

# Hyperspectral and Multispectral Image Fusion using Optimized Twin Dictionaries

Xiaolin Han, Jing Yu, *Member, IEEE*, Jing-Hao Xue, and Weidong Sun, *Member, IEEE*

**Abstract**—Spectral or spatial dictionary has been widely used in fusing low-spatial-resolution hyperspectral (LH) images and high-spatial-resolution multispectral (HM) images. However, only using spectral dictionary is insufficient for preserving spatial information, and vice versa. To address this problem, a new LH and HM image fusion method termed OTD using optimized twin dictionaries is proposed in this paper. The fusion problem of OTD is formulated analytically in the framework of sparse representation, as an optimization of twin spectral-spatial dictionaries and their corresponding sparse coefficients. More specifically, the spectral dictionary representing the generalized spectrums and its spectral sparse coefficients are optimized by utilizing the observed LH and HM images in the spectral domain; and the spatial dictionary representing the spatial information and its spatial sparse coefficients are optimized by modeling the rest of high-frequency information in the spatial domain. In addition, without non-negative constraints, the alternating direction methods of multipliers (ADMM) are employed to implement the above optimization process. Comparison results with the related state-of-the-art fusion methods on various datasets demonstrate that our proposed OTD method achieves a better fusion performance in both spatial and spectral domains.

**Index Terms**—hyperspectral image fusion, optimized twin dictionaries (OTD), spectral dictionary, spatial dictionary.

## I. INTRODUCTION

Hyperspectral images can provide abundant useful spectral information, and are widely used in many applications, such as image classification and target detection [1]-[4]. However, the imaging process to achieve high spectral resolution is at the expense of spatial resolution [5]. Compared with hyperspectral images, multispectral images usually have much higher spatial resolution, but much fewer spectral bands. Fortunately, the spatial information lost in the hyperspectral imaging can be estimated by using a multispectral image over the same scene, through image fusion. In other words, a high-spatial-resolution hyperspectral (HH) image can be obtained, by fusing a low-spatial-resolution hyperspectral (LH)

image with a high-spatial-resolution multispectral (HM) image over the same scene.

Hyperspectral and multispectral image (HS-MS) fusion methods can be roughly divided into six groups: pansharpening based methods, spatial dictionary based methods, spectral dictionary based methods, Bayesian based methods, tensor decomposition based methods, and deep learning based methods. More specifically, pansharpening based methods have been proposed to improve the spatial resolution of a multispectral image by fusing the multispectral image with a panchromatic (PAN) image of higher spatial resolution [6][7]. Using a 23 coefficients polynomial filter, the EXP method [8] has achieved a sole interpolation of the LHS image. Aiazzi *et al.* [9] have proposed a component substitution based pansharpening method, which is integrated into the Gram-Schmidt spectral sharpening method [10]. Liu *et al.* [11] have proposed a multiresolution analysis based method for pansharpening, utilizing the solar radiation and landsurface reflection model. Furthermore, a hypersharpening technique is involved by Yokoya *et al.* [12], to adapt the pansharpening to the HS-MS fusion problem. For the WorldView-3 data, an improved hypersharpening method of [13] is proposed by Selva *et al.* [14], which introduces the histogram matching operation into the synthesized band variant. Based on the sparse representation theory, Vicinanza *et al.* [15] have proposed a pansharpening method, exploiting the details self-similarity through the scales. Grohnfeldt *et al.* [18] expanded the pansharpening methods proposed in [16][17] to the HS-MS fusion problem, which simplifies the hyperspectral and multispectral image fusion as several pansharpening processes.

For the further improvement of the fusion performance, spatial dictionary based methods have been investigated. Zhao *et al.* [19] have proposed a collaborative representation based fusion method, using a local adaptive spatial dictionary pair that consists of each patch and its neighboring patches, which can also reduce the computational complexity of the method proposed in [18]. Additionally, a spectral unmixing and Bayesian sparse representation based method has been proposed by Ghasrodashti *et al.* [20], to improve the spatial resolution of LH image by using an HM image over the same scene and an additional PAN image from an unrelated scene, in which two weighted spatial dictionaries are created, one from the HM image and the other from the PAN image.

As the spatial dictionary based methods may not perform well in spectral preservation, spectral dictionary based methods have been proposed to further improve the fusion quality in the spectral domain. Relying on the linear mixing model [21], a method of learning spectral dictionary has been proposed by Huang *et al.* [22], where the pure spectral signatures are

X. Han and W. Sun are with the Institute for Artificial Intelligence, Beijing National Research Center for Information Science and Technology, State Key Laboratory of Intelligent Technology and Systems, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: hxl15@mails.tsinghua.edu.cn; wdsun@tsinghua.edu.cn).

J. Yu is with the Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: jing.yu@bjut.edu.cn).

J.-H. Xue is with the Department of Statistical Science, University College London, London, WC1E 6BT, U.K. (e-mail: jinghao.xue@ucl.ac.uk).

This work was supported in part by the National Natural Science Foundation of China (41971294), and Beijing Natural Science Foundation (4172002).

constructed by a trained spectral dictionary. Yokoya *et al.* [23] have proposed a fusion method that decomposes the hyperspectral image into abundances and endmembers, i.e., a special form of spectral dictionary. Simoes *et al.* [24] have proposed an LH and HM image fusion method using total variation regularization, in which the spectral dictionary consists of endmember spectral signatures. Akhtar *et al.* [5] have offered a sparse representation approach with non-negative constraints, where each atom in the spectral dictionary is regarded as a pure spectral signature. Similarly, a non-negative structured sparse representation method has been proposed by Dong *et al.* [25], which exploits the strong spectral correlations in similar neighbors. Han *et al.* [26] have proposed a fusion method by using sparse representation without non-negative constraints, where the spectral dictionary consists of mixed spectral signatures. In addition, to utilize the spectral and spatial information simultaneously, Nezhad *et al.* [27] have proposed a fusion method with spectral unmixing and sparse coding, in which a spectral dictionary is constructed directly by using the endmembers of the LH image, and a spatial dictionary is created by using several HM or PAN images from unrelated scenes. Yi *et al.* [28] have also proposed a fusion method, using a spectral dictionary created from the endmembers of the LH image, but the spatial dictionary is trained using the HM image only.

Using the prior distribution in the observed scene, such as Gaussian priors, Bayesian based methods have been proposed to build various estimators [29]. Wei *et al.* [30] have proposed an HS-MS fusion method with a carefully designed sparse regularization term. In addition, a fast fusion method has also been proposed by solving a Sylvester equation associated with Bayesian estimator [31].

Based on tensor decomposition, Zhang *et al.* [32] have established a low-rank tensor decomposition model for the HS-MS fusion, using two graphs for the spatial correlation and the spectral structure preservation. Li *et al.* [33] have proposed a coupled sparse tensor factorization based fusion method, utilizing a sparse prior on the core tensor.

As for the deep learning based methods, Palsson *et al.* [34] have firstly proposed a 3-D convolutional neural network (CNN) method to fuse LH and HM images into an HH image, by learning the spatial decimation filter. A pyramid fully CNN based method has been proposed by Zhou *et al.* [35], which is made up of encoder and pyramid fusion sub-networks. Han *et al.* [36] have proposed a multi-branch BP neural network based HS-MS fusion method, to establish the spectral mapping for each cluster.

As discussed above, by using the spatial dictionary, the HS-MS fusion methods proposed in [18]-[20] can exploit detailed structure information, while the spectral preservation has not been taken into consideration. On the other hand, for the spectral dictionary based methods, the spectral dictionary is usually created by using the pure spectral signatures of endmembers [22]-[25][27][28], which are somehow difficult to determine, particularly for the small and weak targets in practical applications. Although in the spectral unmixing, constraint based priors can be used to ensure the uniqueness of decomposition [21], but it is quite challenging to design and evaluate a proper prior. To preserve both spectral and spatial information simultaneously, in addition to using the spectral

dictionary, related or unrelated high-spatial-resolution images are often used to create an additional penalty term indirectly [27][28], but the high-frequency spatial information such as edges and textures may be ignored.

To deal with the above issues, this paper proposes a fusion method with optimized twin dictionaries (named as OTD), to obtain the HH image by fusing an LH image with an HM image over the same scene. Firstly, an optimization of the spectral dictionary and its sparse coefficients has been formulated and derived theoretically, by utilizing the spectral information provided by the LH and HM images. Here, using the alternating direction method of multipliers (ADMM) [37] without non-negative constraints, we can consider each atom in the spectral dictionary as a more generalized spectral signature, instead of a pure spectral signature. Then, an optimization of the spatial dictionary and its corresponding sparse coefficients is also derived theoretically, and will be estimated by the K-singular value decomposition (K-SVD) [38] and ADMM [37] methods, utilizing high-frequency information supplied by the image patches from the rest of HM and LH image. In this way, each atom in the spatial dictionary represents the high-frequency information, such as edges and textures. Finally, using the optimized twin dictionaries and their corresponding sparse coefficients, a final fused HH image can be obtained.

The main contributions of this paper are listed as follows.

- 1) A new optimized twin dictionaries scheme for the LH and HM image fusion is proposed for the first time. The scheme can use both of the spectral and spatial information directly.
- 2) An optimized spectral dictionary is derived theoretically, to utilize the spectral information provided by the LH and HM images, where each atom represents a more generalized spectral signature.
- 3) An optimized spatial dictionary representing the high-frequency information is also derived theoretically, to further utilize the remaining spatial information unrepresented by the spectral dictionary.

