/>

2) University of Pavia dataset<sup>2</sup>: This image is captured over Pavia, Italy, by the reflective optics system imaging spectrometer (ROSIS). After omitting the water vapor absorption bands, 103 bands out of the original 115 bands were preserved. Thus, the final size of the reference image is  $610 \times 340 \times 103$ .

3) Washington DC Mall dataset<sup>3</sup>: This image was taken over the National Mall in Washington, D.C., in 1995. It has 210 bands within the range between  $0.4\mu\text{m}$  and  $2.5\mu\text{m}$  captured by the HYDICE sensor. Bands in the  $0.9$  and  $1.4\mu\text{m}$  region where the atmosphere is opaque have been omitted from the data set. The final size of the image is  $1280 \times 307 \times 191$ . We select a  $420 \times 300$ -pixel-size image as the reference image.

## B. Quality Measures

To make a numerical comparison of the proposed method, we use the peak signal-noise-ratio (PSNR), the averaged spectral angle mapper (SAM), the relative dimensionless global error in synthesis (ERGAS), and the cross correlation (CC) as the quality measures. To explain these quality measures, we denote the reference HS image as  $\mathbf{X} \in \mathbb{R}^{B \times N}$  where  $B$  represents the spectral bands and  $N$  is the numbers of pixels.  $\hat{\mathbf{X}} = [\mathbf{x}_1, \dots, \mathbf{x}_B]^T = [\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_B]^T$ . Here,  $\mathbf{x}_i \in \mathbb{R}^{N \times 1}$  is the  $i$ th band ( $i = 1, \dots, B$ ) and  $\mathbf{x}_j \in \mathbb{R}^{B \times 1}$  is the feature of the  $j$ th pixel ( $j = 1, \dots, N$ ).  $\hat{\mathbf{X}}$  denotes the fused HSI.

1) PSNR: The PSNR can reveal the spatial quality of the reconstructed image. The definition for the  $i$ th band is

$$\text{PSNR}(\mathbf{x}_i, \hat{\mathbf{x}}_i) = 10 \cdot \log_{10} \left( \frac{\max(\mathbf{x}_i)^2}{\|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2^2} \right) \quad (36)$$

where  $\max(\mathbf{x}_i)$  represents the maximum pixel intensity in the  $i$ th image of the reference HSI. The larger the PSNR measure is, the better quality the reconstructed image has. The PSNR measure average over all bands is used to measure the quality of the entire fused image.

2) SAM: SAM index is used to evaluate the quality of the spectral structure. It is calculated by the angle among two spectral curves of the fused and reference image. The SAM index of a given pixel is calculated by

$$\text{SAM}(\mathbf{x}_j, \hat{\mathbf{x}}_j) = \arccos \left( \frac{\mathbf{x}_j^T \hat{\mathbf{x}}_j}{\|\mathbf{x}_j\|_2 \|\hat{\mathbf{x}}_j\|_2} \right). \quad (37)$$

SAM value close to zero indicate high spectral quality. We use the average SAM index of all the pixels as the spectral measurement for the fused image.

3) ERGAS: ERGAS shows a global indication of the fusion quality. The definition of ERGAS is

$$\text{ERGAS}(\mathbf{X}, \hat{\mathbf{X}}) = 100d \sqrt{\frac{1}{B} \sum_{i=1}^B \frac{\|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2^2}{(\frac{1}{P} \mathbf{1}_p^T \mathbf{x}_i)^2}} \quad (38)$$

where  $d$  is the downsampling factor and  $\mathbf{1}_p = [1, \dots, 1] \in \mathbb{R}^{N \times 1}$ .

4) CC: This index characterizes the geometric distortion, which is defined as

$$\text{CC}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{B} \sum_{i=1}^B \text{CCS}(\mathbf{x}_i, \hat{\mathbf{x}}_i), \quad (39)$$

Here, CCS represents the cross correlation of a single-band image, calculated by

$$\text{CCS}(\mathbf{x}_i, \hat{\mathbf{x}}_i) = \frac{\sum_{j=1}^P (\mathbf{x}_i^j - \mu)(\hat{\mathbf{x}}_i^j - \hat{\mu})}{\sqrt{\sum_{j=1}^P (\mathbf{x}_i^j - \mu)^2 \sum_{j=1}^P (\hat{\mathbf{x}}_i^j - \hat{\mu})^2}} \quad (40)$$

where,  $\mathbf{x}_i^j$  is the  $j$ th element of  $\mathbf{x}_i$ ,  $\mu = (1/N) \sum_{j=1}^N \mathbf{x}_i^j$  is the sample mean of  $\mathbf{x}_i$ . The ideal value of CC is 1.

## C. Detailed Implementation

In experiments, we compared our result with other state-of-the-art methods, including coupled nonnegative matrix factorization (CNMF) [29], HySure [25], Naive Bayes, Sparse Bayes [23] and the coupled sparse tensor factorization (CSTF) [66]. We implement these competing methods using their published codes and the default parameters. Specifically, the spatial degeneration operators used in CSTF are represented according to the Gaussian kernel. Besides,  $n_w$  and  $n_h$  represent the number of atoms in the spatial dictionaries. These two parameters are decided according to the size of the HSI. In our experiments, we set  $n_w = 320$ ,  $n_h = 160$  for the Moffett field dataset,  $n_w = 600$ ,  $n_h = 300$  for the University of Pavia dataset, and  $n_w = 400$ ,  $n_h = 300$  for the Washington DC Mall dataset. For each reference image, a 4-band HR-MSI is obtained by averaging the bands of the reference image according to the spectral response profiles of the R, G, B, and NIR bands of the IKONOS multispectral sensor. To obtain the LR-HSI, we first blur the reference image band by band with a Gaussian kernel. Then, we downsample the blurred image by selecting one pixel from  $d \times d$  pixels.  $d$  is the downsampling factor which is set to 5 in our experiments. According to literature [11], the size of the blurring kernel and its standard deviation are set to  $9 \times 9$  and 2.12, respectively.

