

Unaligned Hyperspectral Image Fusion via Registration and Interpolation Modeling

Jiacheng Ying[✉], Hui-Liang Shen[✉], and Si-Yuan Cao

Abstract—In satellite remote sensing, the hyperspectral sensor acquires high-spectral-resolution and low-spatial-resolution hyperspectral images (HSIs). Conversely, the multispectral sensor acquires low-spectral-resolution and high-spatial-resolution multispectral images (MSIs). Thus, HSI and MSI fusion is required to promote both spatial and spectral resolutions. Currently, most algorithms are based on the assumption that the HSI and MSI are perfectly aligned. However, this is hardly achievable in real scenarios when the two sensors acquire images from different viewpoints. In this article, we propose a fusion algorithm that consists of two stages, i.e., image registration and image fusion. For image registration, we introduce the normalized edge difference (NED) for image similarity measure considering the different resolutions of the original images. For image fusion, we incorporate the interpolation process in the spatial degradation model to compensate for the interpolation error. Experimental results show that our algorithm performs better than the state of the arts for unaligned image fusion.

Index Terms—Hyperspectral image (HSI), image fusion, image interpolation, image registration, image super-resolution, multispectral image (MSI).

I. INTRODUCTION

HYPERSPECTRAL imaging can record rich spectral information and has been applied in many fields, such as material classification [1], face recognition [2], and geological exploration [3], [4]. However, due to the limitation of physical devices, images captured by hyperspectral sensors are usually of low spatial resolution. Conversely, multispectral sensors can capture high spatial resolution but low-spectral-resolution images. Hence, many image fusion algorithms [5]–[7] have been introduced to improve both the spatial and spectral resolutions. The aim of this work is to reconstruct a high-resolution hyperspectral image (HR-HSI) by fusing the low-resolution HSI (LR-HSI) and high-resolution multispectral image (HR-MSI).

Most of the current algorithms are implicitly based on the hypothesis that the HSI and the MSI are perfectly aligned.

Manuscript received March 27, 2021; revised April 23, 2021; accepted May 13, 2021. Date of publication May 27, 2021; date of current version January 5, 2022. This work was supported by the National Natural Science Foundation of China under Grant 81973751. (Corresponding author: Hui-Liang Shen.)

Jiacheng Ying and Si-Yuan Cao are with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: yingjiacheng@zju.edu.cn; karlcao@hotmail.com).

Hui-Liang Shen is with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China, and also with the Ningbo Research Institute, Zhejiang University, Ningbo 315100, China (e-mail: shenhl@zju.edu.cn).

Digital Object Identifier 10.1109/TGRS.2021.3081136

However, this situation can be hardly met if without special optical design. On the remote sensing satellite, two types of sensors acquire image pairs from different viewpoints, which will eventually cause image misalignment.

In this work, we propose an HSI super-resolution algorithm based on image registration and interpolation modeling. We assume that the spatial relationship between LR-HSI and HR-MSI can be depicted by affine transform. This assumption is reasonable since the satellite is far away from the ground scene to be imaged. Our algorithm contains a registration stage and a fusion stage. In the registration stage, we first apply spectral mapping and spatial upsampling such that the LR-HSI is of the same size as HR-MSI. With this treatment, the original registration problem is transformed to an MSI clear-blur pair registration problem. Generally, the edge difference (ED) between two aligned images is smaller than that between two unaligned images [8]. Hence, we define the normalized ED (NDE) to measure the misregistration error of two images. We also build edge image pyramids to perform multiscale registration. In the fusion stage, we apply a special strategy to utilize the affine parameters to construct a spatial degradation model from HR-HSI to LR-HSI. We also construct the spectral degradation model using the spectral mapping function of the image sensors. We decompose the HR-HSI to the spectral basis matrix and corresponding coefficients' matrix. Our method is based on the prior that each pixel can be represented linearly using a small number of spectral bases. We obtain the spectral basis matrix by nonnegative matrix factorization (NMF) of LR-MSI and compute the corresponding coefficients' matrix by solving a Sylvester equation [9] with L2-norm regularization.

We notice that Nie *et al.* [10] and Zhou *et al.* [11] also deal with the HSI super-resolution problem in the case of image misalignment. These works first estimate the transform parameters and then directly fuse the registered images. We note that, due to the transform bias and interpolation error, direct fusion is optimal for HSI super-resolution. Instead, we incorporate the interpolation process of the affine transform into the spatial degradation model in the image fusion stage. As can be seen in the experiment, this treatment can considerably improve the accuracy of image fusion.

In summary, the main contributions of this work are given as follows.

- 1) We propose a two-stage HS image super-resolution algorithm. Compared with the state of the arts, our algorithm takes into consideration that the remote sensing HSI and

MSI are unaligned, which is a more general situation in real-world applications.

- 2) For image registration, we introduce a new registration objective function called NED that is effective for clear-blur image similarity measurement.
- 3) For image fusion, we incorporate the interpolation process into the spatial degradation model. Compared with the regular treatment in previous works, our treatment effectively reduces the influence of transform bias and interpolation error in HR-HSI reconstruction.

II. RELATED WORK

In the following, we briefly review the work on image registration and fusion that closely relates to HSI super-resolution.

A. Image Registration

The objective of image registration is to find the best spatial transform between two images. Approximately, the registration algorithms can be classified into two categories: feature-based ones and intensity-based ones.

Feature-based algorithms detect feature points in images and perform image alignment by feature points matching. Scale-invariant feature transform (SIFT) [12] is a very common feature descriptor and has been widely used for image registration. It has good stability and is not sensitive to scaling, rotation, and luminance change. The speeded-up robust feature (SURF) descriptor [13] is a 64-D feature vector that is half of SIFT and can accelerate SIFT. It is good at dealing with blurred and rotated images. The Harris corner detector [14] extracts the intersection points between contours as features and is robust to viewpoint change. Brown *et al.* [15] extract SIFT feature points in images and use K-D tree and random sample consensus (RANSAC) [16] for image matching. Fan *et al.* [17] employ the Harris corner filter for pixel classification, which simultaneously improves the robustness of multimodal image registration and reduces computation complexity.

Intensity-based algorithms consider every pixel and introduce some measures that represent the image consistency for image alignment. Maes *et al.* [18] register images by employing the mutual information, which reaches the maximum when two images are perfectly aligned. Irani and Anandan [19] introduce the local normalized-correlation similarity, which measures the consistency of two images under different luminance. Myronenko and Song [20] align images by minimizing the residual complexity of two images, which can be represented by the sparseness of discrete cosine transform (DCT) coefficients. Chen *et al.* [8] introduce normalized total gradient (NTG) to represent the sparseness of gradient of the difference image, by minimizing which multispectral image alignment is achieved. Reddy and Chatterji [21] present phase correlation (PC) technique to cover translation, rotation, and scaling. Bulent *et al.* [22] apply diffeomorphic registration by optimizing a similarity measure for fine-tuned alignment after using SIFT and SURF for initial image alignment.

B. Image Fusion

Most HSI fusion algorithms are based on the assumption that two images are aligned. Guided by a secondary HR-MSI, the spatial resolution of the LR-HSI is improved. The typical algorithms in image fusion include matrix decomposition, tensor factorization, spectrum mapping, and deep learning.

Matrix decomposition algorithms are based on the assumption that the spectrum vector of each pixel can be represented as a linear combination of several spectral bases. The SRIF algorithm [23] employs principal component analysis in extracting spectral dictionary and incorporates gradient regularization in estimating the corresponding coefficients. Simoes *et al.* [5] extract the subspace of HSI as the spectral basis and employ the alternating direction method of multipliers (ADMM) to solve an optimization problem, with vector total variation of coefficients gradient as the regularizer. Yokoya *et al.* [24] use NMF to unmix both the HSI and MSI into endmember and abundance matrices and iteratively optimize one matrix with the other one fixed.

Tensor factorization algorithms decompose the HSI into a core tensor and three corresponding dictionaries. Li *et al.* [6] employ a regularizer to model the correlation between space and spectrum domains that improve the sparseness of the core tensor. Dian *et al.* [25] group the similar cubes together for the learning of dictionaries in three modes and use sparse coding to estimate core tensor. Ding *et al.* [52] deploy a block-term factorization model with multirank terms, which allows prior information incorporated for performance improvement.

Spectrum mapping algorithms map spectrum vectors of each pixel from low dimension to high dimension. Arad and Ben-Shahar [26] generate a large number of RGB atoms and hyperspectral atoms to construct the RGB image and HSI, respectively, with the weights of atoms shared. Similarly, Aeschbacher *et al.* [27] train the nearest neighbor relations between image pixels for HSI reconstruction.

Deep learning algorithms achieve super-resolution by constructing neural networks to train the mapping function from LR images to HR images in a supervised manner. Wang *et al.* [28] train a multiunit fusion network and iteratively plus the residual output of former units together as the final output, to increase the accuracy of the constructed images. Zhang *et al.* [29] build a two-stage SR network. The first stage uses a supervised fusion module to obtain a rough result, and the second stage deploys an unsupervised adaption module to further improve the image quality. Dian *et al.* [30] build a CNN network to train the relation between roughly constructed HR-HSIs and the residual of HR-HSIs, which can improve the accuracy of image reconstruction.

III. PROBLEM FORMULATION

We assume that the HR-MSI, denoted as $\mathbf{X} \in \mathbb{R}^{M \times N \times l}$, includes $M \times N$ pixels in l bands. The LR-HSI $\mathbf{Y} \in \mathbb{R}^{m \times n \times L}$ includes $m \times n$ pixels in L bands, with $L \gg l$. The scale factor is denoted as b , relating the spatial dimensions by $M = b \times m$ and $N = b \times n$. Our objective is to estimate the HR-HSI $\mathbf{Z} \in \mathbb{R}^{M \times N \times L}$ from $\mathbf{X}$ and $\mathbf{Y}$.