The rest of this paper is organized as follows. Section II presents the proposed OTD method. Experimental results and analyses on different datasets are provided in Section III, followed by the conclusions in Section IV.

## II. PROPOSED METHOD

### A. Problem Formulation

The HH image  $\mathbf{X} \in \mathbb{R}^{\lambda_X \times N}$  can be estimated by fusing an LH image  $\mathbf{Y}_H \in \mathbb{R}^{\lambda_X \times n}$  with an HM image  $\mathbf{Y}_M \in \mathbb{R}^{\lambda_Y \times N}$  over the same scene, where  $\lambda_X$  and  $\lambda_Y$  ( $\lambda_X \gg \lambda_Y$ ) denote the numbers of spectral bands, and  $n$  and  $N$  ( $N \gg n$ ) denote the numbers of pixels per band, in  $\mathbf{Y}_H$  and  $\mathbf{Y}_M$ , respectively. Moreover, the observed images  $\mathbf{Y}_H$  and  $\mathbf{Y}_M$  can be expressed as degraded versions of the desired HH image  $\mathbf{X}$  as

$$\mathbf{Y}_H = \mathbf{X}\mathbf{H} + \mathbf{N}_H, \quad (1)$$

$$\mathbf{Y}_M = \mathbf{L}\mathbf{X} + \mathbf{N}_M, \quad (2)$$

where  $\mathbf{H} \in \mathbb{R}^{N \times n}$  denotes the blurring and down sampling operator,  $\mathbf{L} \in \mathbb{R}^{\lambda_Y \times \lambda_X}$  denotes the spectral response function

of the  $\mathbf{Y}_M$ , and  $\mathbf{N}_H$  and  $\mathbf{N}_M$  denote the zero-mean Gaussian noises in the observation model, which is a popular assumption in the imaging process modelling [31][25]. As the estimation of  $\mathbf{X}$  is an ill-posed inverse problem, the sparse representation framework is adopted for its efficiency in solving Eq.(1) and Eq.(2). That is, the HH image  $\mathbf{X}$  can be represented as a linear combination of the spectral dictionary  $\mathbf{D}_s \in \mathbb{R}^{\lambda \times K}$ ,

$$\mathbf{X} = \mathbf{D}_s \mathbf{A} + \mathbf{E}, \quad (3)$$

where  $\mathbf{A} \in \mathbb{R}^{K \times N}$  denotes the spectral sparse coefficients matrix, and  $\mathbf{E} \in \mathbb{R}^{\lambda \times N}$  denotes the error matrix that cannot be represented by the spectral dictionary  $\mathbf{D}_s$ . Here, we found that the remaining spatial information, which cannot be fully represented by the spectral dictionary  $\mathbf{D}_s$ , can still be dug out from the error matrix  $\mathbf{E}$ , i.e.,  $\mathbf{E}$  can be further represented by a spatial dictionary  $\mathbf{D}_p \in \mathbb{R}^{B_p \times K_p}$  with image patches of size  $\sqrt{B_p} \times \sqrt{B_p}$  as

$$\mathbf{E} = \mathbf{D}_p \circ \alpha, \quad (4)$$

where  $\alpha \in \mathbb{R}^{K_p \times N_p}$  is the sparse coefficients matrix in the image patch domain. Symbol “ $\circ$ ” denotes the equivalent expression of multiplication in the image patch domain as shown in Fig.1, which is a popular symbol in image blocking [39][40]. Substituting Eq.(4) into Eq.(3) leads to the following equation:

$$\mathbf{X} = \mathbf{D}_s \mathbf{A} + \mathbf{D}_p \circ \alpha. \quad (5)$$

Therefore, the HH image  $\mathbf{X}$  can be expressed by the twin dictionary denoted by  $\mathbf{D}_s$  and  $\mathbf{D}_p$ . We can see that  $\mathbf{D}_s$  and  $\mathbf{D}_p$  are not two completely independent dictionaries:  $\mathbf{D}_s$  is derived from the whole image of  $\mathbf{Y}_H$  and  $\mathbf{Y}_M$ , while  $\mathbf{D}_p$  is derived by only using the remaining spatial information of  $\mathbf{E}$  that cannot be represented by the  $\mathbf{D}_s$ . This is the reason why we called them as twin dictionaries. Furthermore, each atom in the spectral dictionary  $\mathbf{D}_s$  can be considered as a more generalized spectral signature, no longer a pure spectral signature; the spatial dictionary  $\mathbf{D}_p$  can represent the high-frequency spatial information such as edges and textures, utilizing the remaining spatial information unrepresented by the spectral dictionary. Substituting Eq.(3) and Eq.(4) into Eq.(1) and Eq.(2) leads to the following equations:

$$\mathbf{Y}_H = \mathbf{D}_s \mathbf{A} \mathbf{H} + \mathbf{E}_H, \quad (6)$$

$$\mathbf{Y}_M = \mathbf{L} \mathbf{D}_s \mathbf{A} + \mathbf{E}_M, \quad (7)$$

$$\mathbf{E}_H = \mathbf{D}_p \circ \alpha \mathbf{H} + \mathbf{N}_H, \quad (8)$$

$$\mathbf{E}_M = \mathbf{L} \mathbf{D}_p \circ \alpha + \mathbf{N}_M, \quad (9)$$

where  $\mathbf{E}_H \in \mathbb{R}^{\lambda \times n}$  and  $\mathbf{E}_M \in \mathbb{R}^{\lambda \times N}$  denote the error matrices that cannot be expressed by using the spectral dictionary  $\mathbf{D}_s$  in  $\mathbf{Y}_H$  and  $\mathbf{Y}_M$ , respectively. In addition, Eq.(8) and Eq.(9) can be rewritten in the image patch domain:

$$\varepsilon_H = \mathbf{P} \mathbf{D}_p \alpha + n_H, \quad (10)$$

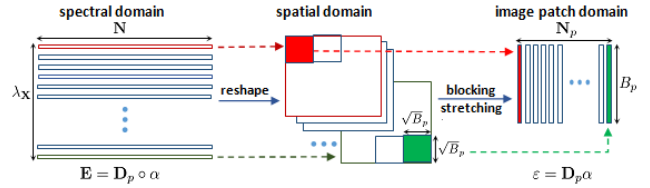


Fig.1 Equivalent expression of multiplication denoted by symbol “ $\circ$ ”.

$$\varepsilon_M = \mathbf{D}_p \alpha \mathbf{L}_p + n_M, \quad (11)$$

where  $\varepsilon_H$  and  $\varepsilon_M$  denote the stretched patches of  $\mathbf{E}_H$  and  $\mathbf{E}_M$ ,  $\mathbf{L}_p$  and  $\mathbf{P}$  denote the equivalent expressions of  $\mathbf{L}$  and  $\mathbf{H}$ , and  $n_H$  and  $n_M$  denote the equivalent expressions of  $\mathbf{N}_H$  and  $\mathbf{N}_M$  in the image patch domain, respectively.

In the following subsections, the two parts of our proposed method will be discussed in more detail. Firstly, spectral dictionary  $\mathbf{D}_s$  and spectral sparse coefficients  $\mathbf{A}$  are optimized by using the spectral information provided by the observed image  $\mathbf{Y}_H$  and  $\mathbf{Y}_M$ . Secondly, spatial dictionary  $\mathbf{D}_p$  and spatial sparse coefficients  $\alpha$  are optimized by utilizing the remaining spatial information left by the spectral dictionary, i.e.  $\mathbf{E}_H$  and  $\mathbf{E}_M$ .

### B. Spectral Dictionary Optimization

To leverage the high-spectral resolution of  $\mathbf{Y}_H$  and the high-spatial resolution of  $\mathbf{Y}_M$ , the spectral dictionary  $\mathbf{D}_s$  and coefficients  $\mathbf{A}$  can be optimized following Eq.(6) and Eq.(7). Considering the sparsity of  $\mathbf{A}$  and replacing the  $l_0$ -norm of  $\mathbf{A}$  with a natural convex approximation, i.e.  $l_1$ -norm, we can get a convex relaxation as

$$\arg \min_{\mathbf{D}_s, \mathbf{A}} \|\mathbf{Y}_M - \mathbf{L} \mathbf{D}_s \mathbf{A}\|_F^2 + \eta \|\mathbf{Y}_H - \mathbf{D}_s \mathbf{A} \mathbf{H}\|_F^2 + \lambda_1 \|\mathbf{A}\|_1, \quad (12)$$

where  $\eta$  is used to balance the trade-off between the spatial error and the spectral error,  $\lambda_1$  denotes the regularization parameter to achieve the tradeoff between the representation error and the solution sparsity. The above biconvex optimization can be efficiently solved by solving the following two subproblems iteratively, a strategy widely used in various applications [41][42] with convergence guarantee [43]:

$$\mathbf{D}_s^{k+1} = \arg \min_{\mathbf{D}_s} \mathcal{J}(\mathbf{A}^k, \mathbf{D}_s), \quad (13)$$

$$\mathbf{A}^{k+1} = \arg \min_{\mathbf{A}} \mathcal{J}(\mathbf{A}, \mathbf{D}_s^{k+1}), \quad (14)$$

where  $\mathcal{J}(\mathbf{A}, \mathbf{D}_s)$  denotes the objective function in Eq.(12), and  $k$  denotes the iteration number.