For our proposed method, we set the balance parameter  $\lambda = 100$ . During the initialization procedure, we set the multipliers  $\mathcal{M} = 0$ ,  $\mathcal{F}_p = 0$  and  $\mu = 10^{-4}$ . The factor matrices are all randomly initialized. The auxiliary parameter  $\rho$  and the maximum iteration number are set to 1.01 and 10, respectively. The window size of each patch is  $d_w = d_h = 5$ . Besides, we set the rank  $R = 80$  and the number of patches in one cluster  $N_p = 300$ . The sensitivity of the two parameters will be analyzed in the following set of experiments.

## D. Experimental Results

Fig. 3 shows the results of various methods for the Moffett field dataset. We show the false color images which are produced by the 70th, 30th and 5th band images in Fig. 3(b). The false color image of HR-MSI is shown in Fig. 3(c). For comparison purposes, the results by the competing methods are shown in Fig. 3(d)-Fig. 3(h). Visually, all the methods can successfully recover the spatial details by fusing the HR-MSI

<sup>2</sup>[http://www.ehu.es/ccwintco/index.php?title=Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes)

<sup>3</sup><https://engineering.purdue.edu/biehl/MultiSpec/hyperspectral.html>

, but there exist some areas that are not recovered well by different methods. For example, CNMF and HySure produce overly smooth spatial structures in the middle-right farmland region. Bayes methods produce some spectral distortions in the up-middle region of the scene. There are some rectangle structures in the result of CSTF. This observation is because that, the CSTF method models the whole HSI as a low-rank tensor without taking into account nonlocal similarities. Specifically, the proposed method gives better spatial structures and spectral characteristics than other methods do by using the nonlocal tensor decomposition. To evaluate the performance of the proposed method in terms of spectral preservation, Fig. 4(a)-(c) show the spectral curves of 3 pixels representing 3 different classes. We can observe that the spectral curves produced by our proposed method are closest to the reference curves. In Fig. 4(d), the comparisons of PSNR measure in every band are shown. Our proposed method can achieve the highest PSNRs in most bands of interest.

Furthermore, the quantitative results are shown in TABLE I. The best values are labeled in bold. As reported in [67], the methods used for comparison achieved promising performances in terms of reconstruction quality. However, the comparison results in TABLE I demonstrate that the proposed method outperforms them in terms of both spectral and spatial measures.

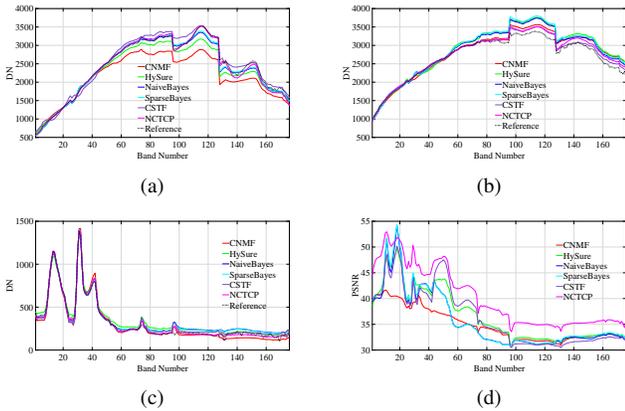


Fig. 4. Spectral curves comparisons of pixels in 3 different classes and the PSNR comparison in each band of Moffet field dataset. (a)-(c) Spectral curves of 3 pixels representing 3 classes. (d) PSNR comparison in each band.

TABLE I

PERFORMANCES OF THE HSI-MSI FUSION METHODS ON MOFFET FIELD: PSNR, ERGAS, SAM, CC

Method	PSNR	ERGAS	SAM	CC
CNMF	35.90	3.76	6.86	0.98
HySure	37.50	3.33	6.43	0.98
NaiveBayes	36.73	2.73	5.64	0.99
SparseBayes	36.88	2.62	5.62	0.99
CSTF	37.35	3.67	8.30	0.98
NCTCP	<b>41.21</b>	<b>2.36</b>	<b>4.61</b>	<b>0.99</b>

The results for the University of Pavia dataset are shown in Fig. 5 and Table II. In this dataset, we have achieved the best performances in all quality measures except PSNR.

SparseBayes is 0.65 dB higher than the proposed NCTCP in PSNR. This result denotes that the Bayes and CSTF methods perform slightly better in preserving the spatial structure. However, the proposed NCTCP method is more efficient in preserving the spectral information. Additionally, the spectral curves of 3 pixels belonging to 3 different classes are shown in Fig. 6(a)-(c). In the first 80 bands, the spectral curves of the proposed method match well with the reference curves. In the last few bands, other methods cannot well approximate the reference curves, but the proposed method generates spectral curves closest to the reference curves. In Fig. 6(d), the PSNR comparison in all bands is shown. The Bayes-based methods achieve higher PSNRs than the proposed method does between Band 10 to Band 60. With proper prior, the Bayes method can well preserve the spatial structure. However, the proposed method can achieve the best spectral preservation with promising spatial preservation.

TABLE II

PERFORMANCES OF THE HSI-MSI FUSION METHODS ON UNIVERSITY OF PAVIA: PSNR, ERGAS, SAM, CC

Method	PSNR	ERGAS	SAM	CC
CNMF	31.48	3.84	5.85	0.92
HySure	36.61	3.34	5.41	0.92
NaiveBayes	37.76	3.21	5.99	0.93
SparseBayes	<b>37.80</b>	3.17	5.93	0.93
CSTF	37.19	2.95	6.06	0.94
NCTCP	37.15	<b>2.36</b>	<b>4.38</b>	<b>0.96</b>

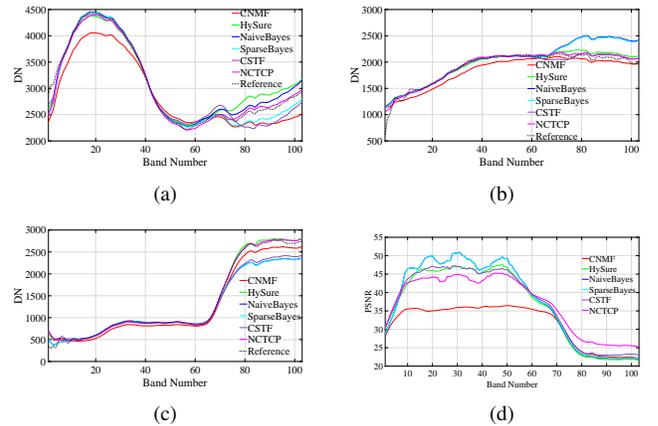


Fig. 6. Spectral curves comparisons of pixels in 3 different classes and the PSNR comparison in each band of University of Pavia dataset. (a)-(c) Spectral curves of 3 pixels representing 3 classes. (d) PSNR comparison in each band.