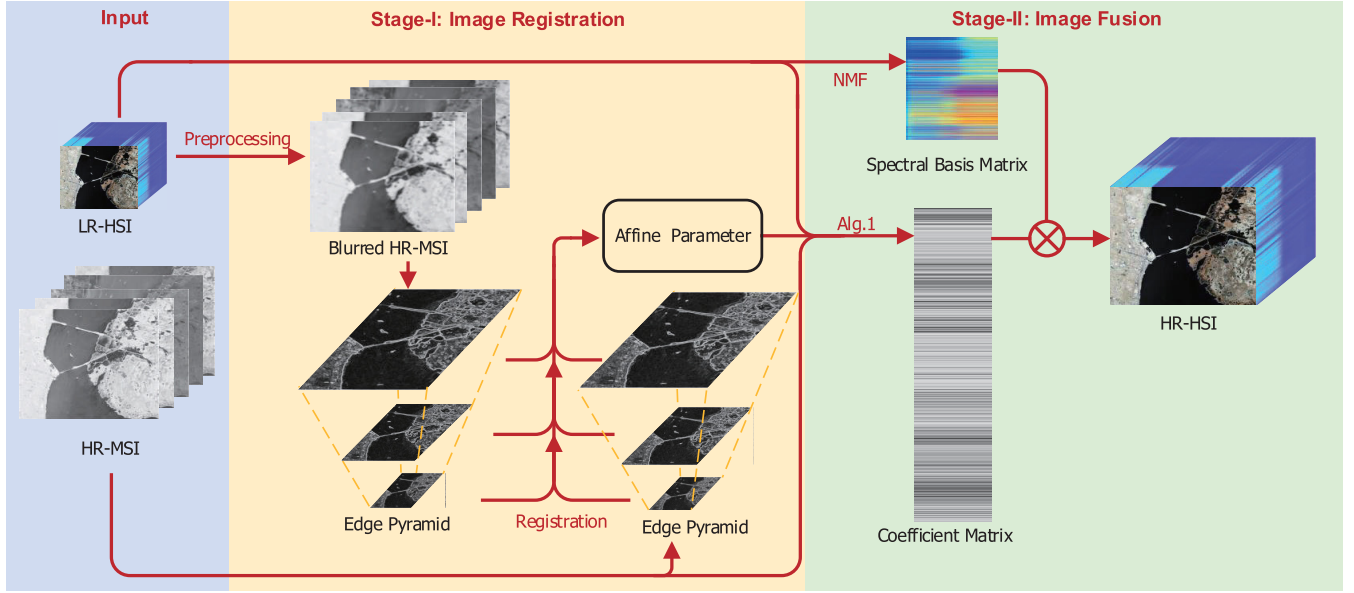


Fig. 1. Flowchart of our proposed algorithm that contains two stages. In the registration stage, image preprocessing and image pyramids techniques are applied to estimate the affine parameter between two unaligned images. In the fusion stage, spectral basis matrix and corresponding coefficient matrix are solved, respectively, using NMF and Algorithm 1 to reconstruct HR-HSI.

The matrix forms of $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ are $\hat{\mathbf{X}} \in \mathbb{R}^{l \times MN}$, $\hat{\mathbf{Y}} \in \mathbb{R}^{L \times mn}$, and $\hat{\mathbf{Z}} \in \mathbb{R}^{L \times MN}$. Then, $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ can be linearly modeled with $\hat{\mathbf{Z}}$ as

$$\hat{\mathbf{X}} = \mathbf{R}\hat{\mathbf{Z}} \quad (1)$$

$$\hat{\mathbf{Y}} = \hat{\mathbf{Z}}\mathbf{B}\mathbf{S} \quad (2)$$

where $\mathbf{R} \in \mathbb{R}^{l \times L}$ denotes the spectral degradation function mapping the hyperspectral pixels to the multispectral ones; $\mathbf{B} \in \mathbb{R}^{MN \times MN}$ represents the deblurring matrix obtained from the point spread function (PSF) of the hyperspectral sensors; and $\mathbf{S} \in \mathbb{R}^{MN \times mn}$ is the spatial mapping matrix that maps the HR-HSI to the LR-HSI in the spatial domain. We note that $\mathbf{R}$ and PSF are the parameters of the hyperspectral sensor and are supposed to be known.

As the spatial relation from $\mathbf{Z}$ to $\mathbf{Y}$ is identical to that from $\mathbf{X}$ to $\mathbf{Y}$, we compute $\mathbf{S}$ by aligning $\mathbf{X}$ and $\mathbf{Y}$ in the registration stage of our algorithm. We first apply spectral mapping and spatial upsampling to $\mathbf{Y}$ to generate a blurred HR-MSI, which is also of size $M \times N \times l$. Then, the different scales' registration problem is transferred to a clear-blur image pair registration problem. Let $\mathcal{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ denote the transform from the clear MSI to the blurred MSI, and

$$\mathcal{T}(\mathbf{x}) = \begin{pmatrix} x & y & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 \end{pmatrix} \mathbf{A} \quad (3)$$

where $\mathbf{x} = (x, y)^\top$ is the original pixel position, $\mathcal{T}(\mathbf{x}) = (u, v)^\top$ is the transformed pixel position, and $\mathbf{A} = (a_1, a_2, \dots, a_6)^\top \in \mathbb{R}^{6 \times 1}$ is the affine parameter.

To be more specific, Fig. 1 illustrates the flowchart of our algorithm, which contains two stages. In the registration stage, the input LR-HSI is first converted to a blurred HR-MSI. Next, the clear HR-MSI and the blurred HR-MSI, which are unaligned, are used to construct the edge pyramids. Then,

the affine parameter is optimized and transferred from the bottom layer to the top layer. In the fusion stage, the spectral basis matrix is solved by applying NMF, and the coefficient matrix is solved using Algorithm 1. Finally, the HR-HSI is reconstructed by multiplying two matrices.

IV. STAGE I: IMAGE REGISTRATION

The aim of image registration in our algorithm is to find the best affine transformation from HR-MSI $\mathbf{X}$ to LR-HSI $\mathbf{Y}$. More details will be elaborated on as follows.

A. Image Preprocessing

As the size of HR-MSI $\mathbf{X}$ is inconsistent with that of LR-HSI $\mathbf{Y}$ in the spectral and spatial domains, we apply spectral mapping and spatial upsampling to the LR-HSI $\mathbf{Y}$ to generate a blurred HR-MSI $\tilde{\mathbf{Y}} \in \mathbb{R}^{M \times N \times l}$ as

$$\tilde{\mathbf{Y}} = [\mathbf{R} \circ \mathbf{Y}]_{\uparrow b} \quad (4)$$

where $\circ$ denotes the spectral mapping operation, $[\cdot]_{\uparrow}$ the upsampling operation (here, we use the bicubic manner), and b the scale factor. Here, we do not choose to downsample the HR-MSI $\mathbf{X}$ to the size of $m \times n \times l$ because the registration error in the $m \times n$ sized pictures is b times as large as that in the $M \times N$ sized pictures. Hence, the multiscale registration problem is transferred to a clear-blur image pair ($\mathbf{X}$ and $\tilde{\mathbf{Y}}$) registration problem.

B. Edge Difference

The image edge can well reflect the structure information of an image. A clear-blur image pair, whose pixels' intensities are different, have similar edge structures. Generally, when two images are perfectly aligned, the ED reduces to the minimum.

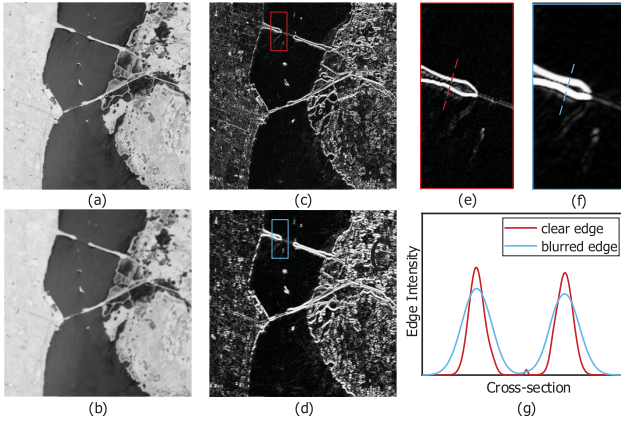


Fig. 2. Edge information in a clear-blur image pair. (a) Clear image. (b) Blurred image. (c) Edge of clear image. (d) Edge of blurred image. (e) Details of clear edge. (f) Details of blurred edge. (g) Cross-sectional distributions of clear and blurred edges.

Therefore, we define the ED of clear-blur image pair $\mathbf{I}_c$ and $\mathbf{I}_b$ as

$$ED(\mathbf{I}_c, \mathbf{I}_b) = \sum_j \|\mathbf{E}_c - \mathbf{E}_b\|_1 \quad (5)$$

where $\|\cdot\|_1$ denotes the L1-norm, $\mathbf{E}_c$ and $\mathbf{E}_b$ are the edges of $\mathbf{I}_c$ and $\mathbf{I}_b$, respectively, and j denotes the image band. Here, we use the L1-norm to compute ED instead of L2-norm because L1-norm can promote the sparseness of ED and is robust to noise. The edge $\mathbf{E}$ of the image $\mathbf{I}$ is computed using

$$E(x, y, j) = \sqrt{\partial_x^2 I(x, y, j) + \partial_y^2 I(x, y, j)} \quad (6)$$

where $E(x, y, j)$ is the edge intensity at pixel position (x, y) at the band j , and $\partial_x I(x, y, j)$ and $\partial_y I(x, y, j)$ are the image gradients computed using the spatial convolution kernels $[-1, 0, 1]$ and $[-1, 0, 1]^T$, respectively.

Fig. 2 illustrates the edge information in a clear-blur image pair. It is observed that the clear image edge is sharper than the blurred image edge. Fig. 2(g) presents the cross-sectional distributions of the two kinds of edges. It is obvious that the distribution of clear image edge is of a higher crest and narrower width.

For demonstration, we use two 1-D Gaussian models with different variances to simulate the registration of clear-blur edge pair in Fig. 3. We compute the ED when the narrow Gaussian model moves from displacement -7 to 7 . It is observed that ED reaches the minimum when there is no displacement.

C. Normalized Edge Difference

To limit the range of the measure ED to $[0, 1]$, we define NED of the clear-blur image pair $\mathbf{I}_c$ and $\mathbf{I}_b$ as

$$NED(\mathbf{I}_c, \mathbf{I}_b) = \frac{\sum_j \|\mathbf{E}_c - \mathbf{E}_b\|_1}{\sum_j (\|\mathbf{E}_c\|_1 + \|\mathbf{E}_b\|_1)} \quad (7)$$

where the denominator is the absolute sum of two image edges, and the numerator is the absolute sum of ED. If we

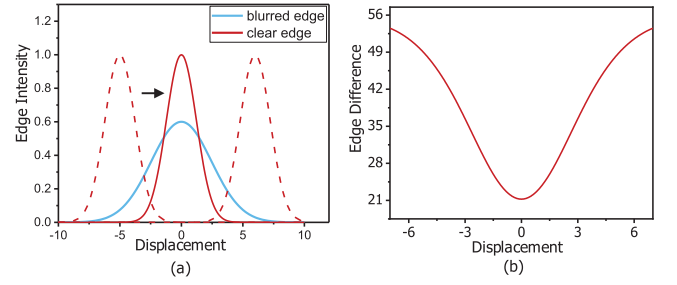


Fig. 3. Demonstration of clear-blur image registration using 1-D Gaussian models with different variances. (a) Simulation of image registration by moving the narrower Gaussian model (clear edge) with respect to the wider Gaussian model (blur edge). (b) Trend of ED with respect to the displacement.

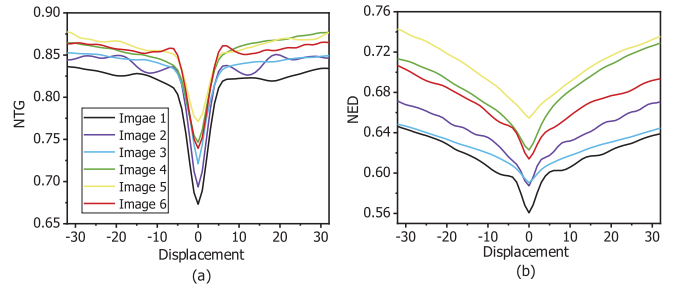


Fig. 4. Trends of (a) NTG and (b) NED values with respect to the displacement on six images.

treat the L1-norm as an energy operator, NED can be regarded as the relative energy of ED.

We note that NTG [8] has a similar form to our NED measure. We compare the two measures by imposing displacements on the clear images of six clear-blur pairs in both horizontal and vertical directions. Fig. 4 presents the NTG and NED values with respect to the displacement. The NTG measure is more likely to fall into the local optimum when the displacements are larger than 10 pixels. In comparison, NED has a wider capture range that is beneficial to find the real solution.