With a fixed spectral sparse coefficients matrix  $\mathbf{A}$ , the spectral dictionary  $\mathbf{D}_s$  can be optimized according to Eq.(13). More specifically, Eq.(13) can be rewritten as

$$\arg \min_{\mathbf{D}_s} \|\mathbf{Y}_M - \mathbf{L} \mathbf{D}_s \mathbf{A}\|_F^2 + \eta \|\mathbf{Y}_H - \mathbf{D}_s \mathbf{A} \mathbf{H}\|_F^2, \quad (15)$$

where  $\mathbf{A}_H = \mathbf{A} \mathbf{H} \in \mathbb{R}^{K \times n}$ . Note that it is much more challenging to minimize the sum of the two terms in Eq.(15)

TIP-20724-2019.R2

than minimizing each term separately. Here, the splitting strategy is exploited, and the alternating direction method of multipliers (ADMM) [37] technique is used to solve this convex problem, which has been widely used in various applications [25][26][42]. Firstly, we reformulate Eq.(15) as

$$\begin{aligned} \arg \min_{\mathbf{D}_s} & \|\mathbf{Y}_M - \mathbf{LZ}\|_F^2 + \eta \|\mathbf{Y}_H - \mathbf{D}_s \mathbf{A}_H\|_F^2 \\ \text{s.t. } & \mathbf{Z} = \mathbf{D}_s \mathbf{A} \end{aligned} \quad (16)$$

where the splitting variable is  $\mathbf{Z} = \mathbf{D}_s \mathbf{A}$ . Then, the augmented Lagrangian function of Eq.(16) becomes

$$\begin{aligned} \mathcal{L}(\mathbf{D}_s, \mathbf{Z}, \mathbf{V}_1) = & \|\mathbf{Y}_M - \mathbf{LZ}\|_F^2 + \eta \|\mathbf{Y}_H - \mathbf{D}_s \mathbf{A}_H\|_F^2 \\ & + \mu_1 \|\mathbf{D}_s \mathbf{A} - \mathbf{Z} + \frac{\mathbf{V}_1}{2\mu_1}\|_F^2, \end{aligned} \quad (17)$$

where  $\mathbf{V}_1 \in \mathbb{R}^{\lambda \times \times N}$  denotes the Lagrangian multiplier ( $\mu_1 > 0$ ). Similarly to the optimization process of Eq.(12), the above augmented Lagrangian function can be minimized by solving the following two subproblems iteratively:

$$\begin{aligned} \arg \min_{\mathbf{Z}} & \mathcal{L}(\mathbf{D}_s^{(k)}, \mathbf{Z}, \mathbf{V}_1^{(k)}) \\ \arg \min_{\mathbf{D}_s} & \mathcal{L}(\mathbf{D}_s, \mathbf{Z}^{(k+1)}, \mathbf{V}_1^{(k)}) \end{aligned} \quad (18)$$

which can be rewritten as

$$\begin{aligned} \arg \min_{\mathbf{Z}} & \|\mathbf{Y}_M - \mathbf{LZ}\|_F^2 + \mu_1 \|\mathbf{D}_s^{(k)} \mathbf{A} - \mathbf{Z} + \frac{\mathbf{V}_1^{(k)}}{2\mu_1}\|_F^2 \\ \arg \min_{\mathbf{D}_s} & \eta \|\mathbf{Y}_H - \mathbf{D}_s \mathbf{A}_H\|_F^2 + \mu_1 \|\mathbf{D}_s \mathbf{A} - \mathbf{Z}^{(k+1)} + \frac{\mathbf{V}_1^{(k)}}{2\mu_1}\|_F^2, \end{aligned} \quad (19)$$

where the Lagrangian multiplier is updated by

$$\mathbf{V}_1^{(k+1)} = \mathbf{V}_1^{(k)} + \mu_1 (\mathbf{D}_s \mathbf{A}^{(k+1)} - \mathbf{Z}^{(k+1)}). \quad (20)$$

Let  $\partial \mathcal{L} / \partial \mathbf{Z}$  and  $\partial \mathcal{L} / \partial \mathbf{D}_s$  equal to zero respectively, the optimization problem in Eq.(19) can be written as

$$\begin{aligned} 2\mathbf{L}^T \mathbf{LZ} - 2\mathbf{L}^T \mathbf{Y}_M + \mu_1 (2\mathbf{Z} - 2(\mathbf{D}_s \mathbf{A} + \frac{\mathbf{V}_1}{2\mu_1})) &= 0 \\ \eta(\mathbf{D}_s \mathbf{A}_H \mathbf{A}_H^T - \mathbf{Y}_H \mathbf{A}_H^T) + \mu_1 (\mathbf{D}_s \mathbf{A} \mathbf{A}^T + (\frac{\mathbf{V}_1}{2\mu_1} - \mathbf{Z}) \mathbf{A}^T) &= 0 \end{aligned} \quad (21)$$

Then, the spectral dictionary  $\mathbf{D}_s$  can be optimized analytically as

$$\begin{aligned} \mathbf{Z} = & (\mathbf{L}^T \mathbf{L} + \mu_1 \mathbf{I})^{-1} [\mathbf{L}^T \mathbf{Y}_M + \mu_1 (\mathbf{D}_s \mathbf{A} + \frac{\mathbf{V}_1}{2\mu_1})] \\ \mathbf{D}_s = & [\eta \mathbf{Y}_H \mathbf{A}_H^T + \mu_1 (\mathbf{Z} - \frac{\mathbf{V}_1}{2\mu_1}) \mathbf{A}^T] (\eta \mathbf{A}_H \mathbf{A}_H^T + \mu_1 \mathbf{A} \mathbf{A}^T)^{-1} \end{aligned} \quad (22)$$

### C. Spectral Sparse Coefficients Optimization

With the optimized spectral dictionary  $\mathbf{D}_s$ , Eq.(14) can be rewritten as

$$\arg \min_{\mathbf{A}} \|\mathbf{Y}_M - \mathbf{D}_L \mathbf{A}\|_F^2 + \eta \|\mathbf{Y}_H - \mathbf{D}_s \mathbf{A}_H\|_F^2 + \lambda_1 \|\mathbf{A}\|_1 \quad (23)$$

where  $\mathbf{D}_L = \mathbf{L} \mathbf{D}_s$  denotes the degradation in the spectral domain, i.e., the low spectral resolution dictionary. Similarly to the optimization of Eq.(15), the splitting strategy is also used to minimize each term in Eq.(23) separately, instead of the sum of the three terms. To apply ADMM [37], we reformulate Eq.(23) as

$$\begin{aligned} \arg \min_{\mathbf{A}} & \|\mathbf{Y}_M - \mathbf{D}_L \mathbf{S}\|_F^2 + \eta \|\mathbf{Y}_H - \mathbf{B}_H\|_F^2 + \lambda_1 \|\mathbf{A}\|_1 \\ \text{s.t. } & \mathbf{S} = \mathbf{A}, \mathbf{B} = \mathbf{D}_s \mathbf{S} \end{aligned} \quad (24)$$

where  $\mathbf{S} = \mathbf{A}$  and  $\mathbf{B} = \mathbf{D}_s \mathbf{S}$  are the splitting variables. Then, by using Lagrangian multipliers  $\mathbf{V}_2 \in \mathbb{R}^{\lambda \times \times N}$  and  $\mathbf{V}_3 \in \mathbb{R}^{K \times N}$  ( $\mu_2 > 0$ ), the coefficients matrix  $\mathbf{A}$  can be estimated by ADMM analytically:

$$\begin{aligned} \mathbf{A} = & \text{soft}(\mathbf{S}^{(k)} + \frac{\mathbf{V}_3^{(k)}}{2\mu_2}, \frac{\lambda_1}{2\mu_2}) \\ \mathbf{B} = & [\eta \mathbf{Y}_H \mathbf{H}^T + \mu_2 (\mathbf{D}_s \mathbf{S}^{(k)} + \frac{\mathbf{V}_2^{(k)}}{2\mu_2})] (\mu_2 \mathbf{I} + \eta \mathbf{H} \mathbf{H}^T)^{-1} \\ \mathbf{S} = & (\mathbf{D}_L^T \mathbf{D}_L + \mu_2 \mathbf{I} + \mu_2 \mathbf{D}_s^T \mathbf{D}_s)^{-1} [\mathbf{D}_L^T \mathbf{Y}_M + \mu_2 (\mathbf{A} - \frac{\mathbf{V}_3^{(k)}}{2\mu_2}) + \mu_2 \mathbf{D}_s^T (\mathbf{B} - \frac{\mathbf{V}_2^{(k)}}{2\mu_2})] \end{aligned} \quad (25)$$

### D. Spatial dictionary Optimization

When the spectral dictionary  $\mathbf{D}_s$  and the sparse codes  $\mathbf{A}$  are obtained, the spatial dictionary  $\mathbf{D}_p$  can be estimated by utilizing the remaining spatial information provided by  $\mathbf{E}_H$  and  $\mathbf{E}_M$ . Since the equivalent matrixes  $\mathbf{L}_p$  and  $\mathbf{P}$  in the image patch domain are challenging to obtain, we proposed a spatial dictionary optimization method without using  $\mathbf{L}_p$  and  $\mathbf{P}$  as a prior. Then, Eq.(10) can be rewritten as follows with spectral degradation:

$$\varepsilon_H \mathbf{L}_p = \mathbf{P} \mathbf{D}_p \alpha \mathbf{L}_p + n_H \mathbf{L}_p. \quad (26)$$

As shown in Eq.(11) and Eq.(26), the spatial dictionary  $\mathbf{D}_p$  and the degraded spatial dictionary  $\mathbf{P} \mathbf{D}_p$  can sparsely represent the  $\varepsilon_M$  and  $\varepsilon_H \mathbf{L}_p$ , respectively. Hence the spatial dictionary  $\mathbf{D}_p$  can be estimated by minimizing the following equation with  $\alpha_p = \alpha \mathbf{L}_p$ :

$$\arg \min_{\mathbf{D}_p} \|\varepsilon_M - \mathbf{D}_p \alpha_p\|_F^2 + \beta \|\varepsilon_H \mathbf{L}_p - \mathbf{P} \mathbf{D}_p \alpha_p\|_F^2 + \lambda \|\alpha_p\|_0 \quad (27)$$

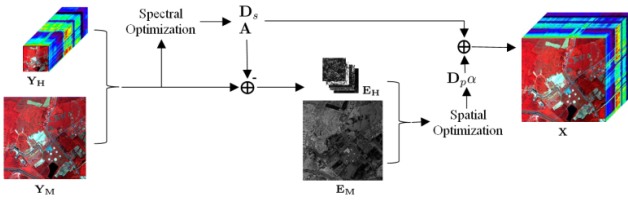


Fig.2 Overall framework of our proposed OTD method.