The last dataset is Washington DC Mall. The results are shown in Fig. 7 and Table III. In this dataset, we can observe that the proposed NCTCP method performs the best in all quality measures. The Bayes methods are better than the other 3 methods, but are 1.6dB lower than our method in terms of PSNR. CNMF, HySure and CSTF show poor performances in preserving the spatial structures. 3 pixels' spectral curves belonging to different classes are shown in Fig. 8(a)-(c). In Figs. 8(a) and 8(b), the spectral curves of the proposed method are closest to the reference spectral curves. For a few

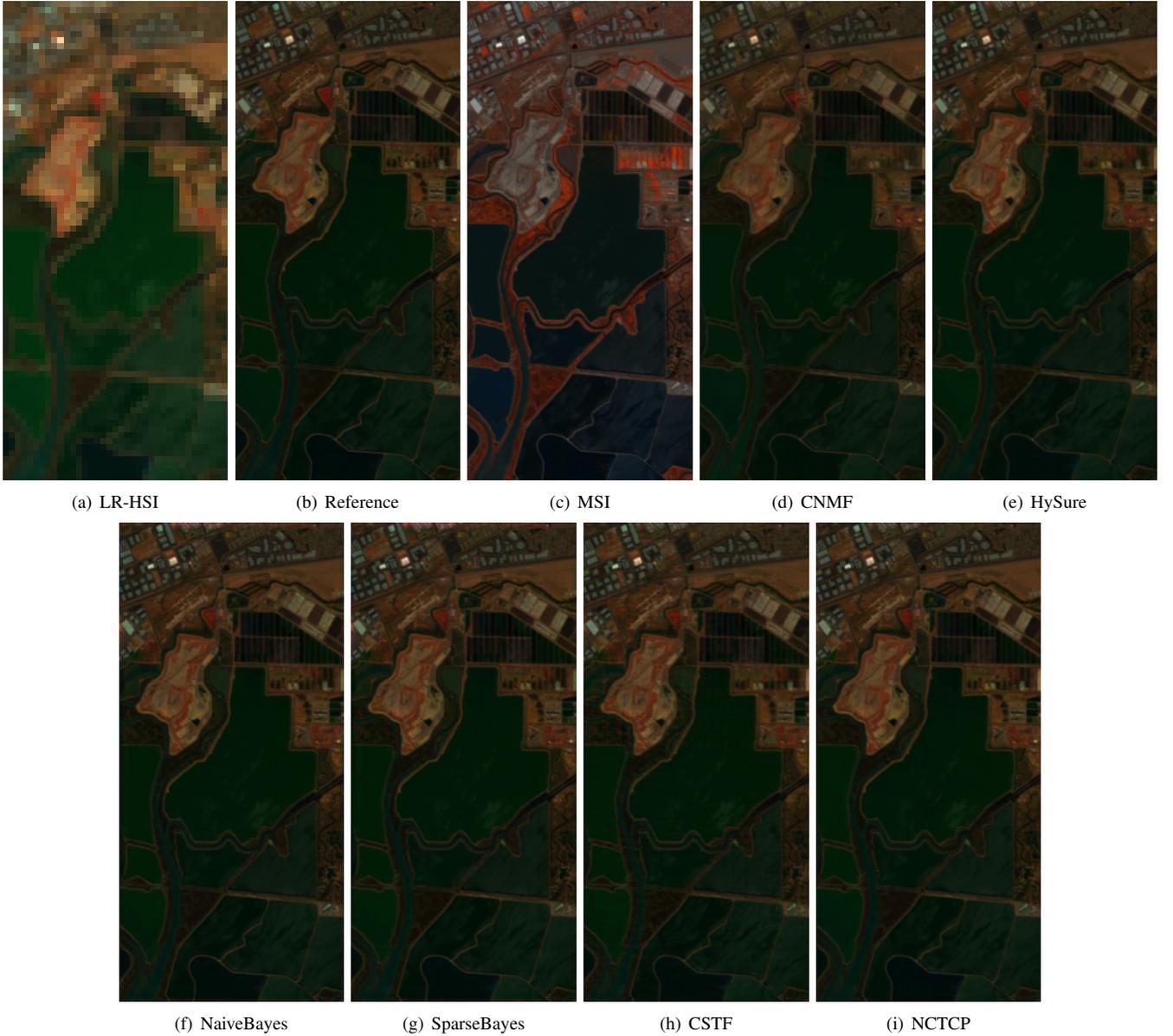


Fig. 3. HSI-MSI fusion results (Moffet field dataset).

bands in Fig. 8(c), HySure and Bayes methods outperform the proposed method. Referring to the PSNR comparison in Fig. 8(d), the proposed method obtains the highest PSNRs in most bands, indicating that both spatial and spectral structures of Washington DC Mall dataset are well preserved by the proposed method.

Through the experiments on the three synthetic datasets, we can conclude that the proposed NCTCP method achieves best HSI-MSI fusion results among all competing methods.