Consequently, the affine parameter $\mathbf{A}$ of the clear-blur image pair registration problem can be solved by minimizing NED as

$$\mathbf{A} = \arg \min_{\mathbf{A}} J(\mathbf{A}) \quad (8)$$

where

$$J(\mathbf{A}) = NED(\mathbf{I}_c(\mathcal{T}(\mathbf{x}, \mathbf{A})), \mathbf{I}_b(\mathbf{x})). \quad (9)$$

To compute $\mathbf{A}$, we initialize it as $\mathbf{A}_0 = (1, 0, 0, 0, 1, 0)^T$ and update it by gradient descent

$$\mathbf{A}^{t+1} = \mathbf{A}^t - \gamma \nabla_{\mathbf{A}}^t J \quad (10)$$

where the step size $\gamma = 0.5/(\max\{M, N\})$. $\nabla_{\mathbf{A}}^t J$ denotes the current gradient value. More details about the derivation of $\nabla_{\mathbf{A}} J$ can be found in the Appendix.

D. Edge-Image Pyramids

Only using the gradient descent method to minimize NED may cause the affine parameter to fall into the local optimum.

To solve this problem, we perform multiscale registration by building edge-image pyramids, whose lower layers are downsampled from the upper ones. We fix the downsampling factor as $\alpha = 1.5$ and require the short side of the bottom layer larger than 16 pixels. Starting from the bottom layer, we first initialize the affine parameter, update it using (10), and transfer it to the upper layer.

Let $\mathbf{A}^s = (a_1^s, a_2^s, \dots, a_6^s)^\top$ denote the affine parameter at the s th layer; we have

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a_1^s x + a_2^s y + a_3^s \\ a_4^s x + a_5^s y + a_6^s \end{pmatrix}. \quad (11)$$

By transferring the affine parameter from the s th layer to the $(s+1)$ th layer with α times size enlargement, we have

$$\begin{pmatrix} u & x \\ v & y \end{pmatrix} \rightarrow \begin{pmatrix} \alpha u & \alpha x \\ \alpha v & \alpha y \end{pmatrix}. \quad (12)$$

Hence, we need to update a_3^s and a_6^s to αa_3^s and αa_6^s , respectively, for parameter transfer. Then, we use (10) to further optimize $\mathbf{A}^{s+1}$ in the $(s+1)$ th layer. In this way, the affine parameter is solved in a coarse-to-fine manner.

We note that the employment of edge-image pyramids is essential to the image registration process. The parameter transfer between adjacent layers makes the algorithm more accurate and robust and also accelerates its convergence.

V. STAGE II: IMAGE FUSION

Based on the spectrum unmixing principle, each pixel spectrum $\hat{\mathbf{z}}_i \in \mathbb{R}^L$ in the HR-HSI $\hat{\mathbf{Z}}$ can be represented as the linear combination of a few spectral bases as

$$\hat{\mathbf{z}}_i = \mathbf{V} \mathbf{e}_i \quad (13)$$

where $\mathbf{V} \triangleq (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k)$ denotes the spectral basis matrix, in which each column is a spectral basis. $\mathbf{e}_i \in \mathbb{R}^k$ is the coefficient vector. Combining all the pixels, we have

$$\hat{\mathbf{Z}} = \mathbf{V} \mathbf{E} \quad (14)$$

where $\mathbf{E} \triangleq (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{MN})$ denotes the corresponding coefficient matrix. With (14), we rewrite (1) and (2) as

$$\hat{\mathbf{X}} = \mathbf{R} \mathbf{V} \mathbf{E} \quad (15)$$

and

$$\hat{\mathbf{Y}} = \mathbf{V} \mathbf{E} \mathbf{B} \mathbf{S}. \quad (16)$$

As mentioned in Section III, $\mathbf{R}$ and $\mathbf{B}$ are determined by the hyperspectral sensor and are, thus, known. $\mathbf{S}$ is the spatial mapping matrix.

Most of the current algorithms treat the registration and fusion as two independent parts. These algorithms first warp LR-HSI (or warp HR-MSI) to register misaligned images and then fuse the two images. With this treatment, $\mathbf{S}$ is constructed as a downsampling matrix, as shown in Fig. 5(a). Different from the previous work, we employ an interpolation modeling treatment to construct the matrix $\mathbf{S}$ using the affine parameter obtained from Section IV. More specifically, a pixel (x, y) in $\hat{\mathbf{Y}}$ is mapped from subpixels in $\hat{\mathbf{Z}}$. We apply interpolation in a bilinear manner (the reason will be stated in Section V-C)

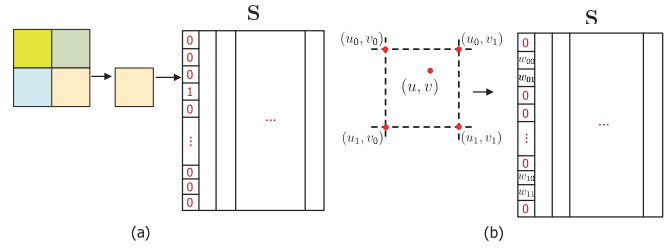


Fig. 5. Construction of matrix $\mathbf{S}$. (a) Traditional treatment without interpolation modeling. (b) Our treatment with interpolation modeling.

by filling different interpolation weights $\{w_{00}, w_{01}, w_{10}, w_{11}\}$ of nearest pixels $\{(u_0, v_0), (u_1, v_0), (u_0, v_1), (u_1, v_1)\}$ into the corresponding positions when constructing $\mathbf{S}$, as illustrated in Fig. 5(b). With our modeling treatment, LR-HSI does not need to be warped, and thus, the negative influence of interpolation error and registration bias to the reconstructing HSIs can be reduced.

A. Computing Spectral Basis Matrix

Physically, different spectral bases present different material endmembers, and a pixel in remote sensing images usually contains one or several endmembers [31]. It is reasonable to assume that the spectral basis matrix and the corresponding coefficients' matrix are nonnegative. We employ NMF [24] to solve the spectral bases by promoting the sparse distribution of the coefficients.

Let

$$\mathbf{E}_Y = \mathbf{E} \mathbf{B} \mathbf{S}. \quad (17)$$

Equation (16) can be rewritten as

$$\hat{\mathbf{Y}} = \mathbf{V} \mathbf{E}_Y \quad (18)$$

which demonstrates that LR-HSI $\hat{\mathbf{Y}}$ can also be represented by spectral bases. Since $\mathbf{E}$, $\mathbf{B}$, and $\mathbf{S}$ are all nonnegative and sparse, $\mathbf{E}_Y$ is also a nonnegative and sparse matrix. We employ NMF to estimate the spectral basis matrix, formulated as

$$\begin{aligned} \mathbf{V} &= \arg \min_{\mathbf{V}} \|\mathbf{E}_Y\|_1 \\ \text{s.t. } &\|\hat{\mathbf{Y}} - \mathbf{V} \mathbf{E}_Y\|_F^2 \leq \epsilon, \quad \mathbf{V} \geq \mathbf{0}, \quad \mathbf{E}_Y \geq \mathbf{0} \end{aligned} \quad (19)$$

where $\|\cdot\|_F$ denotes Frobenius-norm, and ϵ is the tolerance to noise. Here, we treat matrix $\mathbf{E}_Y$ as a vector in the computation of L1-norm. We solve this problem using the SParse Modeling Software (SPAMS) [32].

B. Solving the Coefficient Matrix

With the estimated spectral basis matrix $\mathbf{V}$, we further solve the coefficient matrix $\mathbf{E}$ for reconstructing $\mathbf{Z}$. We formulate the problem as

$$\begin{aligned} \mathbf{E} &= \arg \min_{\mathbf{E}} \Psi(\mathbf{E}) \\ &= \arg \min_{\mathbf{E}} \|\hat{\mathbf{Y}} - \mathbf{V} \mathbf{E} \hat{\mathbf{S}}\|_F^2 + \eta \|\hat{\mathbf{X}} - \hat{\mathbf{R}} \mathbf{E}\|_F^2 + \gamma \|\mathbf{E}\|_F^2 \end{aligned} \quad (20)$$

where $\hat{\mathbf{S}} = \mathbf{B} \mathbf{S}$ and $\hat{\mathbf{R}} = \mathbf{R} \mathbf{V}$. The first and second terms of the right-hand side are to minimize the reconstruction error, and

the third term is a regularizer. The parameters η and γ control the balance of these terms; their influence will be analyzed in Section V-C.

The problem (20) can be solved by convex optimization. Let the derivation of $\Psi(\mathbf{E})$ be zero; we have

$$\mathbf{V}^\top \mathbf{V} \mathbf{E} \hat{\mathbf{S}} \hat{\mathbf{S}}^\top + (\eta \hat{\mathbf{R}}^\top \hat{\mathbf{R}} + \gamma \mathbf{I}) \mathbf{E} = \mathbf{V}^\top \hat{\mathbf{Y}} \hat{\mathbf{S}}^\top + \eta \hat{\mathbf{R}}^\top \hat{\mathbf{X}}. \quad (21)$$

Left multiplying $(\mathbf{V}^\top \mathbf{V})^{-1}$ on both sides, (21) can be formulated as a Sylvester equation

$$\mathbf{P} \mathbf{E} + \mathbf{E} \mathbf{Q} = \mathbf{C} \quad (22)$$

where

$$\mathbf{P} = (\mathbf{V}^\top \mathbf{V})^{-1} (\eta \hat{\mathbf{R}}^\top \hat{\mathbf{R}} + \gamma \mathbf{I}) \quad (23)$$

$$\mathbf{Q} = \hat{\mathbf{S}} \hat{\mathbf{S}}^\top \quad (24)$$

and

$$\mathbf{C} = (\mathbf{V}^\top \mathbf{V})^{-1} (\mathbf{V}^\top \hat{\mathbf{Y}} \hat{\mathbf{S}}^\top + \eta \hat{\mathbf{R}}^\top \hat{\mathbf{X}}). \quad (25)$$

To solve this problem, we conduct eigenvalue decomposition on $\mathbf{P}$

$$\mathbf{P} = \mathbf{H} \mathbf{\Sigma} \mathbf{H}^{-1} \quad (26)$$

where $\mathbf{H} \in \mathbb{R}^{k \times k}$ is the eigenvector matrix, each column of which is an eigenvector. $\mathbf{\Sigma} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$, where λ_i denotes the i th eigenvalue.

Inserting (26) into (22) and then left multiplying $\mathbf{H}^{-1}$, we have

$$\mathbf{\Sigma} \mathbf{H}^{-1} \mathbf{E} + \mathbf{H}^{-1} \mathbf{E} \mathbf{Q} = \mathbf{H}^{-1} \mathbf{C}. \quad (27)$$

Let

$$\mathbf{F} \triangleq (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_k)^\top = \mathbf{H}^{-1} \mathbf{E} \quad (28)$$

$$\mathbf{C}' \triangleq (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k)^\top = \mathbf{H}^{-1} \mathbf{C} \quad (29)$$

where $\mathbf{f}_i$ and $\mathbf{c}_i$ are the i th row of $\mathbf{F}$ and $\mathbf{C}'$, respectively. We solve (27) row by row

$$\mathbf{f}_i = \mathbf{c}_i (\lambda_i \mathbf{I} + \mathbf{Q})^{-1}. \quad (30)$$

After recovering $\mathbf{F}$, we compute $\mathbf{E}$ as

$$\mathbf{E} = \mathbf{H} \mathbf{F}. \quad (31)$$

Finally, the HR-HSI can be reconstructed by $\hat{\mathbf{Z}} = \mathbf{V} \mathbf{E}$.