**Algorithm 1** Proposed OTD Algorithm

Input:  $\mathbf{Y}_R, \mathbf{Y}_I, \mathbf{H}, \mathbf{L}, \eta, \beta, \lambda_1, \lambda_2, \mu_1, \mu_2, \mu_3, T_s, T_p$ ;  
 For  $k = 0, 1, \dots, T_s - 1$   
   Compute  $\mathbf{D}^{(k+1)}$  via Eq.(22);  
   Update the Lagrangian multipliers  $\mathbf{V}_1^{(k+1)}$  via Eq.(20);  
   Compute  $\mathbf{A}^{(k+1)}$  via Eq.(25);  
 End  
 $\mathbf{D} \leftarrow \mathbf{D}^{(k+1)}$ ;  
 $\mathbf{A} \leftarrow \mathbf{A}^{(k+1)}$ ;  
 $\mathbf{E}_H = \mathbf{Y}_H - \mathbf{D}_s \mathbf{A} \mathbf{H}$ ;  
 $\mathbf{E}_M = \mathbf{Y}_M - \mathbf{L} \mathbf{D}_s \mathbf{A}$ ;  
 Compute  $\mathbf{D}_p$  via Eq.(28);  
 For  $k = 0, 1, \dots, T_p - 1$   
   Compute  $\alpha^{(k+1)}$  via Eq.(32);  
 End  
 $\alpha \leftarrow \alpha^{(k+1)}$ ;  
 $\hat{\mathbf{X}} \leftarrow \mathbf{D}_s \mathbf{A} + \mathbf{D}_p \alpha$ ;  
 Output:  $\hat{\mathbf{X}}$  as the fused  $\mathbf{X}$ .

which can be further rewritten as

$$\arg \min_{\mathbf{D}_p} \|\varepsilon_p - \tilde{\mathbf{D}}_p \alpha_p\|_F^2 + \lambda \|\alpha_p\|_0, \quad (28)$$

where  $\tilde{\mathbf{D}}_p = \begin{bmatrix} \beta \mathbf{P} \mathbf{D}_p \\ \mathbf{D}_p \end{bmatrix}$ ,  $\varepsilon_p = \begin{bmatrix} \beta \varepsilon_H \mathbf{L}_p \\ \varepsilon_M \end{bmatrix}$ ,  $\beta$  ( $\beta > 0$ ) denotes the weighting factor. The above nonconvex optimization with an  $l_0$  constraint can be efficiently solved by the K-SVD algorithm [38] with the greedy algorithm orthogonal matching pursuit [44], which has been widely used in various applications [22][45]. In this case, the spatial dictionary  $\mathbf{D}_p = [\mathbf{0} \ \mathbf{I}] \tilde{\mathbf{D}}_p$ , and the degraded version  $\mathbf{P} \mathbf{D}_p = 1/\beta [\mathbf{I} \ \mathbf{0}] \tilde{\mathbf{D}}_p$  can be estimated simultaneously. In addition, the degraded sparse coefficients  $\alpha_p = \alpha \mathbf{L}_p$  can be also obtained in the process of the K-SVD algorithm [38].

**E. Spatial Sparse Coefficient Optimization**

With the optimized spatial dictionary  $\mathbf{D}_p$ , the spatial sparse coefficients  $\alpha$  can be optimized by using Eq.(10) and Eq.(11). Considering the sparsity of  $\alpha$  and replacing the  $l_0$ -norm with the  $l_1$ -norm, we can attain a convex relaxation:

$$\arg \min_{\alpha} \|\varepsilon_H - \mathbf{P} \mathbf{D}_p \alpha\|_F^2 + \beta \|\varepsilon_M - \mathbf{D}_p \alpha \mathbf{L}_p\|_F^2 + \lambda_2 \|\alpha\|_1. \quad (29)$$

With the estimated degraded spatial dictionary  $\mathbf{P} \mathbf{D}_p$  and the estimated degraded sparse coefficients  $\alpha \mathbf{L}_p$  in the last

subsection, the above optimization problem can be simplified as

$$\arg \min_{\alpha} \|\varepsilon_H - \mathbf{P} \mathbf{D}_p \alpha\|_F^2 + \lambda_2 \|\alpha\|_1. \quad (30)$$

Similarly to the optimization of Eq.(15) and Eq.(23), the splitting strategy is used to minimize each term separately in Eq.(30). To apply ADMM [37], we reformulate Eq.(30) as

$$\arg \min_{\alpha} \|\varepsilon_H - \mathbf{P} \mathbf{D}_p \alpha\|_F^2 + \lambda_2 \|\gamma\|_1 \quad (31)$$

*s.t.*  $\alpha = \gamma$

where  $\alpha = \gamma$  is the splitting variable. Then, with ADMM, the spatial coefficients matrix  $\alpha$  can be optimized analytically as

$$\gamma = \text{soft}\left(\alpha^{(k)} + \frac{\mathbf{V}_4^{(k)}}{2\mu_3}, \frac{\lambda_2}{2\mu_3}\right)$$

$$\alpha = [(\mathbf{P} \mathbf{D}_p)^T \mathbf{P} \mathbf{D}_p + \mu_3 \mathbf{I}]^{-1} [(\mathbf{P} \mathbf{D}_p)^T \varepsilon_H + \mu_3 \left(\gamma - \frac{\mathbf{V}_4^{(k)}}{2\mu_3}\right)] \quad (32)$$

where  $\mathbf{V}_4 \in \mathbb{R}^{K_p \times N_p}$  denotes the Lagrangian multiplier ( $\mu_3 > 0$ ).

It can be seen that, with the optimized twin dictionaries, i.e.  $\mathbf{D}_s$  and  $\mathbf{D}_p$ , and the corresponding coefficients matrices, i.e.  $\mathbf{A}$  and  $\alpha$ , both spectral and high-frequency spatial information can be integrated into the final HH images analytically by using our proposed method. The overall framework of our proposed OTD method is illustrated in Fig.2, and the overall algorithm for our proposed OTD method is summarized in **Algorithm 1**.

**III. EXPERIMENTAL RESULTS**

To evaluate the performance of our proposed OTD method, extensive experiments are conducted in comparison with the EXP method [8], the improved hypersharpening method using synthesized band (IHSB) [14], the smoothing filtered-based intensity modulation hypersharpening (SFIMHS) method [12], Gram-Schmidt adaptive (GSA) method [9], the collaborative representation using local adaptive dictionary pair (LACRF) method [19], the generalization of simultaneous orthogonal matching pursuit (G-SOMP+) method [5], the hyperspectral superresolution (Hysure) method [24], the fast fusion based on Sylvester equation (FUSE) method [31], the non-negative structured sparse representation (NSSR) method [25], and the non-factorization sparse representation and error matrix estimation (NFSREE) method [26] on different image datasets. The EXP and the IHSB methods are used as the baseline of the pansharpening based methods in most of the visual assessments. Furthermore, to evaluate the quality of the fused HH images obtained by these methods, six full-reference quality metrics are used, including the mean square error (MSE) [46], the peak-signal-to-noise ratio (PSNR) [46], the universal image quality index (UIQI) [31], the spectral angle mapper (SAM) [31], the relative dimensionless global error in synthesis (ERGAS) [31] and the averaged structural similarity (ASSIM)



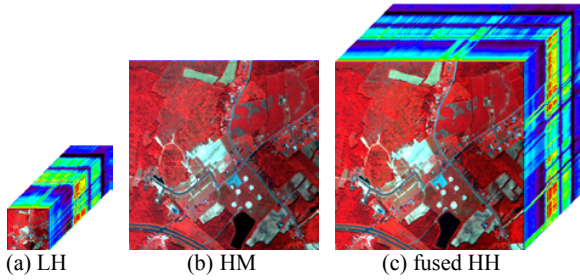


Fig. 3 False color images of (a) the LH image, (b) the HM image, (c) the fused HH image by using our proposed OTD method.

Table 1 Averaged MSE, PSNR, SAM, UIQI, ERGAS and SSIM results on the AVIRIS dataset.