In order to further justify the advantage of the introduced relationship between the HR-MSI and the HR-HSI, we compared our results with the results obtained by the method without using HR-MSI. TABLE IV shows the quality measures of the two methods in all datasets. Since the proposed method significantly outperforms the method without using HR-MSI, we draw the conclusion that the introduced relationship between the HR-MSI and the HR-HSI is beneficial for improving the

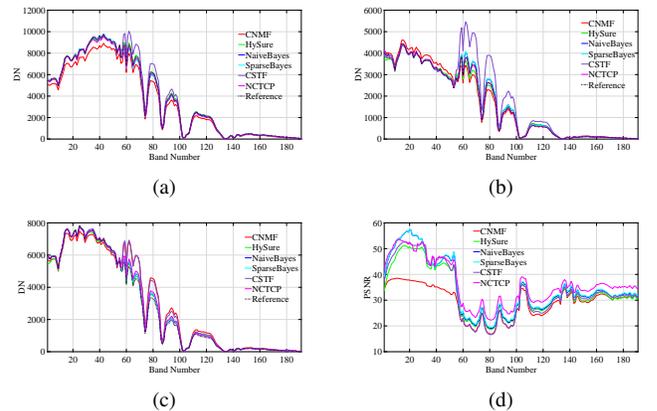


Fig. 8. Spectral curves comparisons of pixels in 3 different classes and the PSNR comparison in each band of Washington DC Mall dataset. (a)-(c) Spectral curves of 3 pixels representing 3 classes. (d) PSNR comparison in each band.

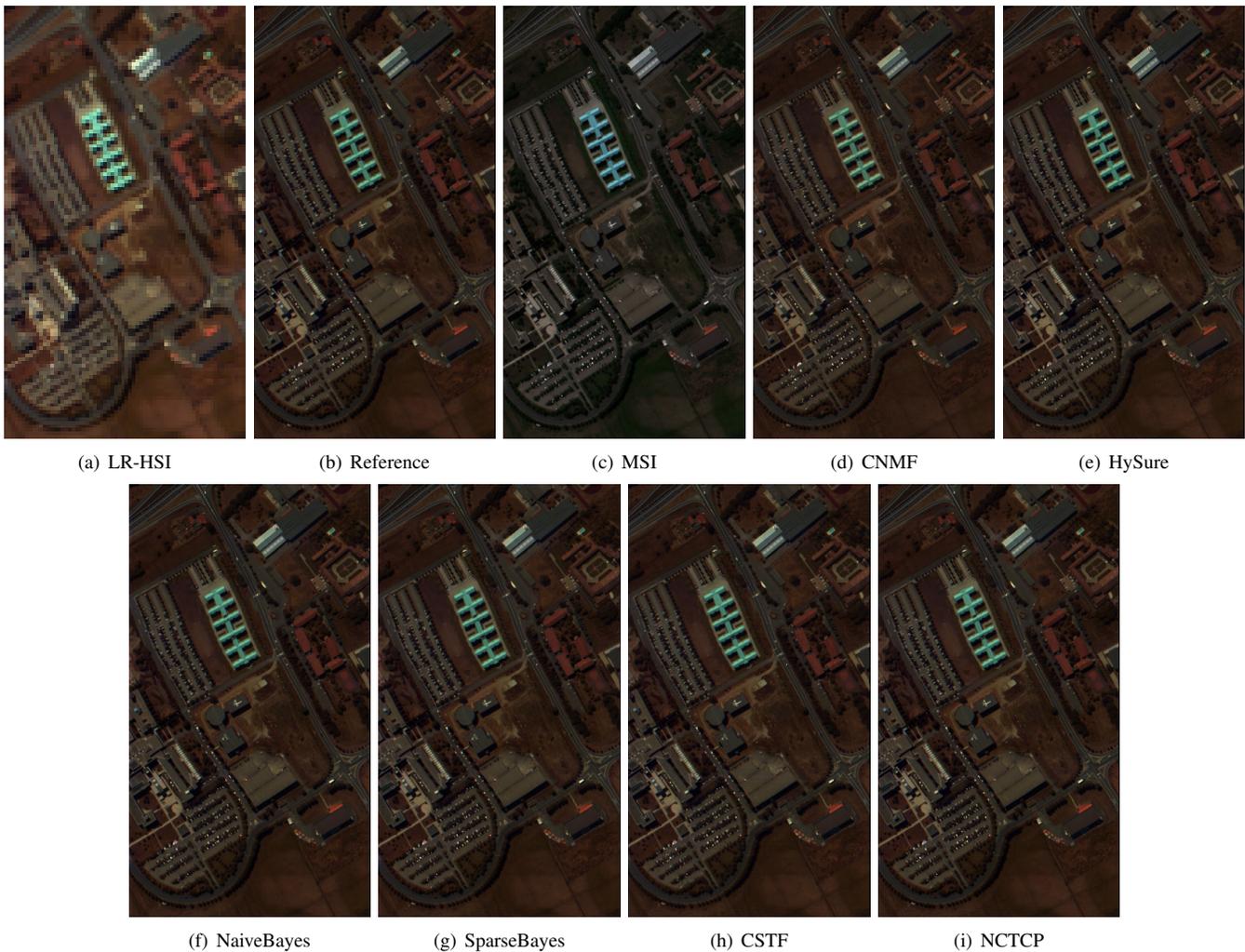


Fig. 5. HSI-MSI fusion results (University of Pavia dataset).

TABLE III  
PERFORMANCES OF THE HSI-MSI FUSION METHODS ON WASHINGTON  
DC MALL : PSNR, ERAGS, SAM, CC

Method	PSNR	ERGAS	SAM	CC
CNMF	45.77	5.09	7.38	0.91
HySure	49.59	4.38	5.87	0.93
NaiveBayes	51.23	3.72	4.61	0.95
SparseBayes	51.43	3.56	4.47	0.95
CSTF	49.74	4.65	7.91	0.92
NCTCP	<b>53.03</b>	<b>2.79</b>	<b>3.88</b>	<b>0.97</b>

TABLE IV  
COMPARISONS OF METHODS WITH AND WITHOUT HR-MSI

Dataset	Method	PSNR	ERGAS	SAM	CC
Moffet field	Without HR-MSI	29.74	5.84	8.86	0.94
	With HR-MSI	<b>41.21</b>	<b>2.36</b>	<b>4.61</b>	<b>0.99</b>
University of Pavia	Without HR-MSI	27.48	5.32	5.77	0.88
	With HR-MSI	<b>37.15</b>	<b>2.36</b>	<b>4.38</b>	<b>0.96</b>
Washington DC Mall	Without HR-MSI	42.90	6.39	6.86	0.85
	With HR-MSI	<b>53.03</b>	<b>2.79</b>	<b>3.88</b>	<b>0.97</b>

reconstruction quality.