To summarize, Algorithm 1 presents the procedure of solving the coefficients' matrix $\mathbf{E}$.

C. Interpolation Manner and Parameter Choosing

Bilinear and bicubic manners are two common alternatives for image interpolation when constructing matrix $\mathbf{S}$. We run our algorithm on six remote sensing hyperspectral datasets in unaligned cases using these two interpolation manners. Table I lists the average SAM and PSNR values of reconstructed HSIs and the running time. It is observed that the qualities of reconstructed HSIs are quite similar. Compared with the bicubic manner, the bilinear manner is computationally more efficient as it only needs to consider four nearest neighbor pixels, and accordingly, its spatial mapping matrix $\mathbf{S}$ is sparser. Hence, we choose the bilinear manner in constructing $\mathbf{S}$.

Algorithm 1 Solving the Coefficients' Matrix $\mathbf{E}$

Input: LR multispectral matrix $\hat{\mathbf{X}}$, HR hyperspectral matrix $\hat{\mathbf{Y}}$, spectral basis matrix $\mathbf{V}$, spectral degradation function $\mathbf{R}$, affine parameter $\mathbf{A}$, PSF.

Output: HR hyperspectral matrix $\hat{\mathbf{Z}}$.

Construct $\mathbf{B}$ using PSF;

Construct $\mathbf{S}$ using $\mathbf{A}$;

Compute $\mathbf{P}$ using (23);

Compute $\mathbf{Q}$ using (24);

Compute $\mathbf{C}$ using (25);

Compute $\mathbf{P}$ using (23);

Compute $\mathbf{H}$ and $\mathbf{\Sigma}$ using (26);

Compute $\mathbf{C}' = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k)^\top$ using (29);

for $i = 1$ **to** k **do**

 | Compute $\mathbf{f}_i$ using (30);

end

Recover $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_k)^\top$;

Compute $\mathbf{E}$ using (31).

TABLE I

AVERAGE SAM AND PSNR VALUES OF THE RECONSTRUCTED HSIS USING BILINEAR AND BICUBIC INTERPOLATION MANNERS FOR IMAGE INTERPOLATION AND THE CORRESPONDING RUNNING TIME. SCALE FACTOR $b = 4$ AND 8, RESPECTIVELY

	$b = 4$		$b = 8$	
	Bilinear	Bicubic	Bilinear	Bicubic
SAM	2.30	2.31	2.68	2.69
RMSE	1.93	1.94	2.26	2.26
PSNR	43.03	43.00	41.74	41.72
Time (s)	21.89	25.31	39.91	65.04

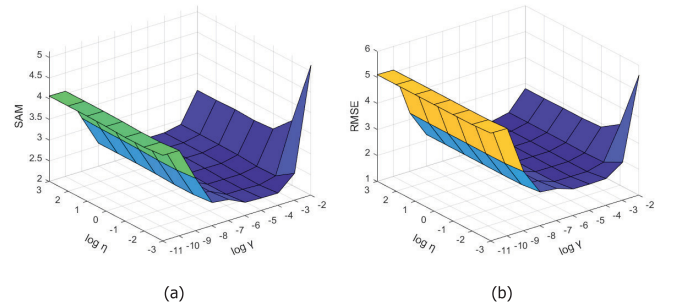


Fig. 6. Trends of (a) average SAM values and (b) RMSE values of the reconstructed HSIs on six datasets with respect to the parameters η and γ .

We determine the appropriate values of the parameters γ and η in (20) from the range $\log_{10} \eta \in \{-3, -2, -1, 0, 1, 2, 3\}$ and $\log_{10} \gamma \in \{-11, -10, -9, -8, -7, -6, -5, -4, -3, -2\}$, respectively. Fig. 6 shows the average SAM and RMSE values of all the reconstructed HSIs with respect to the parameters η and γ . We set the combination $\eta = 0.1$ and $\gamma = 10^{-5}$ that produce the lowest RMSE and SAM values.

VI. EXPERIMENTS

In this section, we first introduce the datasets we used in our experiments. Then, we introduce five quality metrics, one of which evaluates the registration degree and the others evaluate

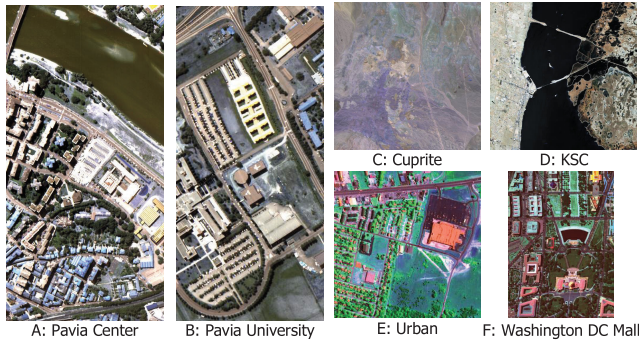


Fig. 7. Six remote sensing datasets used in our experiments, shown as pseudocolor images.

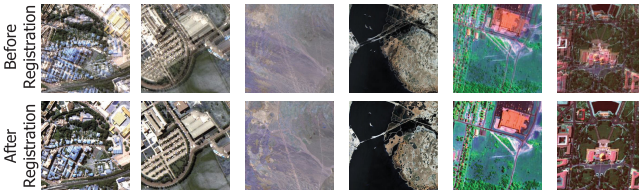


Fig. 8. Blending images before and after registration using our NED-based algorithm on six remote sensing datasets, shown as pseudo-RGBs.

the reconstructed HSIs. Finally, we present the experimental results of our proposed algorithm.

We perform three kinds of experiments. First, we compare our NED measure in the registration stage with NTG [8], SAD [33], SSD [33], and NCC [19]. Then, we evaluate the influence of interpolation error in the fusion stage with or without using our interpolation modeling. Finally, we compare our HSI fusion algorithm with the state-of-the-art ones, including HySure [5], CSTF [6], SRIF [23], uSDN [34], IARF [11], HISRM [50], UAL [29], DBHIF [28], and DHSIS [30]. Similar to our algorithm, HySure, CSTF, and SRIF use degradation models. uSDN builds an unsupervised deep network. IARF and HISRM cope with the misaligned HSI-MSI fusion problem. UAL, DBHIF, and DHSIS are all supervised deep learning algorithms.

A. Datasets

We use six remote sensing datasets (see Fig. 7) to evaluate the algorithms. Datasets A and B were both taken by optics system imaging spectrometer (ROSIS) over Pavia, with the wavelength ranging from 0.43 to 0.86 μm [35]. The scenes are Pavia Center and Pavia University, respectively.¹ We discard the noisy or water absorption channels and keep 93 bands for the experiments. We use a spatial region with 800×400 pixels in dataset A and 600×280 pixels in dataset B as HSI ground truth.

Dataset C was captured by the NASA Airborne Visible Infrared Imaging Spectrometer (AVIRIS) [36], covering the Cuprite in Las Vegas. It is the basic hyperspectral dataset for

spectrum unmixing [37]–[40]. The dataset contains 224 bands ranging from 370 to 2480 nm.² After eliminating noisy channels (1–3 and 221–224) and water absorption channels (107–113 and 153–168), 194 bands remain. We use an image region of 496×496 pixels as HSI ground truth.

Dataset D is the Kennedy Space Center (KSC) [36], [41], which was also taken by NASA AVIRIS over Florida at an altitude of about 20 km, with the wavelength ranging from 400 to 2500 nm.³ Getting rid of noisy channels and water absorption channels, 66 bands are preserved. A 496×496 spatial region is used in the experiments.

Dataset E is Urban [42], [43], one of the most widely used hyperspectral datasets in spectrum unmixing.⁴ It contains 210 bands ranging from 400 to 2500 nm with an increment of 10 nm. Removing the noisy bands and water absorption bands, we remained 162 bands. The 280×280 pixel-sized region is cut out from the image for the experiments.

Dataset F was captured by an airborne hyperspectral data flight line over the Washington DC Mall [44], with the wavelength ranging from 0.4 to 2.4 μm .⁵ We use 191 bands and cut out the image region with 376×272 pixels for the experiments.

To generate HR-MSIs, we use IKONOS-like spectral response filters [45] to map multiple bands to four bands. To generate the LR-HSIs, we blur the ground-truth HSIs spatially and downsampled them with scale factor $b = 4, 8$, and 16, respectively. The $s \times s$ Gaussian filters with the size $s = (3b)/2 - 1$ and variance $\sigma^2 = (3b)/8$ are used to blur the images.

Besides, the CAVE dataset [46] and the Harvard dataset [47] are also used to compare our algorithm with the supervised deep learning algorithms, i.e., UAL [29], DBHIF [28], and DHSIS [30]. The CAVE dataset contains 32 indoor scenes, and the Harvard dataset has 50 outdoor scenes.

B. Quantitative Metrics

Five metrics are used to evaluate our experimental results. The first one is registration error [48], formulated as

$$e = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N \|\tilde{p}(x, y, \mathbf{A}_{\text{opt}}) - \tilde{p}(x, y, \mathbf{A}_{\text{gt}})\|_F^2 \quad (32)$$

where MN is the total number of pixels, x and y denote the coordinate of pixels, $\mathbf{A}_{\text{opt}}$ and $\mathbf{A}_{\text{gt}}$ denote the estimated and ground-truth affine parameters, respectively, and $\tilde{p}$ represents the pixel position after image warping.

The other four metrics evaluate the quality of the reconstructed HS images. The spectral angle mapper (SAM) [49]

² Available at http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

³ Available at http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

⁴ Available at <https://sites.google.com/site/feiyunzhuhomepage/datasets-ground-truths>

⁵ Available at <https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>

¹ Available at http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

computes the average angle difference between two images as

$$\text{SAM}(\mathbf{Z}, \hat{\mathbf{Z}}) = \frac{1}{MN} \sum_{j=1}^{MN} \arccos \left(\frac{\mathbf{Z}_{:,j}^\top \hat{\mathbf{Z}}_{:,j}}{\|\mathbf{Z}_{:,j}\|_2 \|\hat{\mathbf{Z}}_{:,j}\|_2} \right) \quad (33)$$

where $\mathbf{Z} \in \mathbb{R}^{M \times N \times L}$ and $\hat{\mathbf{Z}} \in \mathbb{R}^{M \times N \times L}$ denote the estimated and ground-truth HR-HSIs, respectively. $\mathbf{Z}_{:,j}$ and $\hat{\mathbf{Z}}_{:,j}$ present the spectral vectors of the j th pixel of the images. The smaller the SAM, the better the spectrum reconstruction.

The third metric, the root mean square error (RMSE) [49], is defined as

$$\text{RMSE}(\mathbf{Z}, \hat{\mathbf{Z}}) = \sqrt{\frac{\|\mathbf{Z} - \hat{\mathbf{Z}}\|_F^2}{M \times N \times L}}. \quad (34)$$

The smaller the RMSE, the better the reconstruction quality.