	MSE	PSNR	UIQI	SAM	ERGAS	ASSIM	TIME/s
EXP	14.6650	36.4680	0.8533	3.7260	2.6524	0.8869	<b>0.22</b>
IHSB	22.2070	34.6659	0.9396	1.8821	3.1422	0.9728	2.69
SFIMHS	11.0640	37.6917	0.9071	3.1555	2.4972	0.9553	0.23
GSA	3.5330	42.6494	0.9682	3.0960	1.3792	0.9802	0.79
LACRF	0.9574	48.3199	0.9858	1.8461	0.7789	0.9922	2.87
G-SOMP+	0.8151	49.0188	0.9881	1.7622	0.7758	0.9961	75.33
Hysure	0.2863	53.5630	0.9924	0.9454	0.4604	0.9979	69.05
FUSE	0.3183	53.1025	0.9923	0.9980	0.4704	0.9981	1.89
NSSR	0.6958	49.7059	0.9949	1.5130	0.6320	0.9967	66.13
NFSREE	0.3087	53.2347	0.9921	1.0133	0.4613	0.9976	7.71
OTD	<b>0.1406</b>	<b>56.6519</b>	<b>0.9958</b>	<b>0.7477</b>	<b>0.3149</b>	<b>0.9985</b>	41.77

Table 2 Comparisons with that only using spectral dictionary.

	MSE	PSNR	UIQI	SAM	ERGAS	SSIM	TIME/s
Only $D_s$	0.5452	50.7650	0.9900	1.4503	0.5828	0.9978	<b>7.89</b>
OTD	<b>0.1406</b>	<b>56.6519</b>	<b>0.9958</b>	<b>0.7477</b>	<b>0.3149</b>	<b>0.9985</b>	41.77

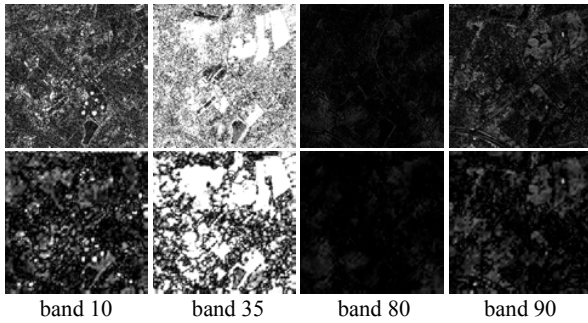


Fig.4 The ground truth (upper row) and the estimation results of  $E$  using spatial dictionary  $D_p$  (lower row) in typical bands.

defined as  $ASSIM = \frac{1}{\lambda_x} \sum_{i=1}^{\lambda_x} SSIM_i$ , where  $SSIM_i$  is the SSIM [46] of band  $i$ . All of the above methods are conducted by using MATLAB R2015b on a computer with a 3.60 GHz CPU and 16 GB RAM.

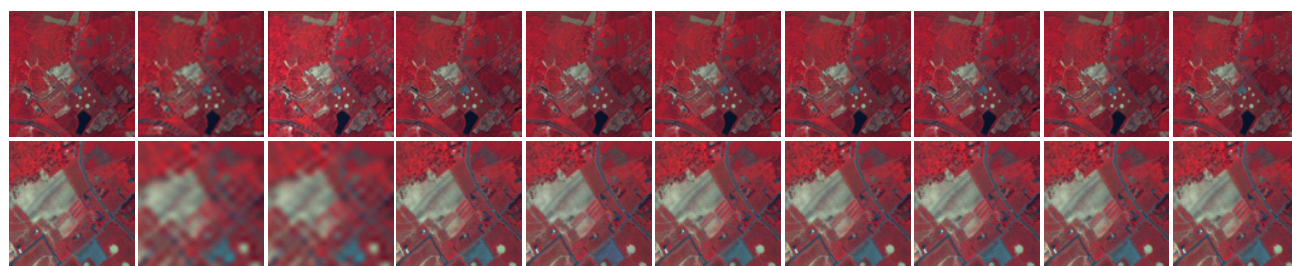
#### A. Comparison Results on the AVIRIS Dataset

In this part of experiment, the hyperspectral image from the AVIRIS airborne system [47] on July 5, 1996 is used as the ground-truth of HH image with a dimension of  $300 \times 300 \times 93$ , and some of the original bands 1-2, 105-115, 150-170, 223-224 have been removed due to the serious effect of water absorption. The IKONOS-like spectral response function [31] is used in this experiment, which covers visible and near infrared

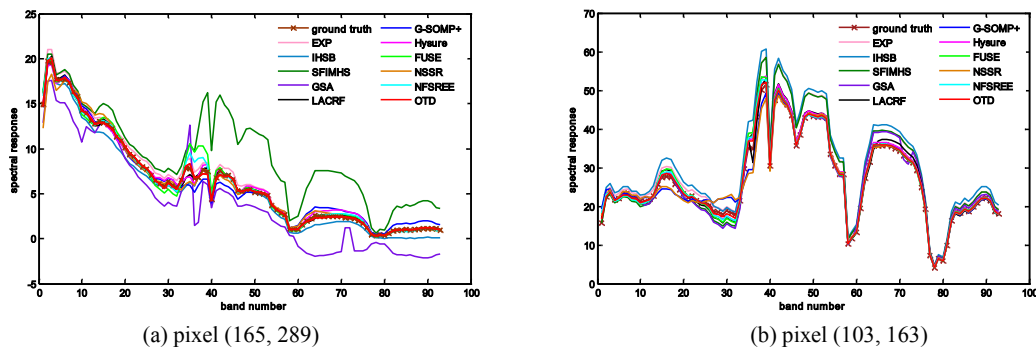
spectrum ( $\lambda_Y = 4$ ). Thus, HM image generated using Eq.(2) with a dimension of  $300 \times 300 \times 4$ , is shown as a false color image in Fig.3(b), in which the first three bands of the generated HM image are corresponding to the blue, green and red channels, respectively. The operator of blurring and down sampling matrix  $H$  is generated with a  $5 \times 5$  Gaussian kernel with standard deviation  $\sigma = 3.0$  and the downsampling rate is 5 in both vertical and horizontal directions. Consequently, LH image generated using Eq.(1) with a dimension of  $50 \times 50 \times 93$ , is shown as a false color image in Fig.3(a), where bands 49, 24 and 11 of the LH image are used as the red, green and blue channels, respectively. In addition, the non-overlapped image patches with a dimension of  $10 \times 10$  are used in the spatial domain; the number of columns in spectral and spatial dictionary  $D_s$  and  $D_p$  are set to 100 and 1000, respectively; the maximum numbers of iterations  $T_s$  and  $T_p$  are both set to 10; the parameters in the augmented Lagrangian functions are  $\lambda_1 = \lambda_2 = 10^{-6}$ ,  $\eta = \beta = 10^{-1}$ ,  $\mu_1 = 10^{-3}$ ,  $\mu_2 = 10^{-3}$ ,  $\mu_3 = 10^{-1}$ . For fair comparisons, all of the other compared methods use the same blurring and down sampling process as described above. In addition, the number of columns in spectral dictionary is set to 100 in the compared G-SOMP+, NSSR, and NSFREE methods, which is the same to that of our proposed method. The other typical parameters of the compared methods are set according to their original articles and codes for better fusion performances, such as the numbers of iterations in the NSSR method [25] is set to 26, the parameters in the augmented Lagrangian functions in the NFSREE method [26] are set to  $\lambda = 10^{-6}$ ,  $\mu = 10^{-3}$ ,  $\eta = 10$ , respectively.

The averaged MSE, PSNR, SAM, UIQI, ERGAS and ASSIM results are shown in Table 1. In addition, to exclude the corrupted pixels at the border, 2 columns of pixels at each border are excluded when calculating the full-reference quality metrics. As can be seen in Table 1, our proposed OTD method shows better fusion results in both spatial and spectral domains than all the other fusion methods. Specifically, SAM in Table 1 is reduced by more than 0.19 by our proposed OTD method, which is much significant for hyperspectral images. Simultaneously, PSNR is also improved by more than 3.1dB by the OTD method, which indicates a better performance in spatial preservation. The computing time in the last column of Table 1 shows that pansharpening based methods in the first four rows, especially the EXP method, have better performance in terms of computing time, while our proposed OTD method is only faster than the G-SOMP+, Hysure and NSSR methods, due to its double complexity of the twin dictionaries optimization. To further evaluate the efficiency of the spatial dictionary in our proposed OTD method, comparisons with the fusion method only using spectral dictionary  $D_s$  are shown in Table 2. As can be seen in Table 2, the spatial dictionary improves the performance of the fused HH image over 0.70 in SAM, and over 5.8dB in PSNR.

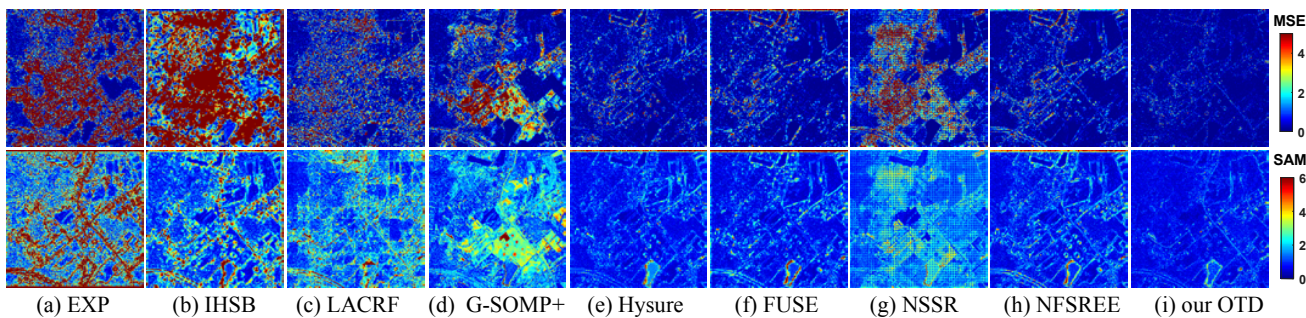
Moreover, Fig.4 shows the ground truth and the estimation of matrix  $E$  using the spatial dictionary  $D_p$  in typical bands. The high similarity between the upper and lower row also indicates the efficiency of the spatial dictionary in our proposed method. Furthermore, Fig.5 shows the false-color visual results of different methods on the fused HH image with one region zoomed. Specifically, bands 49, 24 and 11 in the fused HH



(a) Ground-truth (b) EXP (c) IHSB (d) LACRF (e) G-SOMP+ (f) Hysure (g) FUSE (h) NSSR (i) NFSREE (j) our OTD  
Fig. 5 False-color fusion results on the AVIRIS dataset. Upper row: whole band; lower row: close up.