### E. Parameter Selection

To evaluate the sensitivities of several key parameters in the proposed method, we have varied their values to observe the fluctuation in fusion results, i.e., rank  $R$ , number of patches in one cluster  $N_p$ . Fig. 9 plots the quality measures of the reconstructed Moffet field dataset as a function of the rank  $R$  and  $N_p$ .  $R$  varies from 20 to 120 with step 20 and  $N_p$  varies from 100 to 600 with step 100. It can

be observed that the quality measures are robust against the change of  $N_p$ . When  $R$  lies between 40 and 100, the proposed method achieves the best quality measures. When  $R$  is beyond 100, the proposed method exhibit performance degradation. Considering the comprehensive results, we set  $R = 80$  and  $N_p = 300$ .

### F. Real Data Experiment

In this section, we conduct the experiments on a real dataset. This dataset consists of a LR-HSI and a HR-MSI taken over

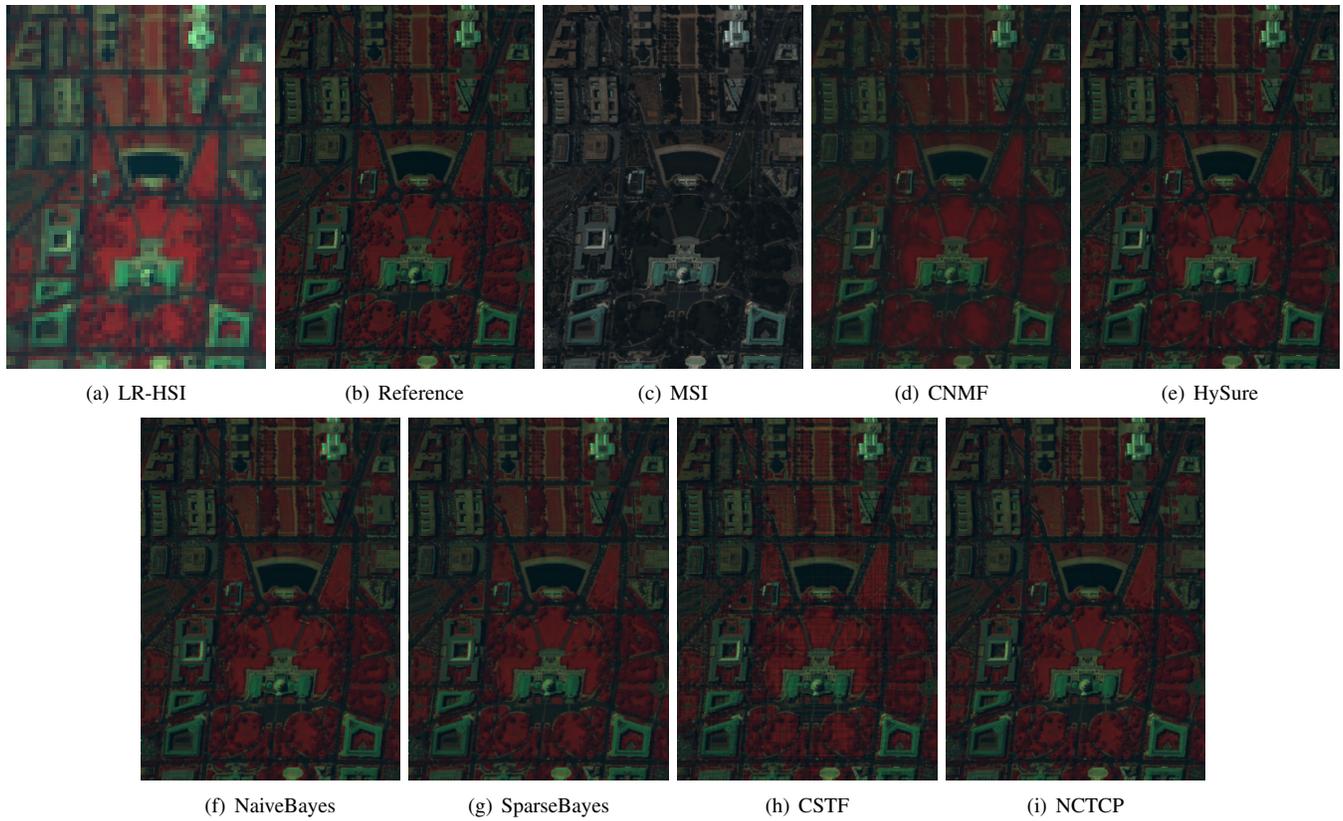


Fig. 7. HSI-MSI fusion results (Washington DC Mall).