The fourth metric is the peak signal-to-noise ratio (PSNR) [24], defined as

$$\text{PSNR}(\mathbf{Z}, \hat{\mathbf{Z}}) = \frac{1}{L} \sum_{i=1}^L 10 \lg \left(\frac{M \times N \times \text{MAX}_i^2}{\|\mathbf{z}_i - \hat{\mathbf{z}}_i\|_F^2} \right) \quad (35)$$

where $\hat{\mathbf{z}}_i$ and $\mathbf{z}_i$ are the ground-truth and estimated images in the i th band, respectively, and MAX_i denotes the maximum intensity in the i th band of the ground truth. The larger the PSNR, the better the spatial quality.

The last metric is ERGAS [49], formulated as

$$\text{ERGAS}(\mathbf{Z}, \hat{\mathbf{Z}}) = \frac{100}{b} \sqrt{\frac{1}{L} \sum_{i=1}^L \left(\frac{\text{RMSE}_i}{\mu_{\hat{\mathbf{z}}_i}} \right)^2} \quad (36)$$

where b is the scale factor of the spatial resolution from the HR-MSI to LR-HSI, RMSE_i is the RMSE value in the i th band, and $\mu_{\hat{\mathbf{z}}_i}$ is the mean of the ground-truth image in the i th band. The smaller the ERGAS, the better the reconstructed image quality.

C. Evaluation of Image Registration

In the following, we evaluate the registration process of our algorithm. We first apply three affine transforms to HR-MSIs and downsample HSIs with $4\times$ and $8\times$ to create LR-HSIs. Then, the unaligned HR-MSI and LR-HSI pairs are put into our image registration algorithm. We impose small, middle, and large deformations on the images using the affine parameters

$$\begin{aligned} \mathbf{A}_1 &= (0.99, 0.05, -5, 0.04, 0.97, -5)^\top \\ \mathbf{A}_2 &= (1.02, 0.03, -10, -0.02, 0.98, -10)^\top \\ \mathbf{A}_3 &= (0.98, 0.03, -15, -0.03, 1.01, -15)^\top. \end{aligned} \quad (37)$$

Fig. 8 illustrates the blending images of six datasets before and after registration using our NED-based algorithm, shown as pseudo-RGBs. It is obvious that ghosts are significantly removed after registration. Fig. 9 shows the image registration results on the Pavia University dataset when the LR images are $8\times$ downsampled. It is observed that our NED performs the best among the five measures with either visual or quantitative evaluation.

Table II lists the registration error on the six remote sensing datasets with scale factor $b = 4$ and 8 . Our NED measure

TABLE II
REGISTRATION ERROR OF THE LR-HSI AND HR-MSI ALIGNMENT USING DIFFERENT ALGORITHMS ON SIX REMOTE SENSING DATASETS WITH SCALE FACTOR $b = 4$ AND 8

	$b = 4$			$b = 8$		
	$\mathbf{A}_1$	$\mathbf{A}_2$	$\mathbf{A}_3$	$\mathbf{A}_1$	$\mathbf{A}_2$	$\mathbf{A}_3$
Dataset A: Pavia Center						
SSD [33]	5.46	3.60	5.22	5.38	3.58	5.15
SAD [33]	0.10	3.70	15.60	0.34	1.83	13.02
NCC [19]	0.33	0.35	0.51	1.27	1.30	2.11
NTG [8]	0.09	0.11	0.15	0.36	0.19	0.45
NED	0.10	0.14	0.13	0.44	0.38	0.45
Dataset B: Pavia University						
SSD [33]	4.16	2.77	3.32	4.23	3.22	3.82
SAD [33]	0.37	0.34	0.91	1.29	2.09	2.28
NCC [19]	1.07	1.15	1.26	3.99	4.07	4.27
NTG [8]	0.19	0.33	0.46	0.68	0.64	2.97
NED	0.12	0.13	0.15	0.36	0.45	0.58
Dataset C: Cuprite						
SSD [33]	7.39	5.09	6.09	9.98	10.16	7.85
SAD [33]	0.46	0.52	0.49	3.52	4.03	2.69
NCC [19]	37.13	32.94	65.24	11.48	7.88	42.79
NTG [8]	0.81	0.77	0.90	4.21	4.22	4.30
NED	0.37	0.47	0.47	4.25	4.84	2.89
Dataset D: KSC						
SSD [33]	5.46	3.60	5.22	5.38	3.58	5.15
SAD [33]	0.29	0.32	0.38	1.62	1.68	2.16
NCC [19]	0.83	0.87	1.05	3.81	3.76	4.12
NTG [8]	0.16	0.18	0.17	0.77	0.57	0.72
NED	0.19	0.17	0.24	0.47	0.45	0.67
Dataset E: Urban						
SSD [33]	3.38	2.31	2.91	4.29	3.77	3.92
SAD [33]	0.73	0.69	0.65	4.03	4.10	4.47
NCC [19]	1.30	10.57	16.93	4.45	7.91	18.55
NTG [8]	0.33	0.52	0.50	4.20	4.31	4.43
NED	0.26	0.30	0.41	0.64	2.09	2.25
Dataset F: Washington DC Mall						
SSD [33]	3.36	2.25	2.87	4.14	3.24	4.04
SAD [33]	0.63	0.74	0.72	2.64	3.64	3.85
NCC [19]	1.23	1.35	1.44	4.29	4.35	4.52
NTG [8]	0.44	0.44	0.46	1.00	3.02	3.15
NED	0.20	0.21	0.22	0.28	0.94	0.91

always produces the lowest registration error on the datasets B, E, and F in all cases. It performs the best or close to best on the rest datasets in most cases. Hence, we consider that our NED measure is suitable for the HSI-MSI fusion problem.

D. Influence of Interpolation Error

We note that there exists an error between the intensity of real pixels and that computed by interpolation. Here, we analyze the influence of the interpolation error on the reconstructing HSIs without or with interpolation modeling. In the case of without interpolation modeling, we warp the LR-HSI using the ground-truth affine parameter and conduct image fusion. In the case of interpolation modeling, we incorporate the ground-truth affine parameter in the fusion process, as discussed in Section V. Table III compares the average

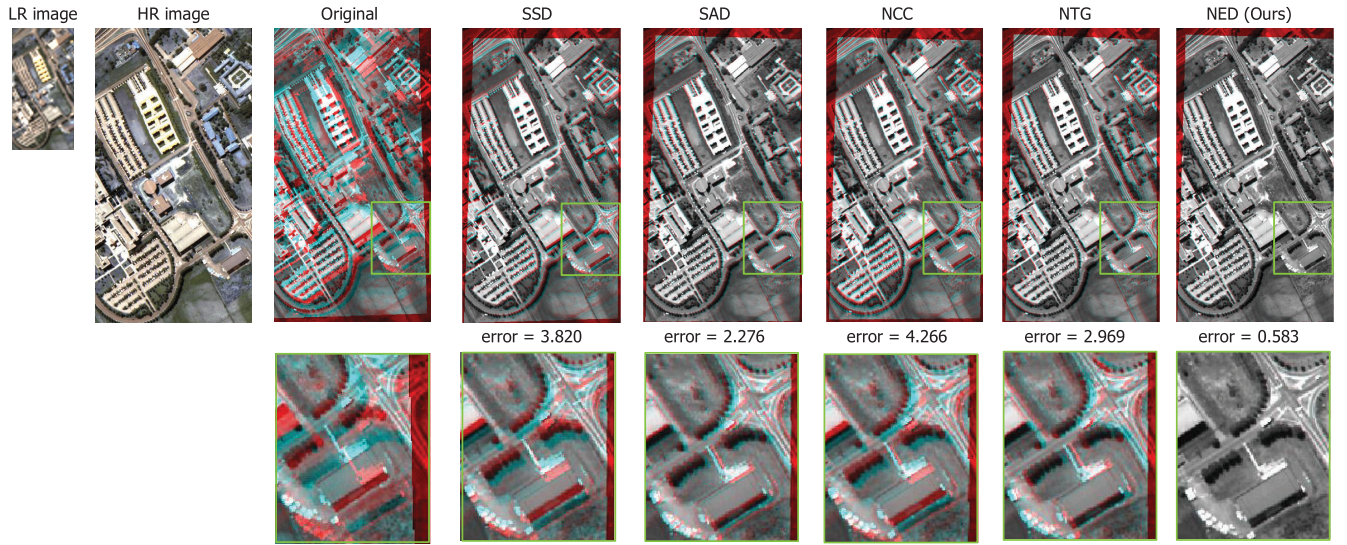


Fig. 9. Image registration results on the Pavia University dataset using SSD [33], SAD [33], NCC [19], NTG [8], and our NED. The LR image and the HR image are unaligned. The spatial size relationship is $8\times$. Blending images of original state and results are shown in pseudocolors. The second row shows the detail regions in red boxes.

TABLE III

AVERAGE SAM, RMSE, PSNR, AND ERGAS OF THE RECONSTRUCTED HSI USING METHODS WITH INTERPOLATION MODELING AND WITHOUT MODELING WITH SCALE FACTOR $b = 4$ AND 8

	$b = 4$				$b = 8$			
	SAM	RMSE	PSNR	ERGAS	SAM	RMSE	PSNR	ERGAS
Dataset A: Pavia Center								
w/o modeling	3.01	1.61	46.07	1.35	4.30	2.47	44.05	0.99
w/ modeling	2.58	1.37	46.83	1.18	3.84	2.27	44.54	0.91
Dataset B: Pavia University								
w/o modeling	2.44	2.20	43.77	1.19	2.62	2.35	43.06	0.67
w/ modeling	2.11	1.97	44.38	1.08	2.23	2.11	43.64	0.61
Dataset C: Cuprite								
w/o modeling	0.73	1.77	44.25	0.44	0.95	2.41	42.75	0.29
w/ modeling	0.68	1.61	44.72	0.41	0.89	2.24	43.14	0.27
Dataset D: KSC								
w/o modeling	2.48	1.49	43.51	2.17	3.04	1.78	41.87	1.33
w/ modeling	1.87	1.16	45.51	1.71	2.35	1.49	43.32	1.15
Dataset E: Urban								
w/o modeling	3.41	3.93	36.74	2.39	3.44	3.70	37.14	1.08
w/ modeling	2.86	3.19	38.32	1.92	3.04	3.25	38.13	0.94
Dataset F: Washington DC Mall								
w/o modeling	3.91	2.34	38.19	2.49	3.54	1.99	38.38	1.17
w/ modeling	3.63	2.20	38.65	2.39	3.21	1.81	39.00	1.09

results on six datasets with scale factor $b = 4$ and 8. It is obvious that, on all metrics, the algorithm with interpolation modeling clearly improves the accuracy of HSI reconstruction compared with the traditional algorithms directly using image warping in image fusion.

In the HSI-MSI registration process, there is also the registration error in addition to the interpolation error. Hence, we then use our registration stage to estimate the affine parameter and compare the algorithms with and without interpolation modeling. Fig. 10(a) shows the PSNR values of

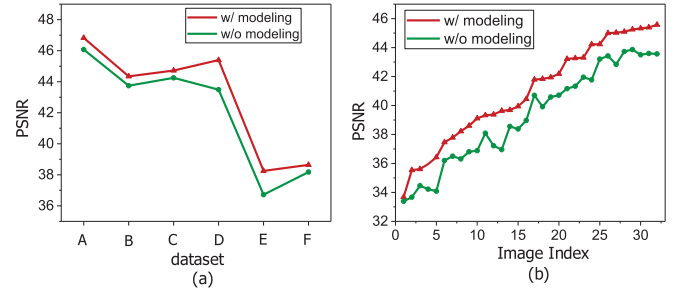


Fig. 10. PSNR values of the reconstructed HSIs on six remote sensing datasets and the CAVE dataset in the cases of with and without interpolation treatments. (a) PSNR values on six datasets. (b) PSNR values of the images in the CAVE dataset. The image index is sorted with respect to the PSNR values with interpolation modeling.

reconstructed HSIs on six remote sensing datasets with LR images $4\times$ downsampled. It is observed that the algorithm with interpolation modeling produces higher PSNR values. To further validate this, we test on a larger dataset, the CAVE dataset, by fusing unaligned RGB and MS image pairs. Fig. 10(b) shows that the algorithm with our interpolation modeling always performs better than that without modeling.

E. Evaluation of Image Fusion

In the following, we evaluate the image fusion process of our algorithm. We compare our algorithm with the state of the arts, including HySure [5], CSTF [6], SRIF [23], uSDN [34], IARF [11], HISRM [50], UAL [29], DBHIF [28], and DHSIS [30], among which the last three algorithms are supervised deep learning ones that need model training. IARF and HISRM deal with the unregistered HSI fusion problem, while the other competitors deal with aligned image pairs. Therefore, for HySure, CSTF, SRIF, uSDN, UAL, DBHIF, and

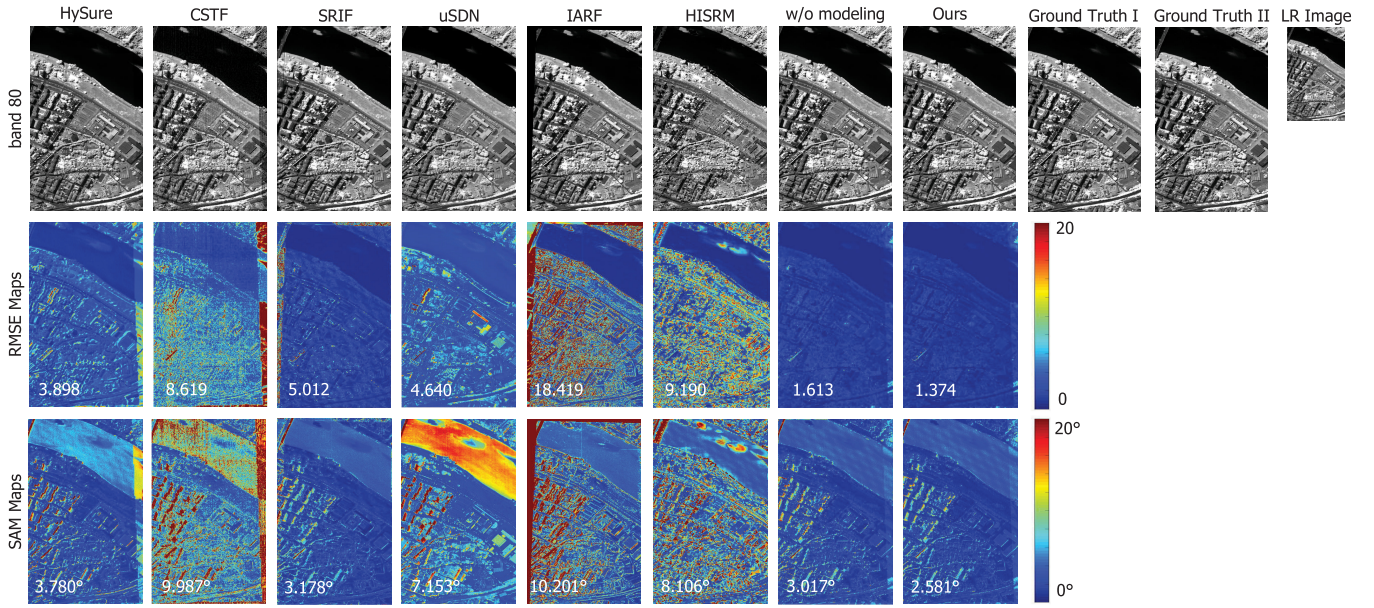


Fig. 11. Reconstructed images on the Pavia Center dataset. Ground Truth I is aligned with the results of HySure, CSTF, uSDN, HISRM, w/o modeling, and ours. Ground Truth II is aligned with the results of SRIF and IARF.

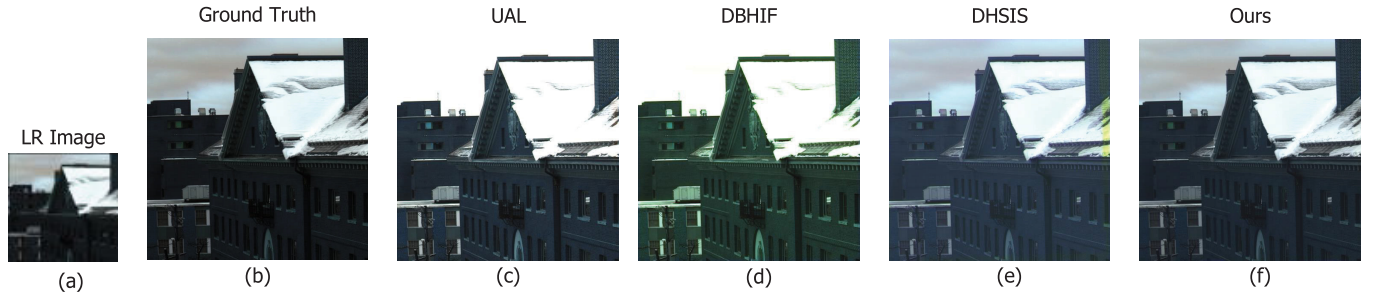


Fig. 12. Reconstructed images on the roof scene in cross-dataset evaluation. (a) LR image. (b) Ground truth. (c) UAL. (d) DBHIF. (e) DHSIS. (f) Ours.

TABLE IV

SAM AND RMSE VALUES OF THE RECONSTRUCTED HSIs PRODUCED BY HYSURE, CSTF, SRIF, USDN, UAL, DBHIF, AND DHSIS WHEN USING DIFFERENT WARPING MANNERS. THE SCALE FACTOR $b = 8$

	Warping HR-MSI		Warping LR-HSI		Validation dataset
	SAM	RMSE	SAM	RMSE	
HySure [5]	6.14	12.17	3.74	4.18	Pavia Center
CSTF [6]	22.01	39.77	11.89	10.09	Pavia Center
SRIF [23]	3.99	7.37	5.33	9.94	Pavia Center
uSDN [34]	10.79	14.37	8.37	6.67	Pavia Center
UAL [29]	6.77	9.35	5.43	2.00	CAVE
DBHIF [28]	6.31	9.22	6.94	2.99	CAVE
DHSIS [30]	7.40	8.93	8.67	4.16	CAVE

TABLE V

BASIC INFORMATION OF HYSURE, CSTF, SRIF, USDN, IARF, HISRM, UAL, DBHIF, AND DHSIS, INCLUDING WARPING OBJECT, DATASETS, AND WHETHER NEED TRAINING

Algorithm	Warping	Datasets	Training
HySure [5]	LR-HSI	Remote sensing	No
CSTF [6]	LR-HSI	Remote sensing	No
SRIF [23]	HR-MSI	Remote sensing	No
uSDN [34]	LR-HSI	Remote sensing	No
IARF [11]	—	Remote sensing	No
HISRM [50]	—	Remote sensing	No
UAL [29]	LR-HSI	CAVE and Harvard	Yes
DBHIF [28]	LR-HSI	CAVE and Harvard	Yes
DHSIS [30]	LR-HSI	CAVE and Harvard	Yes

DHSIS, we need to align the HSI and MSI before performing HSI-MSI fusion.

We first evaluate the impacts of two warping manners (warping LR-HSI or warping HR-MSI) on image fusion since both manners cause interpolation errors. Table IV lists the SAM and RMSE values of the reconstructed HSIs when using different

warping manners on the Pavia Center or CAVE datasets with $8\times$ resolution improvement. It is observed that, for HySure, CSTF, uSDN, and UAL, warping LR-HSI produces smaller SAM and RMSE values. SRIF performs better when warping HR-MSI. For DBHIF and DHSIS, warping LR-HSI produces

TABLE VI
SAM, RMSE, PSNR, AND ERGAS VALUES OF THE RECONSTRUCTED HSIs USING HySURE, CSTF, SRIF, ARAD, USDN, IARF, HISRM,
AND OUR ALGORITHM ON SIX REMOTE SENSING DATASETS WITH SCALE FACTOR $b = 4, 8$, AND 16

	$b = 4$				$b = 8$				$b = 16$			
	SAM	RMSE	PSNR	ERGAS	SAM	RMSE	PSNR	ERGAS	SAM	RMSE	PSNR	ERGAS
Dataset A: Pavia Center												
HySure [5]	3.78	3.90	36.34	3.63	3.74	4.18	35.72	1.87	3.29	5.18	34.18	1.09
CSTF [6]	9.99	8.62	30.42	7.29	11.89	10.09	29.02	4.31	10.72	9.38	29.09	2.06
SRIF [23]	3.18	5.01	34.33	3.97	3.99	7.37	30.99	2.92	8.48	19.52	22.69	3.75
uSDN [34]	7.15	4.64	34.53	4.44	8.37	6.67	32.38	2.75	5.22	11.61	26.59	1.48
IARF [11]	10.20	18.42	23.51	14.18	14.35	24.75	20.95	9.38	18.13	29.16	19.51	5.58
HISRM [50]	8.11	9.19	28.99	8.03	5.42	4.90	34.49	2.18	3.84	3.15	39.61	0.70
Ours	2.58	1.37	46.82	1.18	3.86	2.28	44.52	0.91	4.50	2.54	43.46	0.50
Dataset B: Pavia University												
HySure [5]	2.75	3.63	37.05	2.17	2.61	4.35	35.23	1.33	2.42	4.84	34.60	0.70
CSTF [6]	7.86	10.66	28.55	6.30	8.49	11.28	27.52	3.44	7.61	10.58	27.90	1.59
SRIF [23]	2.18	5.43	33.82	2.99	2.71	7.49	31.01	2.07	11.19	24.90	20.42	3.53
uSDN [34]	3.69	5.48	33.64	3.18	6.66	10.05	28.04	3.05	8.49	14.60	24.60	2.27
IARF [11]	9.29	19.77	22.90	11.09	14.13	26.85	20.11	7.69	18.96	32.12	18.44	4.68
HISRM [50]	7.61	10.58	27.46	6.51	4.52	6.27	32.23	1.88	2.29	4.21	36.21	0.59
Ours	2.12	1.97	44.35	1.09	2.26	2.12	43.62	0.61	2.65	2.41	43.25	0.33
Dataset C: Cuprite												
HySure [5]	1.26	6.77	30.32	1.41	1.25	9.75	27.04	0.98	1.27	3.53	36.86	0.19
CSTF [6]	4.00	20.75	21.27	4.29	4.40	22.79	20.40	2.36	1.39	4.27	35.01	0.23
SRIF [23]	0.66	9.51	27.03	1.90	0.95	14.01	23.66	1.40	1.81	26.35	18.16	1.32
uSDN [34]	1.57	4.02	34.92	0.93	1.82	5.03	32.70	0.56	1.71	6.38	30.34	0.34
IARF [11]	2.34	12.40	25.48	2.46	4.08	27.34	19.71	2.66	4.71	30.55	18.62	1.49
HISRM [50]	2.89	12.47	24.66	2.64	2.48	8.04	29.88	0.90	2.01	5.62	35.67	0.32
Ours	0.68	1.62	44.72	0.41	0.90	2.27	43.06	0.28	1.21	3.22	41.25	0.19
Dataset D: KSC												
HySure [5]	2.90	2.53	38.61	2.58	4.70	3.37	36.34	1.66	5.71	4.32	33.99	1.09
CSTF [6]	7.75	6.12	32.01	5.58	8.32	6.54	31.30	3.02	9.40	7.34	30.45	1.70
SRIF [23]	1.70	3.75	35.14	3.70	2.52	5.81	31.08	2.96	5.69	11.98	24.90	3.01
uSDN [34]	5.79	2.98	36.73	1.66	8.41	3.75	34.52	2.96	7.41	5.08	32.24	1.36
IARF [11]	5.99	10.91	26.14	11.21	8.72	13.86	23.77	7.21	11.93	17.14	22.12	4.39
HISRM [50]	7.31	7.89	28.52	8.14	5.24	4.14	34.06	2.44	3.59	2.64	38.75	1.01
Ours	1.90	1.18	45.39	1.73	2.52	1.57	42.88	1.21	3.66	2.22	39.96	0.87
Dataset E: Urban												
HySure [5]	3.43	6.31	32.11	3.06	3.33	7.25	30.92	1.68	3.06	7.07	31.00	0.82
CSTF [6]	10.28	16.37	24.64	7.40	12.51	18.93	23.16	4.57	8.58	14.67	25.60	1.77
SRIF [23]	2.98	7.23	30.28	4.20	8.49	18.62	22.17	5.33	12.40	26.45	18.80	3.93
uSDN [34]	2.92	7.07	31.07	1.79	9.39	11.76	26.72	2.77	10.55	21.16	21.33	2.50
IARF [11]	10.82	20.73	21.15	12.41	14.76	26.20	18.98	7.95	16.79	29.59	17.86	4.50
HISRM [50]	10.07	16.35	23.55	8.09	6.65	9.01	29.16	2.31	5.87	8.04	30.74	0.97
Ours	2.88	3.22	38.25	1.94	3.19	3.41	37.76	0.99	3.53	3.86	36.88	0.55
Dataset F: Washington DC Mall												
HySure [5]	4.12	3.00	33.77	3.21	5.26	4.22	30.93	2.20	4.96	4.20	31.07	1.07
CSTF [6]	9.67	6.74	24.91	42.67	12.00	8.26	23.05	25.41	12.14	8.41	23.01	9.74
SRIF [23]	2.72	2.95	32.26	3.99	7.62	7.42	24.30	4.96	9.85	9.45	22.20	3.14
uSDN [34]	5.83	3.00	28.94	8.59	10.62	6.17	23.92	6.75	13.10	9.07	20.59	5.72
IARF [11]	11.08	9.19	21.00	47.13	15.68	12.02	18.70	28.23	21.23	14.62	17.02	21.04
HISRM [50]	15.28	10.87	22.89	10.95	12.43	7.36	27.55	3.45	8.94	5.06	31.69	1.27
Ours	3.63	2.21	38.64	2.39	3.38	1.91	38.58	1.14	3.75	2.04	38.90	0.55

lower RMSE values but higher SAM values compared with the HR-MSI warping manner. Our investigation indicates that the algorithms perform similarly on other datasets, and hence, for a fair comparison, we use different warping manners for individual algorithms in the following experiments. For clarity,

we list the configurations of these algorithms in Table V, including warping manner, datasets, and the necessity of model training.

We first compare our algorithm with the competitors that do not need training, including HySure [5], CSTF [6], SRIF [23],

TABLE VII
AVERAGE SAM, RMSE, PSNR, AND ERGAS VALUES OF THE RECONSTRUCTED HSIs FOR THE SAME-DATASET VALIDATION AND THE CROSS-DATASET VALIDATION WITH SCALE FACTOR $b = 4, 8$, AND 16

	$b = 4$				$b = 8$				$b = 16$			
	SAM	RMSE	PSNR	ERGAS	SAM	RMSE	PSNR	ERGAS	SAM	RMSE	PSNR	ERGAS
Same-dataset validation: CAVE to CAVE												
UAL [29]	5.93	2.26	42.41	2.40	6.35	2.63	41.77	1.33	6.02	2.70	42.56	0.65
DBHIF [28]	5.70	2.62	43.98	2.47	6.94	2.99	42.17	1.44	7.43	2.87	41.53	0.72
DHSIS [30]	9.26	4.47	38.19	4.31	8.67	4.16	38.67	4.08	12.20	6.22	34.66	1.54
Ours	7.24	2.72	41.23	3.03	7.79	3.76	39.34	1.90	9.78	5.07	36.93	1.28
Cross-dataset validation: CAVE to Harvard												
UAL [29]	2.63	2.01	41.76	1.75	3.12	2.89	39.99	1.20	3.57	3.19	38.43	0.71
DBHIF [28]	17.16	11.34	28.49	8.07	15.16	9.90	29.11	3.54	6.87	6.61	31.36	1.31
DHSIS [30]	17.59	7.75	31.81	6.65	10.54	4.98	34.29	2.13	5.23	3.90	36.79	0.82
Ours	2.39	1.91	43.75	1.64	3.04	2.29	42.30	0.97	3.80	2.78	41.14	0.60

uSDN [34], IARF [11], and HISRM [50]. For our algorithm, we first perform image registration and then incorporate the interpolation modeling in image fusion. For IARF and HISRM, we directly feed the HSI and MSI into the algorithm. For others, we first align the misaligned images using the corresponding warping manner and then perform HSI-MSI fusion. Table VI lists the SAM, RMSE, PSNR, and ERGAS values of the reconstructed HSIs on six remote sensing datasets. It is observed our algorithm outperforms the competitors in most cases. Fig. 11 shows the reconstructed images of band 80 on the Pavia Center dataset using different algorithms, as well as their RMSE and SAM maps. It is observed that the RMSE and SAM values of our algorithm are much lower than those of competitors.

We further compare our algorithm with the supervised deep learning algorithms, UAL [29], DBHIF [28], and DHSIS [30]. We randomly select 12 scenes in the CAVE dataset for model training. We use the rest 20 scenes in the CAVE dataset for the same-dataset evaluation and 50 scenes of the Harvard dataset for cross-dataset evaluation. It is noted that these three deep learning algorithms all deal with aligned image pairs. We train the models under the circumstance that the HSIs and MSIs are aligned pixel by pixel. We register the HSI and MSI first and input them into the models. Table VII lists the average SAM, RMSE, PSNR, and ERGAS values. It is observed that, in the same-dataset validation, our algorithm performs worse than UAL and DBHIF but better than DHSIS. In cross-dataset validation, our algorithm achieves the best on all metrics, except for the SAM value when the scale factor $b = 16$. Fig. 12 shows the reconstruction results of the roof scene in cross-dataset validation. It is observed that the images reconstructed by UAL and DBHIF are too bright, and the one by DHSIS exhibits color shifts. This indicates that the deep learning algorithms do not generalize well on different datasets, while our algorithm performs satisfactorily in this case.

F. Running Time

In the experiment, HySure [5], CSTF [6], SRIF [23], IARF [11], HISRM [50], and our algorithm are implemented using MATLAB R2018b on a personal computer with 4.10-GHz

TABLE VIII
RUNNING TIMES OF HySURE, CSTF, SRIF, USDN, IARF, HISRM, AND OUR ALGORITHM WHEN RECONSTRUCTING AN HR-HSI WITH 31 SPECTRAL BANDS, 496×496 SPATIAL RESOLUTION, AND $4 \times$ RESOLUTION IMPROVEMENT

Algorithms	Time (s)	Processor	Algorithms	Time (s)	Processor
HySure [5]	106.26	CPU	HISRM [50]	119.38	CPU
CSTF [6]	81.10	CPU	UAL [29]	72.50	GPU
SRIF [23]	22.45	CPU	DBHIF [28]	14.43	GPU
uSDN [34]	725.66	CPU	DHSIS [30]	4.95	GPU
IARF [11]	152.73	CPU	Ours	17.80	CPU

CPU (Intel Core i5-10600K) and 16-GB RAM, and uSDN [34] is implemented using Python on the same computer. Deep learning methods, UAL [29], DBHIF [28], and DHSIS [30], are implemented using Python on GeForce RTX 2070 (GPU). Table VIII lists the running times of these algorithms when reconstructing an HR-HSI with 31 spectral bands, 496×496 spatial resolution, and $4 \times$ resolution improvement. We can see that our algorithm costs the least time compared to competitors running on CPU. Compared to the deep learning ones, our algorithm runs faster than UAL but slower than DBHIF and DHSIS. In general, our algorithm is computationally efficient.

VII. CONCLUSION

In this work, we have proposed an HSI fusion framework for unaligned LR-HSI and HR-MSI pairs. NED is introduced as a new measure to cope with the clear-blur image registration problem. Different from the previous work that directly warps the LR-HSIs, we employ the interpolation modeling treatment by incorporating the affine parameter into the spatial degradation model. Experimental results validate that the interpolation modeling treatment can effectively improve the quality of HSI reconstruction. The proposed image fusion algorithm is useful to real-world applications in which HSI and MSI cannot be assumed to be perfectly aligned.

APPENDIX

A. Affine Transformation

The affine transformation is formulated as

$$\mathcal{T}(x, y, \mathbf{A}) = \begin{pmatrix} a_1x + a_2y + a_3 \\ a_4x + a_5y + a_6 \end{pmatrix} = \mathbf{XA} \quad (38)$$

where

$$\mathbf{A} = (a_1, a_2, \dots, a_6)^\top \in \mathbb{R}^{6 \times 1} \quad (39)$$

$$\mathbf{X} = \begin{pmatrix} x & y & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 \end{pmatrix} \in \mathbb{R}^{2 \times 6}. \quad (40)$$

B. Approximation of L1-Norm

We use the Charbonnier penalty function [51] to approximate the L1-norm

$$\Psi(\mathbf{x}) = \sum_{i=1}^N \sqrt{x_i^2 + \delta^2}, \quad \mathbf{x} \in \mathbb{R}^N \quad (41)$$

where δ is fixed small constant. In this work, we set $\delta = 0.01$. The first derivative of the Charbonnier function is

$$\Psi'(\mathbf{x}) = \left(x_1 / \sqrt{x_1^2 + \delta^2}, \dots, x_N / \sqrt{x_N^2 + \delta^2} \right)^\top. \quad (42)$$

C. Derivation of Gradient

By approximating the L1-norm with the Charbonnier function, NED can be rewritten as

$$J(\mathbf{A}) = \frac{\Psi(\mathbf{E}_b - \mathbf{E}_c(\mathcal{T}(x, y, \mathbf{A})))}{\Psi(\mathbf{E}_b) + \Psi(\mathbf{E}_c(\mathcal{T}(x, y, \mathbf{A})))}. \quad (43)$$

Let

$$P(\mathbf{A}) = \Psi(\mathbf{E}_b - \mathbf{E}_c(\mathcal{T}(x, y, \mathbf{A}))) \quad (44)$$

and

$$Q(\mathbf{A}) = \Psi(\mathbf{E}_b) + \Psi(\mathbf{E}_c(\mathcal{T}(x, y, \mathbf{A}))). \quad (45)$$

We have $J(\mathbf{A}) = (P(\mathbf{A})) / (Q(\mathbf{A}))$ and

$$\nabla_{\mathbf{A}} J = \frac{(\nabla_{\mathbf{A}} P) - J(\nabla_{\mathbf{A}} Q)}{Q}. \quad (46)$$

Then, we compute the first derivations of P and Q with respect to $\mathbf{A}$

$$\nabla_{\mathbf{A}} P = (\nabla_{\mathbf{A}} \mathbf{E}_c) \Psi'(\mathbf{E}_b - \mathbf{E}_c) \quad (47)$$

$$\nabla_{\mathbf{A}} Q = (\nabla_{\mathbf{A}} \mathbf{E}_c) \Psi'(\mathbf{E}_c). \quad (48)$$

Combining (46)–(48), we have

$$\nabla_{\mathbf{A}} J(\mathbf{A}) = \frac{(\nabla_{\mathbf{A}} \mathbf{E}_c) [\Psi'(\mathbf{E}_b - \mathbf{E}_c) - J \Psi'(\mathbf{E}_c)]}{Q}. \quad (49)$$

Let

$$\mathbf{w} = \Psi'(\mathbf{E}_b - \mathbf{E}_c) - J \Psi'(\mathbf{E}_c) \in \mathbb{R}^N. \quad (50)$$

We have

$$\nabla_{\mathbf{A}} J(\mathbf{A}) = \frac{(\nabla_{\mathbf{A}} \mathbf{E}_c) \mathbf{w}}{Q}. \quad (51)$$

Since we only concern the direction of $\nabla_{\mathbf{A}} J(\mathbf{A})$, (51) can be simplified to

$$\nabla_{\mathbf{A}} J(\mathbf{A}) \propto (\nabla_{\mathbf{A}} \mathbf{E}_c) \mathbf{w}. \quad (52)$$

Here, we denote $\nabla_{\mathbf{A}} \mathbf{E}_c = (\nabla_{\mathbf{A}} e_{c1}, \nabla_{\mathbf{A}} e_{c2}, \dots, \nabla_{\mathbf{A}} e_{cN})$, each column of which is computed as

$$\nabla_{\mathbf{A}} e_c = \nabla_{\mathbf{A}} e_c(\mathcal{T}(x, y, \mathbf{A})) = (\nabla_{\mathbf{A}} \mathcal{T}(x, y, \mathbf{A})) (\nabla e_c). \quad (53)$$

According to (38), we have

$$\nabla_{\mathbf{A}} \mathcal{T}(x, y, \mathbf{A}) = \nabla_{\mathbf{A}} \mathbf{X} \mathbf{A} = \mathbf{X}^\top. \quad (54)$$

In (53), ∇e_c denotes the gradients at pixel e_c in the x - and y -directions, formulated as

$$\nabla e_c = \begin{pmatrix} \nabla_x e_c \\ \nabla_y e_c \end{pmatrix}. \quad (55)$$

Combining (52)–(55), we have

$$\nabla_{\mathbf{A}} J(\mathbf{A}) \propto \sum_{i=1}^N w_i \mathbf{X}_i^\top \nabla e_{ci} \quad (56)$$

which completes the deduction of $\nabla_{\mathbf{A}} J$.

REFERENCES

- [1] W. Song, S. Li, L. Fang, and T. Lu, "Hyperspectral image classification with deep feature fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3173–3184, Jun. 2018.
- [2] M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral face recognition with spatio-spectral information fusion and PLS regression," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1127–1137, Mar. 2015.
- [3] G. A. Shaw and H.-H. K. Burke, "Spectral imaging for remote sensing," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 3–28, 2003.
- [4] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
- [5] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image super-resolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, Jun. 2015.
- [6] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4118–4130, Aug. 2018.
- [7] E. Wycoff, T.-H. Chan, K. Jia, W.-K. Ma, and Y. Ma, "A non-negative sparse promoting algorithm for high resolution hyperspectral imaging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 1409–1413.
- [8] S.-J. Chen, H.-L. Shen, C. Li, and J. H. Xin, "Normalized total gradient: A new measure for multispectral image registration," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1297–1310, Mar. 2018.
- [9] V. Kučera, "The matrix equation $AX + XB = C$," *SIAM J. Appl. Math.*, vol. 26, no. 1, pp. 15–25, 1974.
- [10] J. Nie, L. Zhang, W. Wei, C. Ding, and Y. Zhang, "Unsupervised deep hyperspectral super-resolution with unregistered images," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2020, pp. 1–6.
- [11] Y. Zhou, A. Rangarajan, and P. D. Gader, "An integrated approach to registration and fusion of hyperspectral and multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3020–3033, May 2020.
- [12] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [13] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer-Verlag, 2006, pp. 404–417.
- [14] S. Guiming and S. Jidong, "Multi-scale harris corner detection algorithm based on canny edge-detection," in *Proc. IEEE Int. Conf. Comput. Commun. Eng. Technol. (CCET)*, Aug. 2018, pp. 305–309.
- [15] B. Lowe, "Recognising panoramas," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, p. 1218.
- [16] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [17] X. Fan, H. Rhody, and E. Saber, "A spatial-feature-enhanced MMI algorithm for multimodal airborne image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2580–2589, Jun. 2010.
- [18] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.
- [19] M. Irani and P. Anandan, "Robust multi-sensor image alignment," in *Proc. 6th Int. Conf. Comput. Vis.*, Jan. 1998, pp. 959–966.

- [20] A. Myronenko and X. Song, "Intensity-based image registration by minimizing residual complexity," *IEEE Trans. Med. Imag.*, vol. 29, no. 11, pp. 1882–1891, Nov. 2010.
- [21] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, Aug. 1996.
- [22] B. Ayhan, M. Dao, C. Kwan, H.-M. Chen, J. F. Bell, and R. Kidd, "A novel utilization of image registration techniques to process mastcam images in mars rover with applications to image fusion, pixel clustering, and anomaly detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4553–4564, Oct. 2017.
- [23] Z.-W. Pan and H.-L. Shen, "Multispectral image super-resolution via RGB image fusion and radiometric calibration," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1783–1797, Apr. 2019.
- [24] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.
- [25] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5344–5353.
- [26] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 19–34.
- [27] J. Wu, J. Aeschbacher, and R. Timofte, "In defense of shallow learned spectral reconstruction from RGB images," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 471–479.
- [28] W. Wang, W. Zeng, Y. Huang, X. Ding, and J. Paisley, "Deep blind hyperspectral image fusion," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4150–4159.
- [29] L. Zhang, J. Nie, W. Wei, Y. Zhang, S. Liao, and L. Shao, "Unsupervised adaptation learning for hyperspectral imagery super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3073–3082.
- [30] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5345–5355, Nov. 2018.
- [31] L. T. Maloney, "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 3, no. 10, pp. 1673–1683, 1986.
- [32] J. Mairal, F. Bach, J. Ponce, G. Sapiro, R. Jenatton, and G. Obozinski. (2014). *Spams: A Sparse Modeling Software*, V2. 3. [Online]. Available: <http://spams-devel.gforge.inria.fr/downloads.html>
- [33] R. Szeliski, "Image alignment and stitching: A tutorial," *Found. Trends Comput. Graph. Vis.*, vol. 2, no. 1, pp. 1–104, 2007.
- [34] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse Dirichlet-net for hyperspectral image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2511–2520.
- [35] H. J. Kramer, *Observation of the Earth and Its Environment: Survey of Missions and Sensors*. Berlin, Germany: Springer-Verlag, 2002.
- [36] G. Vane, R. O. Green, T. G. Chrien, H. T. Enmark, E. G. Hansen, and W. M. Porter, "The airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 44, nos. 2–3, pp. 127–143, 1993.
- [37] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.
- [38] N. Wang, B. Du, and L. Zhang, "An endmember dissimilarity constrained non-negative matrix factorization method for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 554–569, Apr. 2013.
- [39] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $L_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4282–4297, Nov. 2011.
- [40] F. Zhu, Y. Wang, B. Fan, S. Xiang, G. Meng, and C. Pan, "Spectral unmixing via data-guided sparsity," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5412–5427, Dec. 2014.
- [41] A. L. Neuenschwander, M. M. Crawford, and M. J. Provan, "Mapping of coastal wetlands via hyperspectral aviris data," in *Proc. IEEE Int. Geosci. Remote Sensing. Symp.*, vol. 1, Jul. 1998, pp. 189–191.
- [42] Q. Xiao, S. Austin, E. McPherson, and P. Peper, "Characterization of the structure and species composition of urban trees using high resolution aviris data," in *Proc. AVIRIS Workshop*, 1999, pp. 1–10.
- [43] F. Dell'Acqua, P. Gamba, A. Ferrari, J. A. Palmason, J. A. Benediktsson, and K. Arnason, "Exploiting spectral and spatial information in hyperspectral urban data with high resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 322–326, Oct. 2004.
- [44] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, vol. 29. Hoboken, NJ, USA: Wiley, 2005.
- [45] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109–4121, Nov. 2015.
- [46] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. CVPR*, Jun. 2011, pp. 193–200.
- [47] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.
- [48] S.-Y. Cao, H.-L. Shen, S.-J. Chen, and C. Li, "Boosting structure consistency for multispectral and multimodal image registration," *IEEE Trans. Image Process.*, vol. 29, pp. 5147–5162, 2020.
- [49] L. Loncan *et al.*, "Hyperspectral pansharpening: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.
- [50] Y. Lin, Y. Zheng, Y. Fu, and H. Huang, "Hyperspectral image super-resolution under misaligned hybrid camera system," *IET Image Process.*, vol. 12, no. 10, pp. 1824–1831, Oct. 2018.
- [51] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2432–2439.
- [52] M. Ding, X. Fu, T.-Z. Huang, J. Wang, and X.-L. Zhao, "Hyperspectral superresolution via interpretable block-term tensor modeling," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 641–656, 2021, doi: [10.1109/JSTSP.2020.3045965](https://doi.org/10.1109/JSTSP.2020.3045965).



Jiacheng Ying received the B.Eng. degree in information engineering from Zhejiang University, Hangzhou, China, in 2020, where he is pursuing the Ph.D. degree with the College of Information Science and Electronic Engineering.

His research interests are multispectral imaging and image processing.



Hui-Liang Shen received the B.Eng. and Ph.D. degrees in electronic engineering from Zhejiang University, Hangzhou, China, in 1996 and 2002, respectively.

He was a Research Associate and a Research Fellow with The Hong Kong Polytechnic University, Hong Kong, from 2001 to 2005. He is a Full Professor with the College of Information Science and Electronic Engineering, Zhejiang University. His research interests include multispectral imaging, image processing, computer vision, and machine learning.



Si-Yuan Cao received the B.E. degree from Tianjin University, Tianjin, China, in 2016. He is pursuing the Ph.D. degree with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China.

His research interests are multispectral/multimodal image alignment and image processing.