(a) pixel (165, 289) (b) pixel (103, 163)  
Fig. 6 Fusion results in the spectral domain on (a) pixel (165, 289) and (b) pixel (103, 163) in the AVIRIS dataset.



(a) EXP (b) IHSB (c) LACRF (d) G-SOMP+ (e) Hysure (f) FUSE (g) NSSR (h) NFSREE (i) our OTD  
Fig. 7 Error images in MSE on band 30 (upper row) and SAM for the whole image (lower row) of the AVIRIS dataset.

image are used as the red, green and blue channels, respectively. It can be seen from Fig.5 that our proposed method performs better both in the spatial detail reconstruction and spectral preservation, which is obviously better than the EXP, IHSB and G-SOMP+ methods. The fusion results in the spectral domain on typical pixels are shown in Fig.6. As can be seen in Fig.6, our proposed method shows an outperformance in spectral preservation, which is apparent in comparison with the IHSB, SFIMHS, GSA and FUSE methods.

To easily discern the differences between different methods visually, Fig.7 shows the errors of the fused HH images in MSE and SAM metrics, respectively. In Fig.7, MSE images visualized the magnitude of the error at each pixel in band 30, and SAM images visualized the spatial distribution of spectral angle errors for the whole image. Compared with the related fusion methods only using spatial or spectral dictionary, our proposed OTD method shows an obvious improvement in the fusion performance, particularly in spatial preservation shown in the upper row of column (h) in Fig.7, which indicates the significance of using the spatial dictionary. Our proposed OTD method also shows a better performance in spectral reconstruction, with the lowest SAM than the other methods.

## B. Comparison Results on Two Other Dataset

In this part of experiment, the hyperspectral image from APEX [48] developed by a Swiss-Belgian consortium on behalf of ESA is used with a dimension of  $512 \times 614 \times 224$ . Another hyperspectral image dataset, Pavia Center, acquired with ROSIS (a sensor of DLR) [49], is also used with a dimension of  $1096 \times 1096 \times 102$ . Due to the serious effect of water absorption and computational complexity, just  $300 \times 300 \times 93$  of the APEX and Pavia Center datasets are used as the ground truth of HH images. In addition, the spectral response function  $\mathbf{L}$ , the blurring and down-sampling operator  $\mathbf{H}$  and all of the parameters are the same with Section III-A.

Fig.8 and Fig.9 show the visible fusion results on the APEX and Pavia Center datasets, respectively. The upper row shows the MSE of each pixel in band 30, indicating the distribution of errors in the spatial domain. As can be seen in Fig.8 and Fig.9, our proposed method performs the best in the spatial information preservation with less error in MSE. The lower rows of Fig.8 and Fig.9 show the SAM of each spectrum, which indicates the distribution of spectral distortion. As shown in Fig.8 and Fig.9, the HH image fused by our proposed OTD



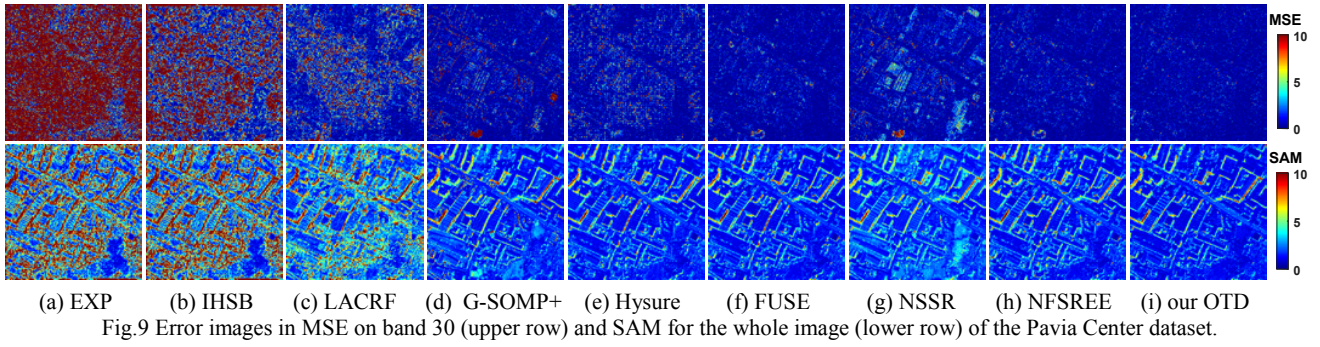
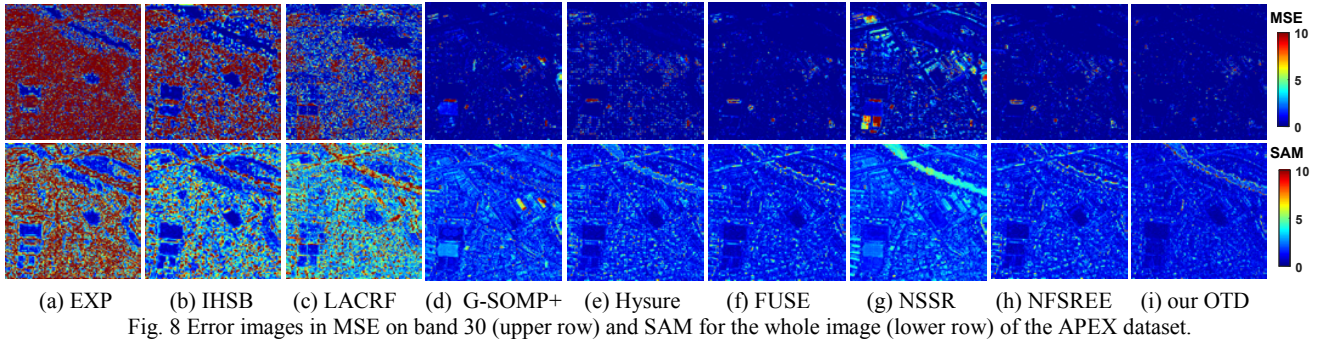


Table 3 Averaged MSE, PSNR, SAM, UIQI, ERGAS and SSIM results on different datasets.

	MSE	PSNR	UIQI	SAM	ERGAS	ASSIM	TIME/s
APEX Dataset							
EXP	261.0444	23.9637	0.8052	8.3134	7.0181	0.6040	0.51
IHSB	46.7318	31.4347	0.9690	4.2644	2.9379	0.9559	2.49
SFIMHS	239.1973	24.3432	0.8550	7.0293	7.0799	0.8350	<b>0.27</b>
GSA	80.3115	29.0830	0.9565	10.5150	4.0725	0.8864	0.75
LACRF	14.9218	36.3926	0.9824	4.6390	2.9087	0.9620	2.45
G-SOMP+	4.2817	41.8146	0.9958	1.7790	1.8289	0.9941	75.32
Hysure	2.5733	44.0258	0.9975	1.5412	1.3261	0.9946	68.57
FUSE	1.6307	46.0069	0.9982	1.3551	1.1759	0.9960	1.58
NSSR	3.4827	42.7117	0.9965	1.8276	1.5976	0.9943	64.53
NFSREE	1.6884	45.8561	0.9981	1.3872	1.1844	0.9955	7.51
OTD	<b>0.9914</b>	<b>48.1682</b>	<b>0.9989</b>	<b>1.1372</b>	<b>1.0070</b>	<b>0.9970</b>	42.81
Pavia Center Dataset							
EXP	154.9523	26.2288	0.8278	5.6700	5.5012	0.6531	0.47
IHSB	25.4769	34.0693	0.9743	3.6085	2.2388	0.9642	2.80
SFIMHS	167.3076	25.8956	0.8547	5.0707	5.6463	0.8096	<b>0.34</b>
GSA	57.6165	30.5253	0.9577	11.2709	3.2626	0.9085	0.66
LACRF	7.5854	39.3310	0.9918	3.7360	1.7774	0.9812	2.55
G-SOMP+	3.9860	42.1254	0.9958	2.1983	1.3676	0.9919	66.56
Hysure	2.2544	44.6005	0.9977	1.9453	1.0087	0.9935	65.18
FUSE	1.5644	46.1872	0.9983	1.9135	0.8691	0.9947	1.37
NSSR	2.6073	43.9690	0.9973	2.2141	1.0918	0.9939	62.23
NFSREE	1.7973	45.5846	0.9981	1.9857	0.8985	0.9943	7.44
OTD	<b>1.1946</b>	<b>47.3584</b>	<b>0.9986</b>	<b>1.7431</b>	<b>0.7779</b>	<b>0.9953</b>	40.49

method has the highest accuracy in spectral preservation. In addition, after excluding the corrupted pixels at the borders, the fusion results on the APEX and Pavia Center datasets are summarized in Table 3. It can be seen that, with the optimized

twin dictionary, our proposed OTD method shows much better fusion performance both in the spatial and spectral domain on these two commonly used datasets.