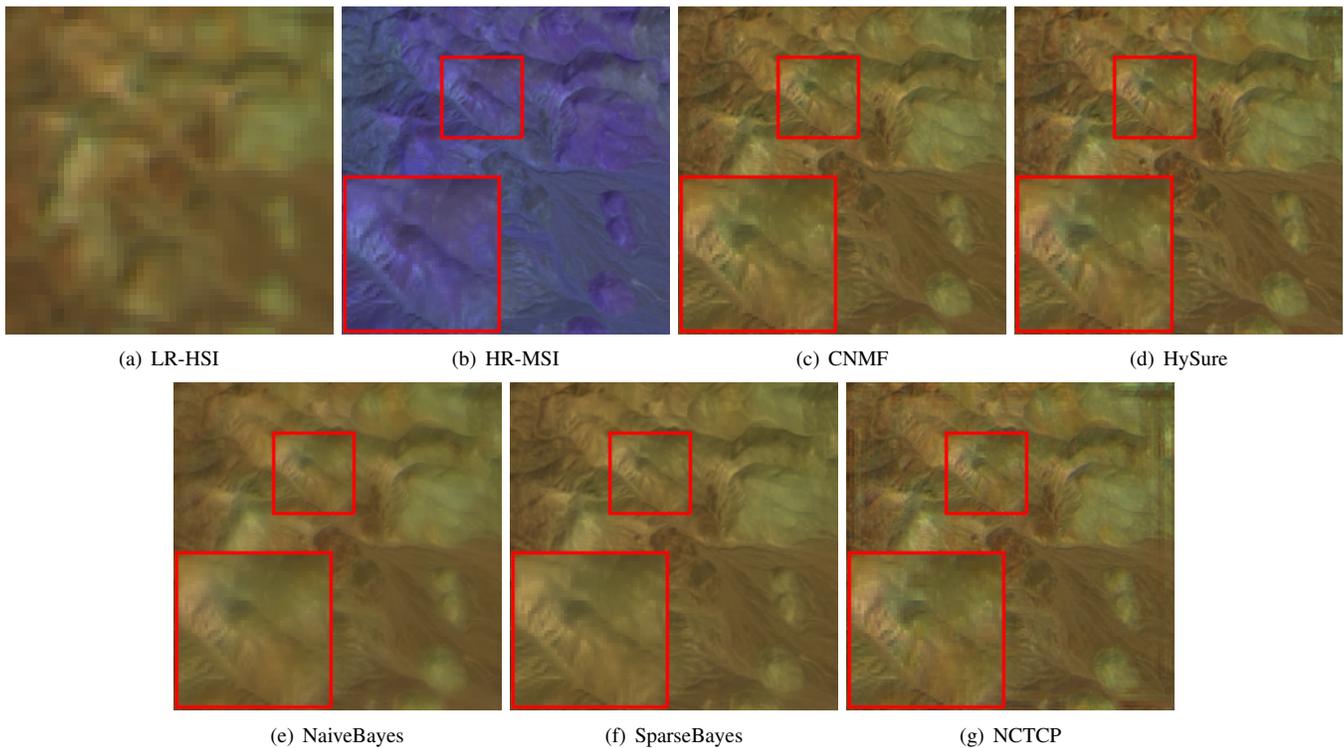


Fig. 10. False color images of real dataset and the reconstructed false color images using the compared methods.

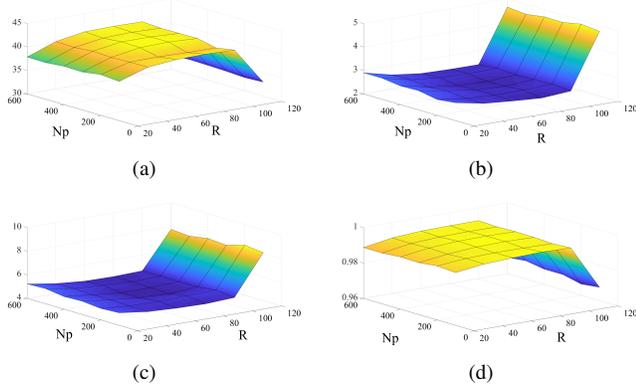


Fig. 9. Parameter analysis on Moffet field dataset. (a) PSNR, (b) ERGAS (c) SAM (d) CC.

the Cuprite district, Nevada, US. The LR-HSI has a size of  $927 \times 379 \times 167$  collected by Hyperion and the HR-MSI has a size of  $3708 \times 1516 \times 8$  collected by WorldView-3. The downsampling ratio is 4. The spectral response and the blurring kernel are not known. However, the CSTF method requires the blurring kernel to be separable to obtain the spatial degeneration operator, and therefore is not used for comparison in this real data experiment. We select a  $50 \times 50 \times 167$  sized LR-HSI and the corresponding  $200 \times 200 \times 8$  sized HR-MSI as our real test dataset. The false color images composed of 3 band LR-HSI and HR-MSI images are shown in Fig. 10(a) and Fig. 10(b), respectively. The spectral response and blurring kernel are estimated using the methods in [25]. The false color images of the reconstructed HR-HSIs are shown in Figs. 10(c)-(g). We can see that our method leads to a reconstructed HR-HSI with clear spatial structures. Fig. 11 provides the recovered spectral curves of 3 randomly selected pixels in the real scene. In addition, the spectral curves of LR-HSI in the corresponding positions are shown for comparison. We can see that our results are close to the spectral curves in LR-HSI, justifying the effectiveness of our proposed method in recovering the real spatial-spectral structure of the scene. To summarize, the proposed method also performs well in real LR-HSI and MR-HSI fusion.

## VI. CONCLUSIONS

A HSI-MSI fusion method, named NCTCP, is proposed to fuse the LR-HSI with the HR-MSI to obtain images having high spatial and spectral resolutions. Nonlocal tensors are formed by nonlocal similar patches. These nonlocal tensors are assumed to be low-rank tensors which can be represented by CP decomposition. In addition, the relationship between the HR-HSI and the HR-MSI is built by sharing the factor matrices representing the spatial and nonlocal similarity dimensions. A nonlocal coupled tensor CP decomposition model is used to reconstruct the HR-HSI. Considering that the real HR-HSI is unknown in prior, we use the HR-MSI to guide the nonlocal patch clustering in the reconstruction process. A smooth order of all the patches is computed for the clustering procedure. Numerical experiments show that the proposed NCTCP can achieve higher reconstruction accuracy and better

visual quality on the test images compared with existing HSI-MSI fusion methods. Future research can be directed towards using higher order tensor to represent the HR-HSI effectively and discover the relationship between HR-MSIs and HR-HSIs. Other tensor representation methods, e.g., Tucker and tensor train, will also be studied in future research line.

## ACKNOWLEDGMENT

The authors would like to thank Dr. Naoto Yokoya for providing us the real dataset. The copyright of the WorldView-3 imagery is European Space Imaging/Digital Globe.

## REFERENCES

- [1] Y. Yuan, X. Zheng, and X. Lu, "Discovering diverse subset for unsupervised hyperspectral band selection," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 51–64, Jan. 2017.
- [2] L. Gao, Q. Du, B. Zhang, W. Yang, and Y. Wu, "A comparative study on linear regression-based noise estimation for hyperspectral imagery," *IEEE J Sel. Topics in Appl. Earth Observ. and Remote Sensing*, vol. 6, no. 2-2, pp. 488–498, 2013.
- [3] W. Sun and Q. Du, "Graph-regularized fast and robust principal component analysis for hyperspectral band selection," *IEEE Trans. Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3185–3195, 2018.
- [4] L. Sun, Z. Wu, J. Liu, L. Xiao, and Z. Wei, "Supervised spectral-spatial hyperspectral image classification with weighted Markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 1490–1503, Mar. 2015.
- [5] B. Du, S. Wang, C. Xu, N. Wang, L. Zhang, and D. Tao, "Multi-task learning for blind source separation," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4219–4231, 2018.
- [6] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 4, pp. 1990–2000, Apr. 2016.
- [7] B. Zhang, S. Li, X. Jia, L. Gao, and M. Peng, "Adaptive markov random field approach for classification of hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 5, pp. 973–977, 2011.
- [8] C. Liu, L. He, Z. Li, and J. Li, "Feature-driven active learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 341–354, 2018.
- [9] Z. Wu, Y. Li, A. Plaza, J. Li, F. Xiao, and Z. Wei, "Parallel and distributed dimensionality reduction of hyperspectral data on cloud computing architectures," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 6, pp. 2270–2278, Jun. 2016.
- [10] J. Peng, W. Sun, and Q. Du, "Self-paced joint sparse representation for the classification of hyperspectral images," *IEEE Trans. Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1183–1194, 2019.

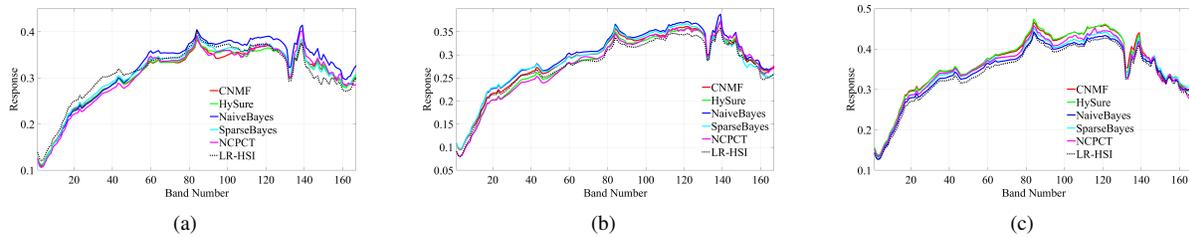


Fig. 11. Spectral curves of the reconstructed HR-HSIs using the real dataset, (a) Pixel [40,180] in compared results and Pixel [10,45] in LR-HSI, (b) Pixel [120,120] in compared results and Pixel [30,30] in LR-HSI, (c) Pixel [168,60] in compared results and Pixel [42,15] in LR-HSI.

- [11] L. Loncan, L. B. de Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes, J. Y. Tourneret, M. A. Veganzones, G. Vivone, Q. Wei, and N. Yokoya, "Hyperspectral pansharpening: A review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 27–46, Sep. 2015.
- [12] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 grs-s data-fusion contest," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007.
- [13] G. Vivone, L. Alparone, J. Chanussot, M. D. Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2565–2586, May 2015.
- [14] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, M. Selva, and C. Chen, "25 years of pansharpening: a critical review and new developments," *Signal and Image Processing for Remote Sensing*, pp. 533–548, 2011.
- [15] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1301–1312, May 2008.
- [16] W. CARPER, T. LILLESAND, and R. KIEFER, "The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data," *Photogrammetric Engineering and remote sensing*, vol. 56, no. 4, pp. 459–467, 1990.
- [17] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A new look at ihs-like image fusion methods," *Information fusion*, vol. 2, no. 3, pp. 177–186, 2001.
- [18] V. P. Shah, N. H. Younan, and R. L. King, "An efficient pan-sharpening method via a combined adaptive pca approach and contourlets," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1323–1335, 2008.
- [19] J. Liu, "Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *International Journal of Remote Sensing*, vol. 21, no. 18, pp. 3461–3472, 2000.
- [20] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "Mtf-tailored multiscale fusion of high-resolution ms and pan imagery," *Photogrammetric Engineering & Remote Sensing*, vol. 72, no. 5, pp. 591–596, 2006.
- [21] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multi-band images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1117–1127, 2015.
- [22] Q. Wei, N. Dobigeon, and J. Y. Tourneret, "Bayesian fusion of hyperspectral and multispectral images," in *Proc. Speech and Signal Processing (ICASSP) 2014 IEEE Int. Conf. Acoustics*, May 2014, pp. 3176–3180.
- [23] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3658–3668, Jul. 2015.
- [24] Q. Wei, J. M. Bioucas-Dias, N. Dobigeon, and J. Y. Tourneret, "Fusion of multispectral and hyperspectral images based on sparse representation," in *Proc. 22nd European Signal Processing Conf. (EUSIPCO)*, Sep. 2014, pp. 1577–1581.
- [25] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3373–3388, Jun. 2015.
- [26] 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 Transactions on Image Processing*, vol. 25, no. 5, pp. 2337–2352, May 2016.
- [27] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 3631–3640.
- [28] B. Huang, H. Song, H. Cui, J. Peng, and Z. Xu, "Spatial and spectral image fusion using sparse matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 3, pp. 1693–1704, 2014.
- [29] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled non-negative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 528–537, 2012.
- [30] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in

- Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3586–3594.
- [31] J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu, “Advanced deep-learning techniques for salient and category-specific object detection: A survey,” *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 84–100, 2018.
- [32] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, “When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative cnns,” *IEEE Transactions on Geoscience and Remote Sensing*, 2018.
- [33] B. Du, W. Xiong, J. Wu, L. Zhang, L. Zhang, and D. Tao, “Stacked convolutional denoising auto-encoders for feature representation,” *IEEE Transactions on Cybernetics*, vol. 47, no. 4, pp. 1017–1027, 2017.
- [34] G. Cheng and J. Han, “A survey on object detection in optical remote sensing images,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 117, pp. 11–28, 2016.
- [35] J. Han, D. Zhang, S. Wen, L. Guo, T. Liu, and X. Li, “Two-stage learning to predict human eye fixations via sdaes,” *IEEE Transactions on Cybernetics*, vol. 46, no. 2, pp. 487–498, 2016.
- [36] Y. Li, J. Hu, X. Zhao, W. Xie, and J. Li, “Hyperspectral image super-resolution using deep convolutional neural network,” *Neurocomputing*, vol. 266, pp. 29–41, 2017.
- [37] Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang, “A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 978–989, 2018.
- [38] R. Dian, L. Fang, and S. Li, “Hyperspectral image super-resolution via non-local sparse tensor factorization,” in *Proc. IEEE Conf. CVPR*, 2017.
- [39] M. A. Veganzones, J. E. Cohen, R. C. Farias, J. Chanussot, and P. Comon, “Nonnegative tensor CP decomposition of hyperspectral data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2577–2588, May 2016.
- [40] Y. Qian, F. Xiong, S. Zeng, J. Zhou, and Y. Y. Tang, “Matrix-vector nonnegative tensor factorization for blind unmixing of hyperspectral imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 3, pp. 1776–1792, Mar. 2017.
- [41] B. Du, Z. Wang, L. Zhang, L. Zhang, W. Liu, J. Shen, and D. Tao, “Exploring representativeness and informativeness for active learning,” *IEEE Transactions on Cybernetics*, vol. 47, no. 1, pp. 14–26, 2017.
- [42] Y. Chang, L. Yan, H. Fang, S. Zhong, and Z. Zhang, “Weighted low-rank tensor recovery for hyperspectral image restoration,” *arXiv preprint arXiv:1709.00192*, 2017.
- [43] Q. Xie, Q. Zhao, D. Meng, Z. Xu, S. Gu, W. Zuo, and L. Zhang, “Multispectral images denoising by intrinsic tensor sparsity regularization,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 1692–1700.
- [44] R. A. Harshman, “Foundations of the parafac procedure: models and conditions for an” explanatory” multimodal factor analysis,” 1970.
- [45] K. Huang, N. D. Sidiropoulos, and A. P. Liavas, “A flexible and efficient algorithmic framework for constrained matrix and tensor factorization,” *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 5052–5065, Oct. 2016.
- [46] A. P. Liavas and N. D. Sidiropoulos, “Parallel algorithms for constrained tensor factorization via alternating direction method of multipliers,” *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5450–5463, Oct. 2015.
- [47] J. D. Carroll and J.-J. Chang, “Analysis of individual differences in multidimensional scaling via an n-way generalization of eckart-young decomposition,” *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.
- [48] M. A. Veganzones, S. Douté, J. E. Cohen, R. C. Farias, J. Chanussot, and P. Comon, “Nonnegative cp decomposition of multiangle hyperspectral data: a case study on crism observations of martian icy surface,” in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2016 8th Workshop on*. IEEE, 2016, pp. 1–5.
- [49] L. Fang, N. He, and H. Lin, “Cp tensor-based compression of hyperspectral images,” *JOSA A*, vol. 34, no. 2, pp. 252–258, 2017.
- [50] X. Guo, X. Huang, L. Zhang, and L. Zhang, “Hyperspectral image noise reduction based on rank-1 tensor decomposition,” *ISPRS journal of photogrammetry and Remote Sensing*, vol. 83, pp. 50–63, 2013.
- [51] P. Comon, “Tensors : A brief introduction,” *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 44–53, May 2014.
- [52] M. A. Veganzones, J. E. Cohen, R. C. Farias, K. Usevich, L. Drumetz, J. Chanussot, and P. Comon, “Canonical polyadic decomposition of hyperspectral patch tensors,” in *Proc. 24th European Signal Processing Conf. (EUSIPCO)*, Aug. 2016, pp. 2176–2180.
- [53] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.
- [54] 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 Transactions on Image Processing*, vol. 25, no. 1, pp. 274–288, Jan. 2016.
- [55] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tournet, M. Chen, and S. Godsill, “Multiband image fusion based on spectral unmixing,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7236–7249, 2016.
- [56] Y. Zhang, X. Mou, G. Wang, and H. Yu, “Tensor-based dictionary learning for spectral ct reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 36, no. 1, pp. 142–154, 2017.
- [57] K. Hosono, S. Ono, and T. Miyata, “Weighted tensor nuclear norm minimization for color image denoising,” in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 3081–3085.

- [58] J. Xue and Y. Zhao, "Rank-1 tensor decomposition for hyperspectral image denoising with nonlocal low-rank regularization," in *Machine Vision and Information Technology (CMVIT), International Conference on*. IEEE, 2017, pp. 40–45.
- [59] B. Du, M. Zhang, L. Zhang, R. Hu, and D. Tao, "Pltd: Patch-based low-rank tensor decomposition for hyperspectral images," *IEEE Transactions on Multimedia*, vol. 19, no. 1, pp. 67–79, Jan. 2017.
- [60] W. Cao, Y. Wang, J. Sun, D. Meng, C. Yang, A. Cichocki, and Z. Xu, "Total variation regularized tensor rpca for background subtraction from compressive measurements," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4075–4090, Sep. 2016.
- [61] I. Ram, M. Elad, and I. Cohen, "Image processing using smooth ordering of its patches," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2764–2774, Jul. 2013.
- [62] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [63] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation  $ax + xb = c$  [f4]," *Communications of the ACM*, vol. 15, no. 9, pp. 820–826, 1972.
- [64] Q. Wei, N. Dobigeon, and J.-Y. Tournieret, "Fast fusion of multi-band images based on solving a sylvester equation," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4109–4121, 2015.
- [65] Q. Wei, N. Dobigeon, J.-Y. Tournieret, J. Bioucas-Dias, and S. Godsill, "R-fuse: Robust fast fusion of multiband images based on solving a sylvester equation," *IEEE Signal Processing Letters*, vol. 23, no. 11, pp. 1632–1636, 2016.
- [66] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. on Image Process.*, vol. 27, no. 8, pp. 4118–4130, Aug. 2018.
- [67] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 2, pp. 29–56, 2017.