### C. Discussion on Parameter Effects

To evaluate the effect and sensitivity of the key parameters in our proposed method on the fusion performance, some of the key parameters are varied on the AVIRIS, APEX and Pavia Center datasets, such as  $K$ , the number of columns in the spectral dictionary;  $K_p$ , the number of columns in the spatial dictionary;  $\sqrt{B_p}$ , the size of image patches; the regularization parameters  $\eta$ ,  $\beta$ ,  $\lambda_1$  and  $\lambda_2$ ; the signal-to-noise ratio (SNR) for Gaussian noise; and the computing time on parameters  $K$  and  $K_p$ . More importantly, the experimental results of a reversed procedure, i.e., first applying the spatial optimization and then applying the spectral optimization, are also performed and analyzed in this section to show the importance of spectral information.

Fig.10(a) shows the effect of the parameter  $K$  on the PSNR on the three different datasets. The similar trend of the three PSNR curves indicates that our proposed method performs stably when the parameter  $K$  is larger than 100. The effects of regularization parameters  $\eta$  and  $\lambda_1$  on the fusion results are shown in Figs.10(b) and (c) in terms of PSNR. It turns out that, when the parameter  $\eta$  is between  $10^{-1}$  and  $10^0$ , and the parameter  $\lambda_1$  is larger than  $10^{-7}$ , our proposed method has a better fusion performance on the three datasets.

As for the optimization in the spatial domain, Fig.11(a) plots the PSNR curve of the fused HH image as a function of the parameter  $K_p$ . It shows that our proposed method has a stable performance when  $K_p$  is larger than 800 on the three datasets. In addition, the improvement in PSNR is limited, when  $K_p$  is larger than 1000. Figs.11(b) and (d) plot the PSNR

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



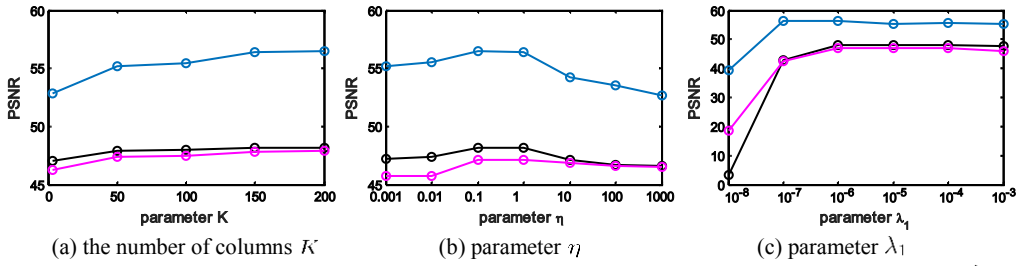


Fig.10 PSNR curves as a function of (a) the number of columns  $K$ , (b) parameter  $\eta$ , and (c) parameter  $\lambda_1$  on the AVIRIS (blue line), APEX (black line) and Pavia Center (red line) dataset.

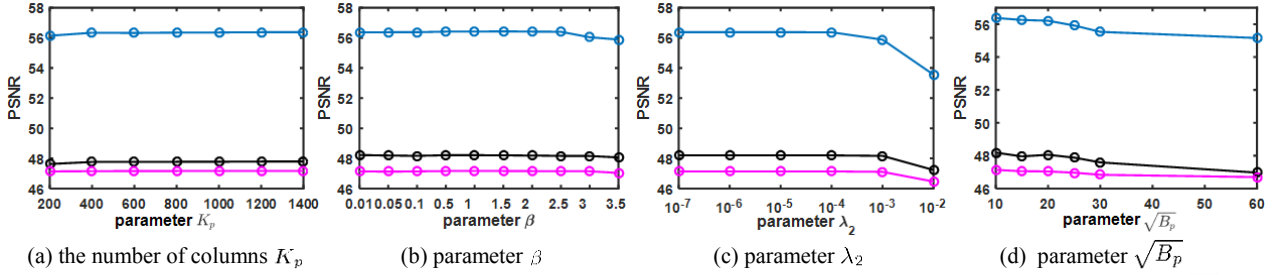


Fig.11 PSNR curves as a function of (a) the number of columns  $K_p$ , (b) parameter  $\beta$ , (c) parameter  $\lambda_2$ , and (d) parameter  $\sqrt{B_p}$  on the AVIRIS (blue line), APEX (black line) and Pavia Center (red line) dataset.

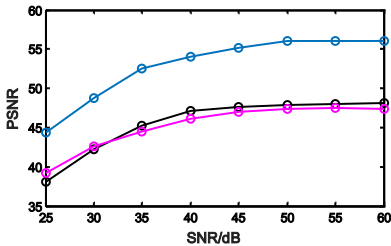


Fig.12 PSNR curves as a function of SNR on the AVIRIS (blue line), APEX (black line) and Pavia Center (red line) dataset.

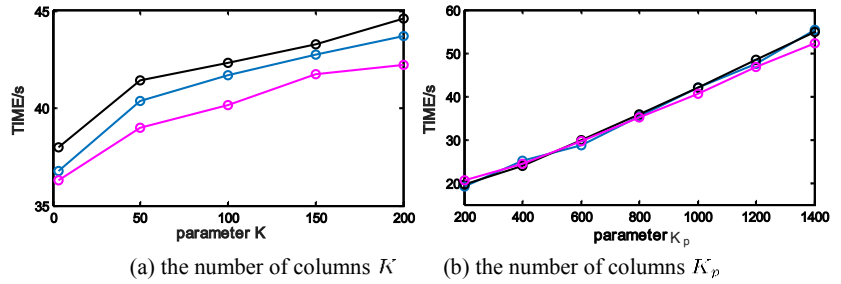


Fig.13 Computing time curves as a function of (a) the number of columns  $K$ , (b) the number of columns  $K_p$  on the AVIRIS (blue line), APEX (black line) and Pavia Center (red line) datasets.

curves as a function of parameters  $\beta$  and  $\lambda_2$ , respectively. It turns out that, when the parameter  $\beta$  is less than 2.5, and the parameter  $\lambda_2$  is less than  $10^{-4}$ , our proposed method shows a better fusion performance even on different datasets. Additionally, Fig.11(d) shows the fusion performance with different size of image patches  $\sqrt{B_p}$  of the HM image, here we should mentioned that, as the downsampling rate is set to 5, the starting patch size 10 of the HM image is determined by the smallest corresponding patch size 2 of the LH image. As can be seen that, with the increase of  $\sqrt{B_p}$ , the PSNR curve shows a downward trend. In this case, a small size of image patches is selected for better fusion performance. Due to the certain similarity between the blue, black and red curves in Figs.10 and 11, it can be indicated that our proposed OTD method has a stable performance on different datasets. Thus, the parameter settings are quite universal and applicable to most common scenarios.

Fig.12 shows the fusion performances of our proposed OTD method with different SNRs. Among the three different datasets we can see that, without a denoising process or a specific consideration of the added noise assumption, our proposed OTD method shows a stable reconstruction ability, when SNR is higher than 45dB.

Table 4 Comparisons with that first applying the spatial optimization on the AVIRIS dataset.

	MSE	PSNR	UIQI	SAM	ERGAS	ASSIM
OTDrev	0.2916	53.4833	0.9948	0.9946	0.4063	0.9975
OTD	<b>0.1406</b>	<b>56.6519</b>	<b>0.9958</b>	<b>0.7477</b>	<b>0.3149</b>	<b>0.9985</b>

As the computing time is also a key parameter for the fusion procedure, Fig.13 shows the computing time curves as a function of the typical parameters  $K$  and  $K_p$ . As can be seen in Fig.13, with more atoms in the spectral or spatial dictionary, the computing time of our OTD method will increase accordingly, while the actual computing time also depends on the parameter settings in Section III-A.

More importantly, in our proposed OTD method, the spectral optimization is followed by the spatial optimization just as the overall framework given in Fig.2, but a reversed procedure, i.e., first applying the spatial optimization and then applying the spectral optimization, is another possible way that can be considered. To show the difference between the two of them and the importance of spectral information for the targeted hyperspectral images, the main issues related to the reversed procedure will be briefly discussed. Here, the reversed procedure is named as OTDrev method, where the spatial

1 optimization utilizes the spatial information supplied by the  
 2 image patches from the LH and HM images, and the spectral  
 3 optimization utilizes the high-frequency spectral information  
 4 provided by the rest of LH and HM image. To apply this  
 5 reversed procedure, necessary modifications on the equations  
 6 in Section II should be considered firstly. Especially, the spatial  
 7 optimization in Eq.(27) and Eq.(29) should be modified as  
 8

$$9 \arg \min_{\mathbf{D}_p} \|\mathcal{Y}_M - \mathbf{D}_p \alpha_p\|_F^2 + \beta \|\mathcal{Y}_H \mathbf{L}_p - \mathbf{P} \mathbf{D}_p \alpha_p\|_F^2 + \lambda \|\alpha_p\|_0, \quad (33)$$

$$10 \arg \min_{\alpha} \|\mathcal{Y}_H - \mathbf{P} \mathbf{D}_p \alpha\|_F^2 + \beta \|\mathcal{Y}_M - \mathbf{D}_p \alpha \mathbf{L}_p\|_F^2 + \lambda_2 \|\alpha\|_1, \quad (34)$$

11 where  $\mathcal{Y}_H$  and  $\mathcal{Y}_M$  denote the stretched patches of  $\mathbf{Y}_H$  and  
 12  $\mathbf{Y}_M$ , respectively. After the spatial optimization, the error  
 13 matrices that cannot be expressed by the spatial dictionary  $\mathbf{D}_p$   
 14 in  $\mathbf{Y}_H$  and  $\mathbf{Y}_M$  will become  $\bar{\mathbf{Y}}_H = \mathbf{Y}_H - \mathbf{D}_p \circ \alpha \mathbf{H}$  and  
 15  $\bar{\mathbf{Y}}_M = \mathbf{Y}_M - \mathbf{L} \mathbf{D}_p \circ \alpha$ , the spectral optimization in Eq.(12)  
 16 can be further modified as

$$17 \arg \min_{\mathbf{D}_s, \mathbf{A}} \|\bar{\mathbf{Y}}_M - \mathbf{L} \mathbf{D}_s \mathbf{A}\|_F^2 + \eta \|\bar{\mathbf{Y}}_H - \mathbf{D}_s \mathbf{A} \mathbf{H}\|_F^2 + \lambda_1 \|\mathbf{A}\|_1. \quad (35)$$

18 With the same dataset and parameter settings in Section III-A,  
 19 Table 4 shows a comparison result between OTD and OTDrev  
 20 methods on the AVIRIS dataset. Compared with the OTDrev  
 21 method, OTD method improves more than 3.1dB in PSNR, and  
 22 reduces more than 0.24 in SAM, but OTDrev method is still the  
 23 second best in most quality metrics than the other remaining  
 24 methods in Table 1. It means that our proposed OTD method,  
 25 which firstly applies spectral optimization, has a better  
 26 performance in both spectral and spatial preservation. That is  
 27 probably because the spectral information is much more  
 28 important than the spatial information for the fusion problem  
 29 between HM and LH image, if our target is set to the high-  
 30 spatial-resolution hyperspectral image.

#### 31 IV. CONCLUSION

32 In this paper, a new hyperspectral image fusion method  
 33 using optimized twin dictionaries was proposed to obtain the  
 34 HH image from fusing an LH and an HM image over the same  
 35 scene. The optimized twin dictionaries, including one spectral  
 36 dictionary and one spatial dictionary, can fully integrate the  
 37 spectral information and the high-frequency spatial information  
 38 into the final HH images simultaneously. The spectral  
 39 dictionary, representing more generalized spectrums, is  
 40 optimized by minimizing an augmented Lagrangian function  
 41 using ADMM without non-negative constraints. The spatial  
 42 dictionary, representing high-frequency information, is learned  
 43 by the K-SVD algorithm from the remaining spatial  
 44 information left by the spectral dictionary. Experimental results  
 45 on various datasets demonstrate the superiority of our proposed  
 46 OTD method in both spatial and spectral preservations than  
 47 the other relative state-of-the-art fusion methods.

#### REFERENCES

- [1] M. Jiang, F. Cao, and Y. Lu, "Extreme learning machine with enhanced composite feature for spectral-spatial hyperspectral image classification," *IEEE Access*, vol. 6, pp. 22645-22654, 2018.
- [2] A. Qin, Z. Shang, J. Tian, T. Zhang, Y. Y. Tang, and J. Qian, "Edge-smoothing-based distribution preserving hyperspherical embedding for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018.
- [3] M. Díaz, R. Guerra, S. López, and R. Sarmiento, "An algorithm for an accurate detection of anomalies in hyperspectral images with a low computational complexity," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 1159-1176, 2018.
- [4] S. Matteoli, M. Diani, and G. Corsini, "Automatic target recognition within anomalous regions of interest in hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 4, pp. 1056-1069, 2018.
- [5] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," *IEEE ECCV*, pp. 63-78, 2014.
- [6] W. J. Carper, T. M. Lillesand and R. W. Kiefer, "The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data." *Photogramm. Eng. Remote Sens.* vol. 56, no.4, pp. 457-467, 1990.
- [7] C. Jiang, H. Zhang, H. Shen and L. Zhang, "A practical compressed sensing-based pan-sharpening method," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 4, pp. 629-633, 2012.
- [8] M. R. Vicinanza, R. Restaino, G. Vivone, M. Dalla Mura and J. Chanussot, "A Pansharpening Method Based on the Sparse Representation of Injected Details," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 1, pp. 180-184, Jan. 2015.
- [9] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS + Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230-3239, Oct. 2007.
- [10] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent 6 011 875 A, Jan. 4, 2000.
- [11] J. G. Liu, "Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461-3472, Jan. 2000.
- [12] 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, June 2017.
- [13] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti, "Hypersharpening: A first approach on SIM-GA data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 3008-3024, 2015.
- [14] M. Selva, L. Santurri and S. Baronti, "Improving Hypersharpening for WorldView-3 Data," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 6, pp. 987-991, June 2019.
- [15] M. R. Vicinanza, R. Restaino, G. Vivone, M. Dalla Mura and J. Chanussot, "A Pansharpening Method Based on the Sparse Representation of Injected Details," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 1, pp. 180-184, Jan. 2015.
- [16] X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2827-2836, May 2013.
- [17] X. X. Zhu, C. Grohnfeldt and R. Bamler, "Exploiting joint sparsity for pansharpening: the J-SparseFI algorithm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2664-2681, 2016.
- [18] C. Grohnfeldt, X. X. Zhu, and R. Bamler. "Jointly sparse fusion of hyperspectral and multispectral imagery." *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*. IEEE, 2013.
- [19] T. Zhao, Y. Zhang, X. Xue and M. He, "Hyperspectral and multispectral image fusion using collaborative representation with local adaptive dictionary pair." *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*, pp. 7212-7215, 2016.
- [20] E. K. Ghasrodashti, A. Karami, R. Heylen, and P. Scheunders, "Spatial resolution enhancement of hyperspectral images using spectral unmixing and bayesian sparse representation." *Remote Sensing*, vol. 9, no. 6, pp. 1-20, 2017.
- [21] N. Keshava, and J. F. Mustard, "Spectral unmixing." *IEEE signal processing magazine*, vol.19, no.1, pp. 44-57, 2002.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- [22] 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.
- [23] N. Yokoya, T. Yairi, A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 528-537, 2012.
- [24] M. Simoes, J. Bioucas-Dias, L. Almeida, J. A. Chanussot, "Convex formulation for hyperspectral image superresolution via subspace-based regularization". *IEEE Trans. Geosci. Remote Sens.*, vol. 53, pp. 3373-3388, 2015.
- [25] 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, 2016.
- [26] X. Han, J. Luo, J. Yu, W. Sun, "Hyperspectral image fusion based on non-factorization sparse representation and error matrix estimation", In *Signal and Information Processing (GlobalSIP), IEEE Global Conference on*, pp. 1155-1159, 2017.
- [27] Z. H. Nezhad, A. Karami, R. Heylen, and P. Scheunders, "Fusion of hyperspectral and multispectral images using spectral unmixing and sparse coding." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 6, pp. 2377-2389, 2016.
- [28] C. Yi, Y. Q. Zhao, and J. C. W. Chan, "Hyperspectral image super-resolution based on spatial and spectral correlation fusion", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no.7, pp. 4165-4177, 2018.
- [29] Q. Wei, N. Dobigeon, and J. Y. Tourneret, "Bayesian fusion of hyperspectral and multispectral images," In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE*, pp. 3176-3180, 2014.
- [30] 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.
- [31] Q. Wei, N. Dobigeon, and J. Y. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Transactions on Image Process*, vol. 24, no. 11, pp. 4109-4121. 2015.
- [32] K. Zhang, M. Wang, S. Yang, and L. Jiao, "Spatial-spectral-graph-regularized low-rank tensor decomposition for multispectral and hyperspectral image fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 4, pp. 1030-1040, 2018.
- [33] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4118-4130, 2018.
- [34] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 639-643, 2017.
- [35] F. Zhou, R. Hang, Q. Liu, and X. Yuan, "Pyramid Fully Convolutional Network for Hyperspectral and Multispectral Image Fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol.12, pp. 1549-1558, 2019
- [36] X. Han, J. Yu, J. Luo, and W. Sun, "Hyperspectral and Multispectral Image Fusion using Cluster-based Multi-branch BP Neural Networks," *Remote Sensing*, vol.11, no. 10, pp. 1173, 2019.
- [37] 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.
- [38] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311-4322, 2006.
- [39] Z. X. Pan, J. Yu, C. Xiao, and W. D. Sun, "Single image super resolution based on adaptive multi-dictionary learning." *Acta Electron. Sin.*, vol. 43, pp. 209-216, 2015.
- [40] M. Elad, and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries." *IEEE Transactions on Image processing*, vol. 15, no. 12, pp. 3736-3745, 2006.
- [41] E. Wycoff, T. H. Chan, K. Jia, W. K. Ma, & Y. Ma, "A non-negative sparse promoting algorithm for high resolution hyperspectral imaging," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1409-1413, 2013.
- [42] C. H. Lin, F. Ma, C. Y. Chi, & C. H. Hsieh. "A convex optimization-based coupled nonnegative matrix factorization algorithm for hyperspectral and multispectral data fusion", *IEEE Transactions on Geoscience and Remote Sensing*, 2017.
- [43] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization", *Journal of optimization theory and applications*, vol. 109, no. 3, pp. 475-494, 2001.
- [44] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit." *Cs Technion*, vol. 40, no. 8, pp. 1-15, 2008.
- [45] Z. X. Pan, J. Yu, H. J. Huang, S. X. Hu, A. W. Zhang, H. B. Ma, and W. D. Sun, "Super-resolution based on compressive sensing and structural self-similarity for remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 9, pp. 4864-4876, 2013.
- [46] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity." *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600-612, 2004.
- [47] <http://aviris.jpl.nasa.gov>.
- [48] <http://www.apex-esa.org/content/free-data-cubes>.
- [49] [http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes).