

# An integrated approach to registration and fusion of hyperspectral and multispectral images

Yuan Zhou, *Member, IEEE*, Anand Rangarajan, *Member, IEEE*, and Paul D. Gader, *Fellow, IEEE*

**Abstract**—Combining a hyperspectral (HS) image and a multispectral (MS) image—an example of image fusion—can result in a spatially and spectrally high-resolution image. Despite the plethora of fusion algorithms in remote sensing, a necessary prerequisite, namely registration, is mostly ignored. This limits their application to well registered images from the same source. In this paper, we propose and validate an integrated registration and fusion approach. The registration algorithm minimizes a least-squares (LSQ) objective function with the point spread function (PSF) incorporated together with a nonrigid freeform transformation applied to the HS image and a rigid transformation applied to the MS image. It can handle images with significant scale differences and spatial distortion. The fusion algorithm takes the full, high resolution HS image as an unknown in the objective function. Assuming that the pixels lie on a low-dimensional manifold invariant to local linear transformations from spectral degradation, the fusion optimization problem leads to a closed form solution. The method was validated on the Pavia University, Salton Sea, and the Mississippi Gulfport datasets. When the proposed registration algorithm is compared to its rigid variant and two mutual information-based methods, it has the best accuracy for both the nonrigid simulated dataset and the real dataset, with an average error less than 0.15 pixels for nonrigid distortion of maximum 1 HS pixel. When the fusion algorithm is compared with current state-of-the-art algorithms, it has the best performance on images with registration errors as well as on simulations which do not consider registration effects.

**Index Terms**—hyperspectral image analysis, nonrigid registration, image fusion, point spread function

## I. INTRODUCTION

**H**YPERSPECTRAL (HS) images have important applications in agriculture, forestry, geoscience, and astronomy, as its sensor can capture the reflectance at hundreds of wavelengths ranging from visible to shortwave infrared, thus leading to analysis of the spectra of materials on the surface. However, due to limited amount of incident energy, it also suffers from a low spatial resolution, such that multiple objects are recorded within a single pixel. For example, the Hyperion system onboard the Earth Observing 1 (EO-1) satellite launched in 2000 acquires data covering wavelengths 0.4  $\mu\text{m}$  to 2.5  $\mu\text{m}$  with a 30 m spatial resolution [1], the AVIRIS sensor

covers the same spectral range with a 18 m spatial resolution [2]. On the other hand, multispectral (MS) (e.g. color) images are recorded with only a few bands, but they can have a much better spatial resolution. For example, the commercial satellite QuickBird launched in 2001 can collect panchromatic imagery at 61 centimeter spatial resolution and MS imagery at 2.5 m.

These two types of images can be combined—via image fusion—to produce a spatially *and* spectrally high-resolution image [3], [4]. Most research on this topic focuses only on the fusion process, in which it is assumed that the images are already registered. This limits their application to only well registered datasets that are collected at the same time and from the same aircraft or satellite. In this work, we will investigate registration and fusion together, without the strict assumption about collection time and devices. Also, we are especially interested in the case where the spatial resolution difference is large and the registration accuracy requirement is high.

Common fusion applications include land cover mapping, mineral mapping, identifying plant species, object detection, etc. In our special case, we can also build ground truth abundance maps for unmixing HS images. A widely used approach to obtain ground truth for material proportions is to refer to the region of a high resolution color image corresponding to the HS counterpart, where the two images could come from different sources. This approach has been used by multiple endmember spectral mixture analysis (MESMA) and the following work [5], [6], [7]. It is worth noting that this type of registration relies on coordinate information stored in the image data, which may lead to shifts in the matched regions due to inaccuracy of pixel-level coordinates [8]. An accurate registration can mitigate this effect and the combined high-resolution pixels can be classified to obtain the ground truth for material distribution. Another application is spatially calibrating distorted airborne HS images resulting from instability of the collection process. In this case, a spatially accurate high-resolution MS image can guide the calibration of the HS image.

### A. Related work on registration

Despite a plethora of work on fusion, there is little mention of the registration process. Previous work focused on registration are usually aimed at overcoming small scale differences as in [9], or lack fine scale accuracy [10]. For example, Yokoya et al. separate pixels with prominent features, e.g. lakes, rivers from those without [10]. For the first kind, control points with

Y. Zhou is with the Department of Radiology and Biomedical Imaging, Yale University, New Haven, CT, USA. A. Rangarajan and P. D. Gader are with the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, USA. E-mail: zhouyuanzxcv@gmail.com, {anand.pgader}@cise.ufl.edu. The work is partially supported by NSF IIS 1743050 to A.R. The data and code of the proposed approach are available online (<https://github.com/zhoyuanzxcv/Hyperspectral>).

correspondence are found through correlation. For the second kind, affine transformations are used for interpolation. It can be difficult to achieve subpixel accuracy in the case of large spatial scale differences when using control points.

In the image registration community, there is little previous work aimed at significant scale differences despite the fact that remote sensing is a major application area of image registration [11]. Registration methods can be mainly categorized as intensity-based or feature-based. Intensity-based methods calculate a metric based on the intensities of the images, e.g. least squares [12] or mutual information (MI) [13], [14]. Sometimes the intensities are transformed to the frequency domain, in which case a direct solution can be obtained by phase correlation [15], [16], but this only applies to simple transformations. Feature-based approaches are more common in remote sensing as the images are usually very large and therefore registering a number of feature points is more efficient than comparing the intensities at all the pixels. These methods usually select some control points through low-level computer vision techniques, such as scale-invariant Fourier transforms [17] or the Harris corner detector [18], and approximate the nonrigid transformation through thin-plate splines [19] or Gaussian radial basis functions [20]. For example, in [21], a constraint from local linear embedding on the feature points is used in the objective function for registering various airborne images. Bentoutou et al. select control points in the reference image by edge detection, find the corresponding points in the test image by template matching, and use thin-plate splines to approximate the warping for nonrigid registration of SPOT satellite images [22]. In [23], the Harris corner detector is used to select feature points along with an MI objective for registering airborne infrared images. In all these applications, the remote sensing images are of similar spatial scale—a limitation we seek to overcome in the present work.

### B. Related work on fusion

Assuming that the images are already registered, various techniques including component substitution, Bayesian methods, and matrix factorization, are proposed for image fusion [3], [4]. Component substitution transforms the spatially upsampled spectral data into components in another space, substitutes a related component with the high-resolution image, and transforms them back to the original spectral space. One example is the guided filter PCA (GFPCA) algorithm, which used a guided filter [24] in the PCA domain and won the 2014 IEEE data fusion contest [25]. It upsamples the HS image using cubic interpolation, and projects the data to a few principal components using PCA. Then in local windows, these components are substituted by linear combinations of the bands of the high-resolution image.

Previous work tends to assume that the high-resolution HS image follows the linear mixing model (LMM), which means that the spectrum of each pixel is a linear combination of some material spectra (*endmembers*) with their sub-pixel fractions as coefficients (*abundances*) [26]. Hence they are interested in estimating the endmembers and abundances, which can be

used to reconstruct the full resolution image [27], [28]. For example, coupled nonnegative matrix factorization (CNMF) alternately updates the endmembers and abundances using multiplicative update rules [28]. Bayesian methods give a least-squares objective function with fitting errors weighted by noise covariance matrices [29], [30], [31], [32]. Several priors are added into this formulation for regularization. For example, a Gaussian prior is used in [31] where the abundances are assumed to follow a Gaussian distribution whose mean is the upsampled HS image. In [32] the prior forces the abundances to be a sparse linear combination of elements of an over-complete dictionary. Since the Bayesian formulation is only useful insofar as it introduces a noise covariance matrix into the least-squares term, we can directly use this term with some constraints. For example, HySure minimizes a similar data fidelity term under a spatial smoothness constraint on the abundances [33].

Despite the popularity of the LMM in fusion, nonlinearity and endmember variability have been well studied to better represent the real data in unmixing research [34], [35]. When endmember variability is considered, each pixel may be formed by a different endmember set, which breaks the low rank assumption of the LMM. To solve this issue, there is research that uses the LMM in a local manner. Veganzones et al. partition the image into overlapping local regions with the LMM used only locally [36]. The final result is obtained by combining locally reconstructed images. Recently, there is emerging interest in directly reconstructing the high-resolution HS image without the LMM assumption. Especially, the pixels are treated as lying on a low-dimensional manifold with regularization constraints derived from the MS image [37], [38].

### C. Our contribution

Registration and fusion of HS and MS images are usually studied separately, i.e. the validation of fusion is achieved without regard to registration effects, while the varying registration error is not viewed in regard to fusion error. A common way of generating a dataset for evaluation is through spatial and spectral degradation of an existing HS dataset. This facilitates evaluation of the fusion process. However, it isolates the evaluation to this ideal situation, while in practice the performance of fusion depends largely on registration. Furthermore, if we view these two topics separately, they have their own characteristics. For fusion, the widely used LMM assumption may not be sufficient to accurately represent the image as the presence of endmember variability attests. For registration, significant scale difference is seldom studied and spatial distortion exists in airborne HS images. Fig. 1 uses an example to show the challenges for registration, where the HS image is obtained from the AVIRIS portal and the MS (color) image is obtained from Google Earth. Fig. 1(a) shows the huge scale difference such that a point spread function (PSF) may be needed in the registration. Fig. 1(b) shows neighboring pixels having similar spectra, which is probably caused by spatial calibration error and hence nonrigid transformation may be needed for the HS image.

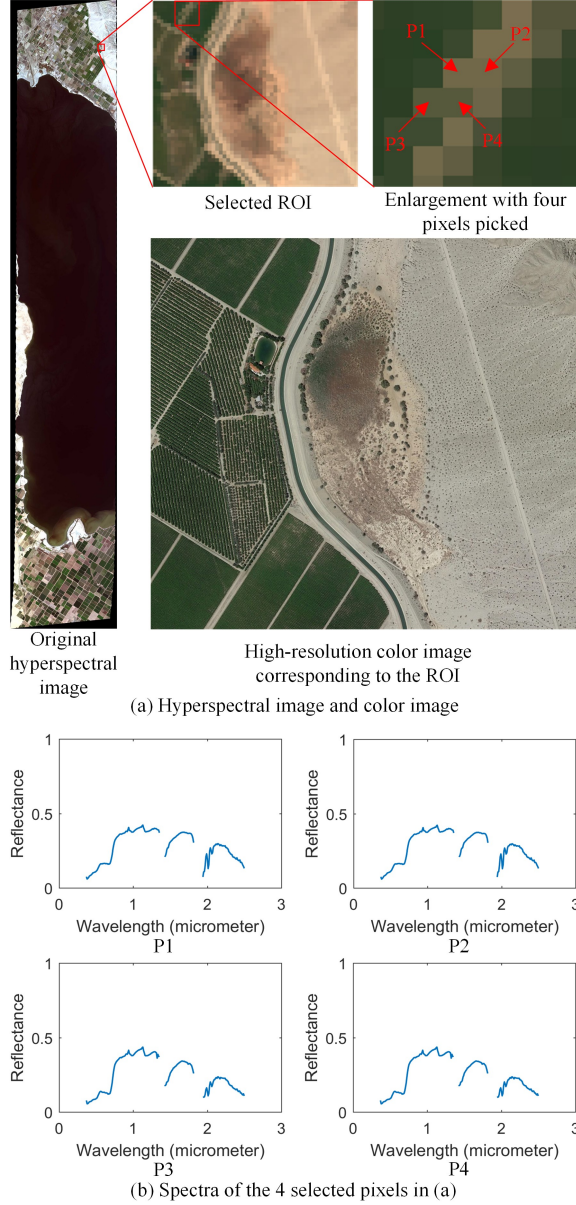


Figure 1. Challenges faced in registering an HS image and an MS image. In (a) the ROI of the Salton Sea dataset has 56 by 51 pixels with 224 bands (shown as a color image by extracting the bands close to 650 nm, 540 nm and 470 nm) while the MS image has 738 by 674 pixels with 3 bands. The scale difference is about 10. In (b), the manually picked pixels P1 and P2 (P3 and P4) have the same spectra. To register the MS image to the HS image, one way is that the road should lie exactly between P1 and P2, and between P3 and P4. However, considering the environment pixels, a more proper explanation is that P1 and P2 (P3 and P4) are spectra corresponding to the same location.

Considering these problems, the contributions of this paper are threefold. First, we propose a least-squares (LSQ) objective function including the PSF for registration. We apply a rigid transformation to the MS image and a nonrigid transformation to the HS image, and estimate them simultaneously. The model can handle significant scale difference and spatial distortion with high accuracy. Second, we propose a fusion objective function directly involving the high-resolution HS image. It assumes that the reconstructed pixels lie on a low-dimensional manifold where every point can be reconstructed by its neigh-

bors in the same way as the pixels from the MS image. The resulting algorithm turns out to be easy to implement, fast to execute and also capable of keeping spectral details. Third, by comparing several registration and fusion algorithms, we validate them in both the previous way (evaluate registration and fusion performance separately) and our integrated way (evaluate fusion performance in presence of registration error). A prior work focusing on registration was published in [39].

## II. PROBLEM FORMULATION

We will first make the following physical assumptions. Let  $\mathcal{D} \subset \mathbb{R}^2$  be the image domain,  $\mathbf{x} = [x, y]^T \in \mathcal{D}$ ,  $I : \mathcal{D} \rightarrow \mathbb{R}_+^B : I(\mathbf{x}) = [I_1(\mathbf{x}), I_2(\mathbf{x}), \dots, I_B(\mathbf{x})]^T$  be the HS image with  $B$  bands and  $I_k : \mathcal{D} \rightarrow \mathbb{R}_+$  be the image at the  $k$ th band such that

$$I_k(\mathbf{x}) = \int_{\mathbb{R}^2} g(\mathbf{y} - \mathbf{S}\mathbf{x}) r_k(\mathbf{y}) d\mathbf{y} + n_k(\mathbf{x}), \quad k = 1, \dots, B \quad (1)$$

where  $\mathbf{S} = \text{diag}(s_1, s_2) \in \mathbb{R}_+^{2 \times 2}$  is the scaling matrix,  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$  is the PSF that is assumed to be positive and normalized, i.e.

$$g(\mathbf{x}) \geq 0, \forall \mathbf{x}, \int_{\mathbb{R}^2} g(\mathbf{x}) d\mathbf{x} = 1. \quad (2)$$

The PSF usually takes the form of a Gaussian function or a constant function over a circular region around the origin [40].  $r_k : \mathbb{R}^2 \rightarrow \mathbb{R}_+$  denotes the fine scale reflectance at the  $k$ th wavelength.  $n_k(\mathbf{x})$  is the noise.

Suppose the MS image has  $b$  bands, and denoted by  $I' : \mathcal{D}' \rightarrow \mathbb{R}_+^b : I'(\mathbf{x}) = [I'_1(\mathbf{x}), \dots, I'_b(\mathbf{x})]^T$ . Considering that the HS image has a spectral resolution with much narrower bands than the MS image (e.g. 10 nm for AVIRIS and 5 nm for AVIRIS-NG compared to 60 - 270 nm from IKONOS [41] and Landsat Thematic Mapper [30]) and a MS band covered spectral range can also be covered by multiple HS bands, a MS band can be assumed to be a linear combination of some selected HS bands,

$$I'_l(\mathbf{x}) = h_{0l} + \sum_{i=1}^d h_{il} r_{k_i}(\mathbf{x}) + n'_l(\mathbf{x}), \quad l = 1, \dots, b \quad (3)$$

where  $k_1, k_2, \dots, k_d$  are the selected indices of the  $B$  wavelengths,  $h_{0l}, h_{1l}, \dots, h_{dl}$  are the coefficients of the spectral response function (SRF), and  $n'_l(\mathbf{x})$  is again the noise function. A constant shift  $h_{0l}$  is added because we also consider modeling images from different sources. An example shape of the SRF for the IKONOS satellite sensor can be seen in Fig. 1 of [41], where each MS band covers 100 nm and the panchromatic band covers 600 nm.

Assuming the images are perfectly registered, combining (1) and (3) while ignoring the two noise functions gives

$$\begin{aligned} h_{0l} + \sum_{i=1}^d h_{il} I_{k_i}(\mathbf{x}) &= h_{0l} + \sum_{i=1}^d h_{il} \int_{\mathbb{R}^2} g(\mathbf{y} - \mathbf{S}\mathbf{x}) r_{k_i}(\mathbf{y}) d\mathbf{y} \\ &= \int_{\mathbb{R}^2} g(\mathbf{y} - \mathbf{S}\mathbf{x}) I'_l(\mathbf{y}) d\mathbf{y}, \end{aligned} \quad (4)$$

where the change of summation and integration uses the property in (2). In the context of image registration, we usually

have a reference (fixed) image and a test (moving) image whose coordinates are transformed. In our case, first note that to introduce the PSF, at least translation should be added to the coordinates of the MS image, since we want to combine a correct set of high-resolution pixels into a low-resolution pixel. So we introduce a rigid transformation to the MS image. For the nonrigid transformation, however, we can not add it to the rigid transformation. This is because the high-resolution image has better spatial accuracy while a nonrigid transformation will distort it. Hence, we separate the transformation into two parts, and apply the nonrigid transformation to the HS image. Let  $\mathcal{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be the transformation on the HS image,  $\mathcal{T}' : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be the transformation on the MS image, and

$$\mathcal{T}(\mathbf{x}) = \mathbf{x} + \mathbf{v}(\mathbf{x}), \mathcal{T}'(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{t} \quad (5)$$

where  $\mathbf{v}(\mathbf{x}) = [u(\mathbf{x}), v(\mathbf{x})]^T$  is a nonrigid translation field on the coordinates,  $\mathbf{A}$  contains only rotation since scaling is incorporated in  $\mathbf{S}$  (we ignore shearing here),  $\mathbf{t} = [t_1, t_2]^T \in \mathbb{R}^2$  is the translation vector. Then, given unregistered  $I$  and  $I'$ , the relation (4) becomes

$$h_{0l} + \sum_{i=1}^d h_{il} I_{k_i}(\mathcal{T}(\mathbf{x})) = \int_{\mathbb{R}^2} g(\mathbf{y} - \mathbf{S}\mathbf{x}) I'_l(\mathcal{T}'(\mathbf{y})) d\mathbf{y} \quad (6)$$

for  $l = 1, \dots, b$  and our problem is to find  $\mathcal{T}$  and  $\mathcal{T}'$ .

The major difference of this formulation from traditional registration is the separation of transformations to the two images. Another difference is the introduction of the SRF and the PSF. For the rigid transformation, the scaling is separated from the rotation/translation in order to introduce the PSF. If we ignore the scale  $\mathbf{S}$ , the right hand side is actually convolution. Hence it actually moves the MS image and then performs low-pass filtering and downsampling.

Once we have the two images registered, we can retrieve the original reflectance  $r_k$  from (1) and (3), which is the fusion process. This can be achieved by solving the following functional

$$\begin{aligned} \mathcal{E}(\{r_k\}) = & \sum_{k=1}^B \int \left( I_k(\mathcal{T}(\mathbf{x})) - \int_{\mathbb{R}^2} g(\mathbf{y} - \mathbf{S}\mathbf{x}) r_k(\mathbf{y}) d\mathbf{y} \right)^2 d\mathbf{x} \\ & + \sum_{l=1}^b \int \left( h_{0l} + \sum_{i=1}^d h_{il} r_{k_i}(\mathbf{x}) - I'_l(\mathcal{T}'(\mathbf{x})) \right)^2 d\mathbf{x}. \end{aligned} \quad (7)$$

Note that even if the PSF and the SRF are given, this is still an underdetermined problem. Hence we will use some regularization to estimate the high-resolution  $r_k$ . An overview of the proposed integrated approach is shown in Fig. 2, with details given by the following two sections.

### III. REGISTRATION

#### A. Rigid registration in fine-scale

We will first consider rigid registration in fine-scale and then extend it to the nonrigid case. Let  $I'' : \mathcal{D} \rightarrow \mathbb{R}^b : I''(\mathbf{x}) = [I''_1(\mathbf{x}), \dots, I''_b(\mathbf{x})]^T$  be the transformed, downgraded MS image, i.e.

$$I''_l(\mathbf{x}) = \int_{\mathbb{R}^2} g(\mathbf{y} - \mathbf{S}\mathbf{x}) I'_l(\mathcal{T}'(\mathbf{y})) d\mathbf{y}, l = 1, \dots, b \quad (8)$$

where  $\mathcal{T}'$  is defined in (5). Assume the PSF has the form of a Gaussian,

$$g(x, y) \propto H\left(\rho - \sqrt{x^2 + y^2}\right) e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

where  $\sigma$  determines the shape and  $\rho$  is the radius controlling the range of influence with the Heaviside function  $H$  ( $\rho$  can be obtained from the instantaneous field of view (IFOV) and the flight height). We can minimize the squared  $L_2$  norm of the difference function,

$$\mathcal{E}(\mathbf{S}, \mathbf{A}, \mathbf{t}, \sigma) = \int_{\mathcal{D}} \sum_l \left| h_{0l} + \sum_{i=1}^d h_{il} I_{k_i}(\mathbf{x}) - I'_l(\mathbf{x}) \right|^2 d\mathbf{x}, \quad (9)$$

with respect to  $\mathbf{S}, \mathbf{A}, \mathbf{t}, \sigma$ . We can rewrite the continuous objective function in a discrete form. Let  $\mathbf{Y} \in \mathbb{R}^{N \times B}$  be the discretized version of  $I$ ,  $\mathbf{X}' \in \mathbb{R}^{N \times b}$  be the discretized version of  $I''$ . The selection of  $d$  bands based on the indices  $\{k_i, i = 1, \dots, d\}$  can be encoded in a matrix  $\mathbf{E} \in \mathbb{R}^{B \times d}$  where for the  $i$ th column, only the  $k_i$ th row is one while the others are zero. Let  $\mathbf{H} := [h_{il}] \in \mathbb{R}^{d \times b}$ ,  $\mathbf{h}_0 := [h_{01}, \dots, h_{0b}]^T \in \mathbb{R}^b$ ,  $\tilde{\mathbf{H}} := [\mathbf{h}_0, \mathbf{H}^T]^T \in \mathbb{R}^{(d+1) \times b}$ ,  $\tilde{\mathbf{Y}} := [\mathbf{1}_N, \mathbf{Y}\mathbf{E}] \in \mathbb{R}^{N \times (d+1)}$ , we can rewrite (9) in the following discrete form

$$\mathcal{E}(\mathbf{S}, \mathbf{A}, \mathbf{t}, \sigma) = \|\mathbf{X}' - \tilde{\mathbf{Y}}\tilde{\mathbf{H}}\|_F^2. \quad (10)$$

Eq. (10) not only has the registration parameters as unknown, but also the SRF  $\tilde{\mathbf{H}}$ . We can remove this dependence by solving for it. It is an overdetermined problem to get  $\tilde{\mathbf{H}}$  from (10). A direct solution will introduce non-smooth SRFs. So we add a regularization term to enforce neighboring values to be similar, i.e.

$$\begin{aligned} \mathcal{E}(\tilde{\mathbf{H}}) = & \|\mathbf{X}' - \tilde{\mathbf{Y}}\tilde{\mathbf{H}}\|_F^2 + \frac{\lambda}{2} \sum_{i=1}^d \sum_{j=1}^d w_{ij} \|\mathbf{h}_i - \mathbf{h}_j\|^2 \\ = & \|\mathbf{X}' - \tilde{\mathbf{Y}}\tilde{\mathbf{H}}\|_F^2 + \lambda \text{Tr}(\mathbf{H}^T \mathbf{L} \mathbf{H}), \end{aligned} \quad (11)$$

where  $\mathbf{h}_i := [h_{i1}, \dots, h_{ib}]^T \in \mathbb{R}^b$  (hence  $\tilde{\mathbf{H}} = [\mathbf{h}_0, \mathbf{h}_1, \dots, \mathbf{h}_d]^T$ ).  $w_{ij} = 1$  when  $|i - j| = 1$  and  $w_{ij} = 0$  otherwise.  $\mathbf{L} \in \mathbb{R}^{d \times d}$  is the Graph Laplacian matrix constructed from  $\{w_{ij}\}$  [42].  $\lambda$  is a parameter. Taking the derivative of (11) with respect to  $\tilde{\mathbf{H}}$  and setting it to zero, we have

$$\tilde{\mathbf{H}} = \left( \tilde{\mathbf{Y}}^T \tilde{\mathbf{Y}} + \lambda \mathbf{L}' \right)^{-1} \tilde{\mathbf{Y}}^T \mathbf{X}'. \quad (12)$$

where  $\mathbf{L}' := \text{diag}(0, \mathbf{L}) \in \mathbb{R}^{(d+1) \times (d+1)}$ . Plug  $\tilde{\mathbf{H}}$  in (12) back into (10), the objective function becomes

$$\mathcal{E}(\mathbf{S}, \mathbf{A}, \mathbf{t}, \sigma) = \|\mathbf{X}' - \tilde{\mathbf{Y}} \left( \tilde{\mathbf{Y}}^T \tilde{\mathbf{Y}} + \lambda \mathbf{L}' \right)^{-1} \tilde{\mathbf{Y}}^T \mathbf{X}'\|_F^2.$$

This is the final objective function used in the optimization.

The optimization involves the parameters for the transformation and for the PSF. We can split them into two sets and use block coordinate descent, i.e. at each iteration, we alternately set

$$\mathbf{S}, \mathbf{A}, \mathbf{t} \leftarrow \arg \min_{\mathbf{S}, \mathbf{A}, \mathbf{t}} \mathcal{E}(\mathbf{S}, \mathbf{A}, \mathbf{t}, \sigma), \sigma \leftarrow \arg \min_{\sigma} \mathcal{E}(\mathbf{S}, \mathbf{A}, \mathbf{t}, \sigma).$$

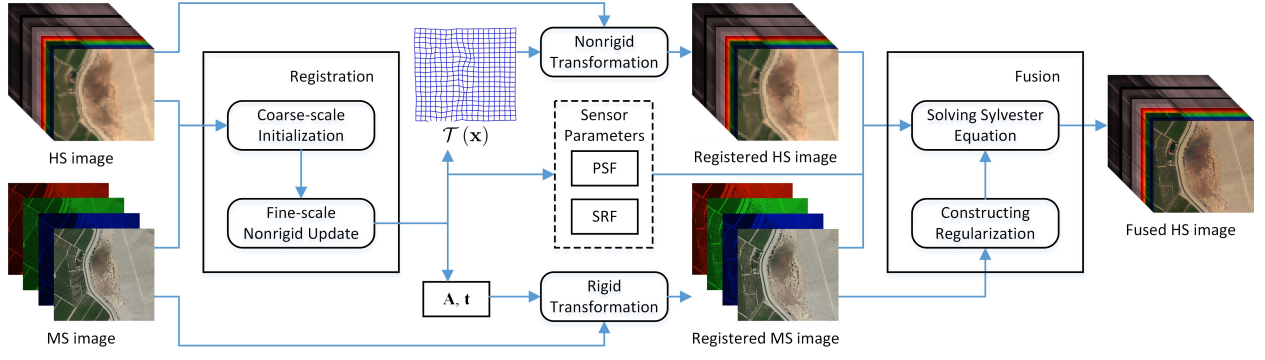


Figure 2. Overview of the proposed integrated approach for registration and fusion.

For the minimization problem with respect to  $\mathbf{S}, \mathbf{A}, \mathbf{t}$ , we can also use block coordinate descent, where each individual minimization is achieved by brutal force search. Given an initial registration (e.g. phase correlation or multi-scale mutual information on the coarse-scale images), the search is restricted to a small neighborhood and we use 5 levels of brutal force search with step sizes reduced by half at each level to achieve a finer grid. It has been shown that block coordinate descent can find the global optimum if the objective function is convex [43]. Since our objective function features a least-squares term, the capture range of convexity could be large enough for a decent initialization. Also, since the search is restricted to a discrete grid, it will quickly converge to a local minimum on this grid.

### B. Nonrigid registration using calculus of variations

We can extend the rigid registration in Section III-A to the nonrigid case by adding the optimization with respect to  $\mathbf{v}(\mathbf{x})$ . Let  $I_l''(\mathbf{x})$  be as defined in (8). With the nonrigid transformation, Eq. (9) is written as

$$\mathcal{E}(\mathbf{S}, \mathbf{A}, \mathbf{t}, \mathbf{v}, \sigma) = \int_{\mathcal{D}} \sum_l \left| h_{0l} + \sum_{i=1}^d h_{il} I_{k_i}(\mathcal{T}(\mathbf{x})) - I_l''(\mathbf{x}) \right|^2 d\mathbf{x}, \quad (13)$$

where  $\mathcal{T}(\mathbf{x}) = \mathbf{x} + \mathbf{v}(\mathbf{x})$ . Suppose  $\tilde{\mathbf{H}}$  is given in (12), Eq. (13) is a functional with respect to the translation field  $\mathbf{v}(\mathbf{x}) = [u(\mathbf{x}), v(\mathbf{x})]^T$ . For optimization with respect to  $\mathbf{v}(\mathbf{x})$ , we also want the nonrigid transformation to be a smooth function, so an additional constraint  $\alpha \int \|\nabla u(\mathbf{x})\|^2 + \|\nabla v(\mathbf{x})\|^2 d\mathbf{x}$  is added to (13). Using calculus of variations, we can obtain the necessary condition for its minimization, i.e. setting the following Euler-Lagrange equation to zero,

$$\frac{\delta \mathcal{E}}{\delta \mathbf{v}} = 2 \sum_l \left\{ \left( h_{0l} + \sum_{i=1}^d h_{il} I_{k_i}(\mathcal{T}(\mathbf{x})) - I_l''(\mathbf{x}) \right) \left( \sum_{i=1}^d h_{il} \nabla I_{k_i}(\mathcal{T}(\mathbf{x})) \right) \right\} - 2\alpha \nabla^2 \mathbf{v}(\mathbf{x}).$$

where  $\nabla$  is the gradient operator and  $\nabla^2$  is the Laplacian operator. Directly solving  $\frac{\delta \mathcal{E}}{\delta \mathbf{v}} = \mathbf{0}$  for  $\mathbf{v}(\mathbf{x})$  is a difficult problem. An easy way is to introduce another time variable to  $\mathbf{v}(\mathbf{x})$ , say  $\mathbf{v}(\mathbf{x}, t)$ , such that we have a partial differential

equation (PDE)  $\frac{\partial \mathbf{v}}{\partial t} = -\frac{\delta \mathcal{E}}{\delta \mathbf{v}}$  to solve. Using forward difference on  $\frac{\partial \mathbf{v}}{\partial t}$ , we have an update rule

$$\mathbf{v}(\mathbf{x}) \leftarrow \mathbf{v}(\mathbf{x}) - \Delta t \frac{\delta \mathcal{E}}{\delta \mathbf{v}}, \quad (14)$$

where  $\Delta t$  should be small enough to ensure a stable solution. Given an initial condition (e.g.  $\mathbf{v}(\mathbf{x}) = \mathbf{0}$ ), we can update  $\mathbf{v}(\mathbf{x})$  according to (14). Once it converges (e.g.  $\max_{\mathbf{x}} |u(\mathbf{x}) - u^{\text{old}}(\mathbf{x})| < 10^{-4}$  and  $\max_{\mathbf{x}} |v(\mathbf{x}) - v^{\text{old}}(\mathbf{x})| < 10^{-4}$  where the thresholds have been specified), we have a solution to  $\frac{\delta \mathcal{E}}{\delta \mathbf{v}} = \mathbf{0}$ .

The implementation involves the discretization of the gradient and the Laplacian operator. We use centered difference for the first order derivative and second centered difference for the Laplacian operator. Considering that the diffusion term  $\nabla^2 \mathbf{v}(\mathbf{x})$  originates from the heat equation and we want the smoothing to be isolated in the domain, the homogeneous Neumann boundary condition is adopted for  $\mathbf{v}(\mathbf{x})$ .

Combining this optimization with the rigid version, we have our final update rules for the least-squares problem (13)

$$\mathbf{S}, \mathbf{A}, \mathbf{t} \leftarrow \arg \min_{\mathbf{S}, \mathbf{A}, \mathbf{t}} \mathcal{E}(\mathbf{S}, \mathbf{A}, \mathbf{t}, \mathbf{v}, \sigma),$$

$$\mathbf{v}(\mathbf{x}) \leftarrow \arg \min_{\mathbf{v}(\mathbf{x})} \mathcal{E}(\mathbf{S}, \mathbf{A}, \mathbf{t}, \mathbf{v}, \sigma),$$

$$\sigma \leftarrow \arg \min_{\sigma} \mathcal{E}(\mathbf{S}, \mathbf{A}, \mathbf{t}, \mathbf{v}, \sigma),$$

where the second minimization problem is solved by the update rule (14) while the remaining two follow Section III-A. In practice, the update of  $\mathbf{S}, \mathbf{A}, \mathbf{t}$  can be stopped after a short period to accelerate convergence.

## IV. FUSION

### A. The underdetermined problem

We will discretize (7) to obtain the actual fusion problem. Consider Eq. (3). Let  $\mathbf{X} \in \mathbb{R}^{N' \times b}$  be the discretized version of  $I' \circ \mathcal{T}'$  following Section III,  $\mathbf{R} \in \mathbb{R}^{N' \times B}$  be the discretized version of  $\{r_k : k = 1, \dots, B\}$  (i.e. the high-resolution reflectance to be estimated). Using  $\mathbf{E}, \mathbf{H}, \mathbf{h}_0$  defined in Section III-A, Eq. (3) can be discretized as  $\mathbf{X} \approx \mathbf{1}_{N'} \mathbf{h}_0^T + \mathbf{R} \mathbf{E} \mathbf{H}$ , where the noise term is ignored. Let  $\tilde{\mathbf{X}} := \mathbf{X} - \mathbf{1}_{N'} \mathbf{h}_0^T \in \mathbb{R}^{N' \times b}$ ,  $\mathbf{F} := \mathbf{E} \mathbf{H} \in \mathbb{R}^{B \times b}$ , the relation can be written as  $\tilde{\mathbf{X}} \approx \mathbf{R} \mathbf{F}$ , where  $\tilde{\mathbf{X}}$  is an adjusted MS image,  $\mathbf{F}$  corresponds to the SRF matrix in other work.

As for the formation of HS images (1), we can also write it in a discretized form. Say  $R$  MS pixels correspond to an HS pixel and  $g$  can be discretized as  $\mathbf{g} := [g_1, \dots, g_R]^T \in \mathbb{R}^R$ . Note that  $\frac{N'}{N} = R$  may not be true as the PSF covered regions may have overlap. Let  $c_n : \{1, \dots, R\} \rightarrow \{1, \dots, N'\}$  map the MS pixel index with respect to the local window to the global index for the  $n$ th HS pixel. Once the two images are registered,  $c_n$  is unique for each HS pixel and represents the correspondence. We can encode  $c_n$  as a correspondence matrix  $\mathbf{C}_n \in \mathbb{R}^{R \times N'}$  such that only the element at the  $l$ th row,  $c_n(l)$ th column is 1 while the other elements are 0. Then Eq. (1) can be discretized as  $\mathbf{y}_n \approx (\mathbf{C}_n \mathbf{R})^T \mathbf{g}$ , where  $\mathbf{Y} := [\mathbf{y}_1, \dots, \mathbf{y}_N]^T \in \mathbb{R}^{N \times B}$  denotes the discretized HS image  $I \circ \mathcal{T}$  and the noise is ignored. Combining all the pixels, we have  $\mathbf{Y} \approx \mathbf{G}\mathbf{R}$ , where  $\mathbf{G} := [\mathbf{C}_1^T \mathbf{g}, \dots, \mathbf{C}_N^T \mathbf{g}]^T \in \mathbb{R}^{N \times N'}$  is the PSF matrix in other works.

Given the registration result, we know  $\mathbf{F}$ ,  $\mathbf{G}$  and  $\tilde{\mathbf{X}}$ ,  $\mathbf{Y}$ . The objective of fusion is to retrieve  $\mathbf{R}$ , which can be written in the following optimization problem

$$\mathcal{E}(\mathbf{R}) = \gamma \|\mathbf{G}\mathbf{R} - \mathbf{Y}\|_F^2 + (1 - \gamma) \|\mathbf{R}\mathbf{F} - \tilde{\mathbf{X}}\|_F^2, \quad (15)$$

where  $\gamma \in (0, 1)$  is a balancing parameter. This problem is underdetermined considering the matrix dimensions. Instead of assuming  $\mathbf{R}$  to be decomposed into a product to two matrices, we use some constraints for regularization.

### B. The regularization term

The regularization term comes from the manifold assumption. Assuming that the original high-resolution HS pixels lie on a smooth low-dimensional manifold, and that every local patch on the manifold has the same geometric properties as the manifold of the MS pixels, we can enforce a constraint that preserves these properties in the reconstruction. Specifically, for the  $i$ th pixel in the adjusted MS image  $\tilde{\mathbf{X}}$ , say  $\mathbf{x}_i$ , we find its closest  $K$  neighbors in the  $b$ -dimensional spectral space and assume that this pixel is a linear combination of its neighbors. Another way to interpret it is that the neighbors are endmembers and the linear coefficients are abundances, which can be seen as a kind of local LMM. To reduce computational complexity, the spatial locations of these neighbors are restricted to a local circular region  $\mathcal{B}_i(\rho)$  centered at the pixel with radius  $\rho$  (pixel) (we reuse the notation  $\rho$  in Section III). Hence it is similar to local linear embedding [44] except that the graph is constructed with a spatial location constraint.

Suppose the  $K$  closest neighbors (in spectrum) in  $\mathcal{B}_i(\rho)$  have indices given by  $\mathcal{M}_i : \{1, \dots, K\} \rightarrow \{1, \dots, N'\}$  which maps its  $j$ th neighbor to the index in the whole image, the coefficients  $\alpha_i = [\alpha_{i1}, \dots, \alpha_{iK}]^T$  can be solved by minimizing

$$\mathcal{E}(\alpha_i) = \left\| \sum_{j=1}^K \mathbf{x}_{\mathcal{M}_i(j)} \alpha_{ij} - \mathbf{x}_i \right\|^2, \text{ s.t. } \alpha_{ij} \geq 0, \sum_j \alpha_{ij} = 1. \quad (16)$$

Solving this traditional unmixing problem for every MS pixel is expensive [45]. Also, considering that  $\{\mathbf{x}_{\mathcal{M}_i(j)}\}$  are similar to  $\mathbf{x}_i$ , this problem may have an identifiability issue. Hence we introduce a regularization term  $\epsilon \|\alpha_i\|^2$  and remove the

positivity constraint. Using Lagrange multipliers, an analytic solution is available to  $\mathcal{E}(\alpha_i)$ :

$$\alpha_i = \frac{(\mathbf{S}_i + \epsilon \mathbf{I})^{-1} \mathbf{1}_K}{\mathbf{1}_K^T (\mathbf{S}_i + \epsilon \mathbf{I})^{-1} \mathbf{1}_K}$$

where  $\mathbf{S}_i \in \mathbb{R}^{K \times K}$  is a matrix whose  $j$ th row,  $k$ th column element is  $(\mathbf{x}_{\mathcal{M}_i(j)} - \mathbf{x}_i)^T (\mathbf{x}_{\mathcal{M}_i(k)} - \mathbf{x}_i)$ .

Once we have the neighbor indices and coefficients, we can assume that the same graph exists for  $\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_{N'}]^T$  and the reverse of the following derivation holds

$$\mathbf{r}_i \approx \sum_j \mathbf{r}_{\mathcal{M}_i(j)} \alpha_{ij} \implies \mathbf{x}_i \approx \sum_j \mathbf{x}_{\mathcal{M}_i(j)} \alpha_{ij} \quad (\mathbf{F}^T \mathbf{r}_i \approx \mathbf{x}_i). \quad (17)$$

Hence we obtain linear relations among  $\{\mathbf{r}_i\}$ . To enforce the relations, the key is to construct a sparse matrix  $\mathbf{D} \in \mathbb{R}^{N' \times N'}$  to calculate  $\|\sum_j \mathbf{r}_{\mathcal{M}_i(j)} \alpha_{ij} - \mathbf{r}_i\|^2$ . For the  $i$ th row of  $\mathbf{D}$ , we set the  $\mathcal{M}_i(j)$ th column to be  $\alpha_{ij}$  for  $j = 1, \dots, K$ , and set the  $i$ th column to be -1. Then, the constraint can be written as  $\|\mathbf{D}\mathbf{R}\|_F^2 = \text{Tr}\{\mathbf{R}^T (\mathbf{D}^T \mathbf{D}) \mathbf{R}\}$ . Note that  $\mathbf{D}$  depends on  $\rho$ . So we add the subscript and use  $\mathbf{D}_\rho$  henceforth. When  $\rho = 1$ , the region  $\mathcal{B}_i(\rho)$  shrinks to the first-order neighborhood of the pixel (adjacent 4 pixels on the image lattice), which spatially regulates the current pixel spectrum like a smoothness constraint. Combining it with the spectral constraint with a large  $\rho$ , the regularization term becomes

$$\mathcal{E}(\mathbf{R}) = \|\mathbf{D}_{\rho_1} \mathbf{R}\|_F^2 + \|\mathbf{D}_{\rho_2} \mathbf{R}\|_F^2 = \text{Tr}(\mathbf{R}^T \mathbf{L} \mathbf{R}) \quad (18)$$

where  $\mathbf{L} = \sum_{\rho \in \{\rho_1, \rho_2\}} \mathbf{D}_\rho^T \mathbf{D}_\rho$ ,  $\rho_1 = 1$ ,  $\rho_2$  is a parameter.

### C. The algorithm

Consider the problem (15) with regularization (18). We have an optimization problem where we minimize

$$\mathcal{E}(\mathbf{R}) = \gamma \|\mathbf{G}\mathbf{R} - \mathbf{Y}\|_F^2 + (1 - \gamma) \|\mathbf{R}\mathbf{F} - \tilde{\mathbf{X}}\|_F^2 + \beta \text{Tr}\{\mathbf{R}^T \mathbf{L} \mathbf{R}\}, \quad (19)$$

where  $\beta$  is a parameter. We ignore the positivity constraint on  $\mathbf{R}$  because in practice, the regularization term along with the data fidelity term makes the solution seldom negative (we will show the sufficiency of our formulation in the experiments).

It can be verified that minimizing (19) is a convex problem. Hence, any local minimum will be its global minimum, and we can take the derivative with respect to  $\mathbf{R}$  and set it to zero to find the solution, which turns out to be *Sylvester equation*

$$\mathbf{B}\mathbf{R} + \mathbf{R}(1 - \gamma)\mathbf{F}\mathbf{F}^T = \mathbf{Z}, \quad (20)$$

where  $\mathbf{B} \in \mathbb{R}^{N' \times N'}$ ,  $\mathbf{Z} \in \mathbb{R}^{N' \times B}$  denote

$$\mathbf{B} := (\gamma \mathbf{G}^T \mathbf{G} + \beta \mathbf{L}), \mathbf{Z} := \gamma \mathbf{G}^T \mathbf{Y} + (1 - \gamma) \tilde{\mathbf{X}} \mathbf{F}^T.$$

Note that  $\mathbf{B}$  is a sparse banded positive definite matrix whose band width is determined by  $\rho_2$ , and  $\mathbf{F}\mathbf{F}^T \in \mathbb{R}^{B \times B}$  is a symmetric matrix with only  $b$  (number of bands for the MS image) nonzero eigenvalues. We can use the classical Krylov subspace method to solve (20) more efficiently [46]. Namely, we decompose  $(1 - \gamma)\mathbf{F}\mathbf{F}^T$  by eigendecomposition, say  $(1 - \gamma)\mathbf{F}\mathbf{F}^T = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$  ( $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_B)$  consists of eigenvalues), which turns (20) into  $\mathbf{B}\mathbf{R}' + \mathbf{R}'\mathbf{\Lambda} = \mathbf{Z}'$ , where  $\mathbf{R}' = \mathbf{R}\mathbf{U}$ ,  $\mathbf{Z}' = \mathbf{Z}\mathbf{U}$ . Then  $\mathbf{R}' = [\mathbf{r}'_1, \dots, \mathbf{r}'_B]$  can be obtained

---

**Algorithm 1** The proposed fusion algorithm
 

---

 Input:  $\mathbf{X}, \mathbf{Y}, \mathbf{E}, \tilde{\mathbf{H}}, \mathbf{G}, \gamma, \beta, \rho_2, K$ 

- 1)  $\gamma \leftarrow (NB(1 - \gamma)/N'b\gamma + 1)^{-1}$ ,  $\beta \leftarrow (b/B)\beta$ .
- 2)  $\mathbf{F} \leftarrow \mathbf{E}\mathbf{H}$ ,  $\tilde{\mathbf{X}} \leftarrow \mathbf{X} - \mathbf{1}_{N'}\mathbf{h}_0^T$ .
- 3) Following Section IV-B, construct  $\mathbf{L}$  from  $\tilde{\mathbf{X}}$ .
- 4)  $\mathbf{Z} \leftarrow \gamma\mathbf{G}^T\mathbf{Y} + (1 - \gamma)\tilde{\mathbf{X}}\mathbf{F}^T$ ,  $\mathbf{B} \leftarrow (\gamma\mathbf{G}^T\mathbf{G} + \beta\mathbf{L})$ .
- 5) Let  $\mathbf{U}$  consist of the eigenvectors of  $(1 - \gamma)\mathbf{F}\mathbf{F}^T$  and  $\{\lambda_i\}$  be the eigenvalues, solve  $(\mathbf{B} + \lambda_i\mathbf{I})\mathbf{r}'_i = \mathbf{z}'_i$  for  $\mathbf{r}'_i$ ,  $i = 1, \dots, B$ , where  $\mathbf{z}'_i$  is the  $i$ th column of  $\mathbf{Z}\mathbf{U}$ .
- 6)  $\mathbf{R} \leftarrow [\mathbf{r}'_1, \dots, \mathbf{r}'_B]\mathbf{U}^T$ .

 Output:  $\mathbf{R}$ 


---

by solving linear system of equations  $(\mathbf{B} + \lambda_i\mathbf{I})\mathbf{r}'_i = \mathbf{z}'_i$  where  $\mathbf{z}'_i$  is the  $i$ th column of  $\mathbf{Z}'$ . Finally  $\mathbf{R}$  is retrieved by  $\mathbf{R}'\mathbf{U}^T$ .

The algorithm is described in Algorithm 1. For the input parameters,  $\mathbf{E}, \tilde{\mathbf{H}}, \mathbf{G}$  are obtained from registration. If the PSF and the SRF are not available while the images are registered (pixel correspondence is known), we can resort to  $\|\mathbf{G}\mathbf{X} - \tilde{\mathbf{Y}}\mathbf{H}\|_F^2$  in (10) to estimate  $\mathbf{g}$  and  $\tilde{\mathbf{H}}$  where  $\mathbf{g}$  can be updated by projected gradient descent and  $\tilde{\mathbf{H}}$  can be updated by (12). The first step in the algorithm is trying to adjust the parameters based on the number of elements in different terms, e.g. suppose  $\gamma NB/(1 - \gamma)N'b = \gamma'/(1 - \gamma')$ , we have  $\gamma = (NB(1 - \gamma')/N'b\gamma' + 1)^{-1}$ . Step 2 calculates the SRF and the adjusted MS image, where  $\mathbf{h}_0$  is the transposed vector of the first row of  $\tilde{\mathbf{H}}$  and  $\mathbf{H}$  comes from the remaining rows. In step 3, we construct  $\mathbf{L}$  based on the adjusted MS image. We use  $\epsilon = 1e - 4$  in calculating  $\alpha_i$ . Step 4 calculates the matrices in the Sylvester equation (20) and the last two steps solve the equation. Since there are only  $b$  nonzero eigenvalues, for zero eigenvalues, we can use Cholesky decomposition to process all the linear equations at one time.

## V. RESULTS

We compared 4 registration algorithms in the experiments. Other than the proposed LSQ rigid registration (referred to as LSQ) in Section III-A and LSQ nonrigid registration with freeform deformation (referred to as LSQ freeform) in Section III-B, we also tried MI as a metric in our registration framework and compared against both rigid and nonrigid versions. The parameters  $\alpha, \lambda, \Delta t$  in our algorithms are fixed as  $\alpha = 0.05$ ,  $\lambda = 10^{-3}N$ ,  $\Delta t = 1$ . The two MI-based algorithms are embedded and implemented in our own framework to handle the significant scale difference. The rigid MI algorithm (referred to as MI) calculates the metric based on the red band of the MS image and the closest band to 650 nm of the HS image. The entropy was calculated by histogramming with 64 bins for the marginal distribution. The nonrigid version (referred to as MI B-spline) uses B-splines to model the deformation, with control points spaced at 8 pixels, then refined to 4 pixels, and finally 2 pixels for the first iteration. The remaining iterations use 2 pixel spaced control points from the previous iteration to further fine-tune the parameters.

The registration error can be calculated for the simulated dataset, where the rigid and nonrigid transformations are

known. Though various parameters are known, we are especially interested in the pixel match error in the HS domain, since they affect the subsequent application. According to [27], the pixel match error should be less than 0.2 pixels in the HS domain for meaningful spectral unmixing and fusion. To calculate this error, we transform the HS pixel coordinates to the corresponding points in the high-resolution image domain according to the ground truth rigid and nonrigid transformations in (6), then apply the estimated transformations to transform them back to the HS domain. These transformed coordinates are compared with the expected coordinates, and Euclidean distances are calculated to represent this error.

For fusion of HS and MS images, we compared the proposed algorithm with GFPCA [25], CNMF [28], Bayesian Naive [31], Bayesian Sparse [32] and HySure [33]. We use color images as MS images in the experiments since they are most easily available in practice. The parameters of our algorithm are  $\gamma = 0.5$ ,  $\beta = 1$ ,  $K = 3$ ,  $\rho_2 = 15$ . For the competing methods, their code was obtained from the pansharpening toolbox [3]. The parameters of GFPCA were tuned to perform best on the first dataset using the proposed registration. For all algorithms, the PSF, SRF, and the starting downsampling position were input if acceptable. We used the same four quality measures in [3], i.e. correlation coefficient (CC), spectral angle mapper (SAM), root mean squared error (RMSE) and erreur relative globale adimensionnelle de synthèse (ERGAS). Given the ground truth high-resolution HS image and the estimated one, CC calculates the correlation coefficient of each band and averages them (best value is 1). SAM calculates the angle between the spectra of these two images while RMSE calculates the size weighted  $L_2$  norm. The calculation of ERGAS involves the scale difference and RMSE weighted by average values of each band. Except CC, the lower the value, the better the result.

### A. Registration and fusion on Pavia University

The Pavia University dataset was recorded by the Reflective Optics System Imaging Spectrometer (ROSIS) during a flight over Pavia, Italy, in July 2002. The scene is around the Engineering School at the University of Pavia. The image features a spatial size of 340 by 610 pixels with a resolution of 1.3 meters/pixel. In the spectral domain, it covers wavelengths from 430 nm to 860 nm by 103 bands. We used it to generate a color image and a low resolution HS image.

The color image took the visible bands and used a Gaussian like SRF covering 120 nm for each band, centered at 650 nm, 540 nm and 470 nm (we set  $\mathbf{h}_0 = \mathbf{0}$  to conform with the assumption of the competing methods). We used the top part of the image with pixel size 340 by 500 for the color image (shown in Fig. 3). Note that the original HS image is already distorted by observing that the middle blue roof is not straight. The HS image was generated by rotating the original image by 0 - 10 degrees and scaling it by  $\mathbf{s} = (4.4, 4.5)$  with a PSF  $\sigma = 10$ ,  $\rho = 3$ . The generated HS image has a size of 50 by 80 pixels with 103 bands. We considered two cases. For the rigid case, only rotation and scaling were applied to generate the HS image. For the nonrigid case,

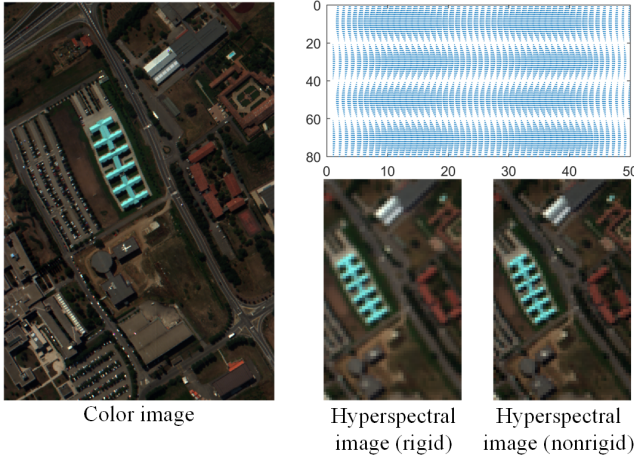


Figure 3. Simulated Pavia University dataset with the color image and HS images (rotated by 5 degrees). The top right corner shows the translation field that was applied to the nonrigidly distorted version. The translation field moves 4 parts of the image by opposite directions, with a maximal magnitude of 1 pixel.

a further nonrigid transformation with  $\mathcal{T}(\mathbf{x}) = \mathbf{x} + \mathbf{v}(\mathbf{x})$ , where  $\mathbf{v}(\mathbf{x}) = \sum_k c_k \mathcal{N}(\mathbf{x} | \mu_k, \sigma^2 \mathbf{I}_2)$ , was applied to the image from the rigid case. We used 8 Gaussian components to simulate the aircraft instability. The translation field  $\mathbf{v}(\mathbf{x})$  and the nonrigidly distorted HS image are shown in Fig. 3. Both the two images were further contaminated by a zero-mean additive Gaussian noise with standard deviation  $\sigma_n$  on all the bands.

For registration, a comprehensive quantitative comparison for all the images and noise levels is shown in Fig. 4. Among the 11 results for each noise level, we found a few large registration errors that dominated the y-axis in the box plots. To better visualize the successful cases, we removed the largest 4 errors for each method. The distribution shows that for the rigid dataset, LSQ gives the least error, while all except MI B-spline have errors below 0.1. For the nonrigid dataset, LSQ freeform performs best with errors below 0.15, followed by errors around 0.2 from MI B-spline.

For fusion, we picked the first dataset in the nonrigid registration case ( $\sigma_n = 0.0001$  in the second row of Fig. 4) such that the random noise does not interfere with our fusion quality measures. We ran fusion algorithms in two cases, with ground truth registration parameters and with estimated registration parameters (from LSQ freeform). The fusion results are shown in Table I, where the mean of multiple images is calculated. Comparing the two main categories, the difference is noticeable even if the registration error is as small as 0.1 pixels. For both registrations, the proposed fusion algorithm has the best result with respect to all the measures.

To test the fusion performance with respect to different registration errors, we generated images with increasing distortion magnitudes (following the same pattern in Fig. 3). LSQ freeform was applied to obtain increasing registration errors and fusion algorithms were run on the registration results. Fig 5 shows the trend with respect to the 4 quality measures, where we also ran the proposed algorithm with a different  $\beta$ . We can see that the proposed, CNMF, and HySure performed

Table I  
FUSION RESULTS ON THE PAVIA REGISTRATION DATASET.

	Mean	CC	SAM	RMSE	ERGAS
GT Reg	Proposed	<b>0.976</b>	<b>2.95</b>	<b>0.021</b>	<b>1.67</b>
	GFPCA	0.938	4.70	0.036	3.96
	CNMF	0.963	3.77	0.026	2.14
	Bayesian Naive	0.929	4.69	0.036	4.24
	Bayesian Sparse	0.912	6.13	0.040	3.94
	HySure	0.964	3.70	0.027	2.25
Est Reg <sup>a</sup>	Proposed	<b>0.930</b>	<b>4.31</b>	<b>0.035</b>	<b>3.94</b>
	GFPCA	0.915	5.16	0.041	4.65
	CNMF	0.913	4.78	0.039	4.15
	Bayesian Naive	0.911	5.06	0.040	4.72
	Bayesian Sparse	0.876	6.35	0.047	5.05
	HySure	0.910	4.66	0.040	4.23

<sup>a</sup> the average time costs for the 6 methods in the estimated case are 13, 9, 5, 0.6, 343, 50 seconds respectively. Despite that our solution is in a much higher dimensional space than the other methods that use a low rank representation, our time cost is still lower than several other methods.

better than the others, which is consistent with the previous comparison [4]. In most cases, the proposed is noticeably better except when the registration error is large enough approaching 0.5 pixels. Note that 0.5 pixels registration error is already unrealistically large, since it means that in average half of the HS pixel covered area is shifted. In practice, such an ill-calibrated HS image rarely exists given the subpixel accuracy of spatial calibration [47].

Comparing the proposed algorithm with different regularization strength, we see that it is quite stable with a changing  $\beta$  when there is no registration error. Comparing the slopes, our algorithm maintains the fusion quality similar to most of the methods when  $\beta = 1$ . An exception is CNMF which almost does not deteriorate with increasing registration errors. Note that the result of CNMF is not deterministic, which may account for the occasional poor performance. Also, because the image is small, the LMM assumption is somewhat valid for a small region and this assumption makes CNMF more stable to increasing registration errors. When  $\beta$  is small (0.1 or 0.01), the proposed fusion algorithm deteriorates faster. This implies that the manifold-based regularization term makes the algorithm more robust to large registration errors, showing an effect similar to the LMM assumption.

### B. Registration on Salton Sea

The Salton Sea dataset was collected by the AVIRIS on-board the ER-2 aircraft (20 km above the ground) on March 31, 2014. The IFOV for one sample is about 1 milliradian (a pixel covers a 20 m diameter region). Given its 16.9 m spatial resolution, there is some overlap between signals of neighboring pixels. We selected a small ROI (56 by 51 pixels with 224 bands) containing vegetation, a river, rooftops and a small part of a hill. The color image (738 by 674 pixels with 3 bands) was obtained from Google Earth, with the original image collected in March, 2015. There is a one year interval between the two images, which may not guarantee exact correspondence for each pixel. We use it to validate the registration algorithms. An initial scale  $s = 10.4$  was estimated between the two images, which means that the PSF

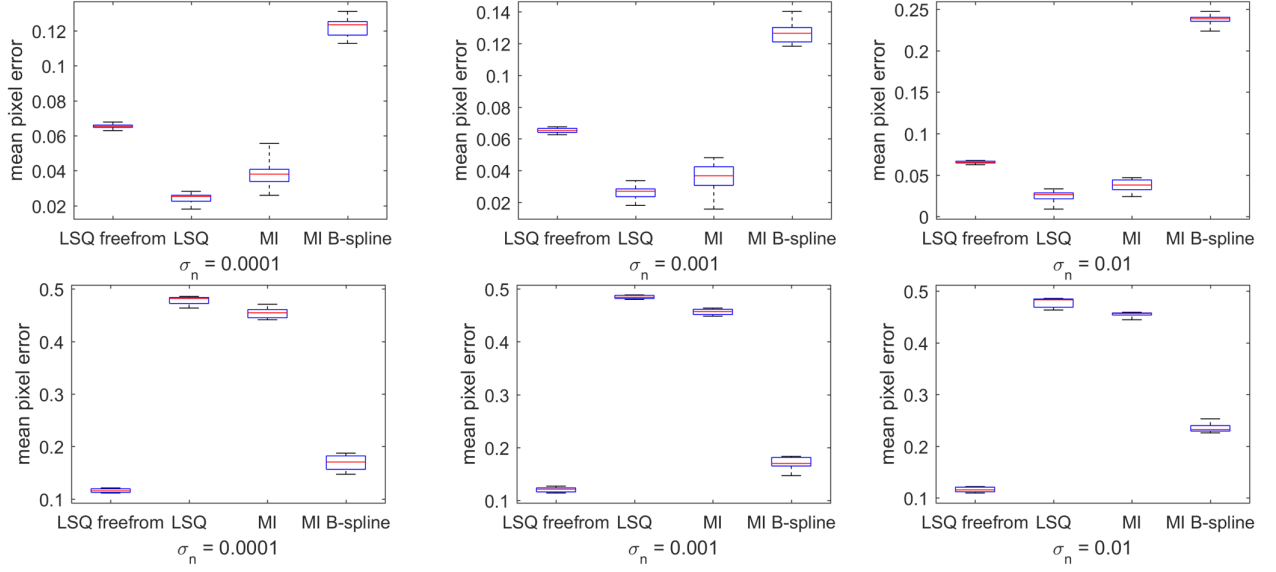


Figure 4. Quantitative comparison of the simulated Pavia dataset. The first (second) row shows the error distribution for the rigid (nonrigid) dataset.

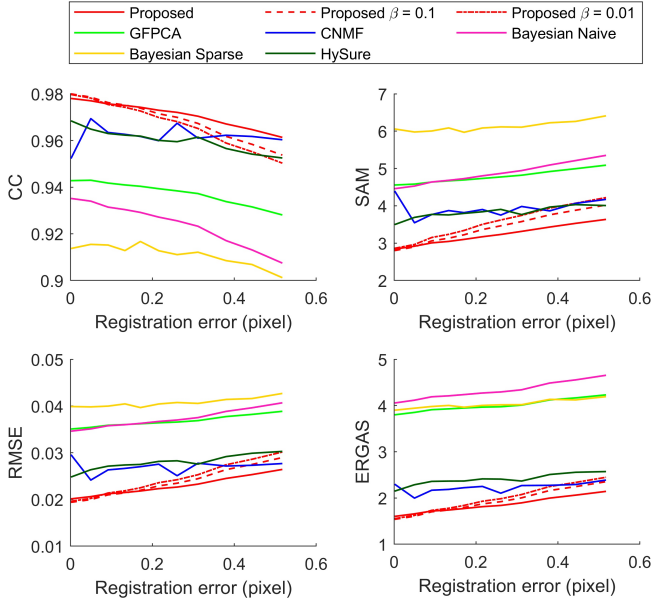


Figure 5. Fusion results on the Pavia dataset with increasing registration errors. To achieve various registration errors, the magnitude of the translation field in Fig. 3 varied from 0 to 4 pixels with an interval of 0.4 pixels. The proposed algorithm was also run with  $\beta = 0.1$  and  $\beta = 0.01$ . The accuracy is slightly higher than the results in Table I because only nonrigid transformation on the HS image is to be estimated by registration.

has  $\rho = \lceil (20/16.9) \times 10.4/2 \rceil = 7$ . The dataset is shown in Fig. 1.

Fig. 6 shows the registered HS and color images. We compared the original HS image and the transformed one visually, and marked three differences in red circles. For the top circle, the original one has a thick road segment, while the transformed one has a thinner road. Consider how narrow the road exhibits in the registered color image, the transformed one gives a better spectra distribution. For the other two circles, the boundary between the vegetation and the road is not smooth

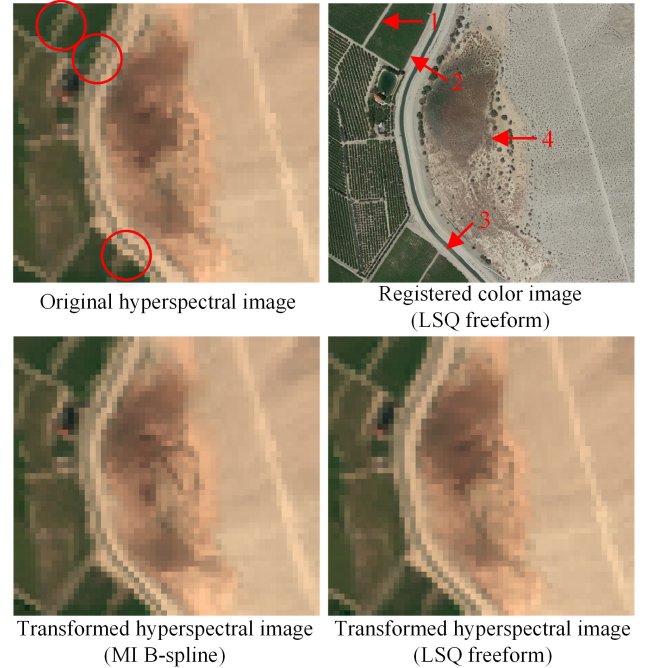


Figure 6. Qualitative results for the Salton Sea dataset. The 3 circle marked areas in the original HS image are improved in the transformed image from LSQ freeform. A detailed region-spectra correspondence for the 4 locations specified in the color image is shown in Fig. 7.

compared to the boundary in the high-resolution color image. In the transformed image, they are more smooth and better correspond to the scene. In comparison, the transformed image from MI B-spline appears to be fuzzier.

Fig. 7 shows the region-spectra correspondence from LSQ freeform, LSQ and MI B-spline (MI not shown due to its similarity to LSQ) for the arrow marked locations in Fig. 6. For the first column, LSQ (MI) has the green region consisting of all vegetation, the blue region consisting of mostly road,

however, the two regions have the same spectrum, which is a bad correspondence. In the LSQ freeform result, the green region has a spectrum between the pure vegetation spectrum (red) and the mostly road spectrum (blue), implying a better correspondence. The same phenomena repeat for the second and third columns. For the first and fourth columns, MI B-spline has very similar spectra for regions with visually different materials, implying that it tends to blur the features.

Since there is no ground truth for this dataset, we need to resort to a new metric for quantitative comparison. Considering that most fusion methods assume the validity of the linear relationship with the PSF and SRF, i.e. Eq. (4), we estimated the SRF, reconstructed the low-resolution color image, and calculated the root mean squared error between the reconstructed and the transformed signals for each pixel. Fig. 8 shows the error maps, where LSQ freeform achieves the least average error. When it comes to correlation coefficient, it also has the highest value 0.981.

### C. Fusion on Mississippi Gulfport

Finally, we validated the performance of fusion algorithms on a dataset using the same simulation as previous research, i.e. spatial and spectral downgrading without considering registration effects. We used the Mississippi Gulfport dataset, which was collected over the University of Southern Mississippi's-Gulfpark Campus [48]. It has 271 by 284 pixels, with a 1 m spatial resolution and 72 bands covering wavelengths from  $0.368 \mu\text{m}$  to  $1.043 \mu\text{m}$ . The scene is shown in Fig. 9, which contains various types of sidewalks, roads, building roofs, concrete, shrubs, trees, and grass. To generate the HS image, we used  $s = (9, 9)$ ,  $\sigma = 10$ ,  $\rho = 5.4$ , resulting in a 11 by 11 pixel blurring kernel. The SRF for the color image was the same as in Section V-A. Noise with standard deviation  $\sigma_n = 0.0001$  was applied to both the images. The resulting HS image was of size 30 by 31 pixels and the color image 270 by 279 pixels. Fig. 9 shows the generated images.

The color images of the fusion results from all the algorithms are shown in Fig. 10. Visually, the proposed, CNMF and HySure have better reconstructions to the original HS image. Bayesian Naive suffers from a strong effect from the smooth Gaussian prior. Similarly, due to the cubic interpolation approximation, GFPCA also has a smooth appearance. Table II shows the quantitative results from all the algorithms. We can see that for most measures, the proposed algorithm has the best performance, with HySure and CNMF following up. Fig. 11 shows the reconstructed spectra at the center pixel of each of the 4 cloths in Fig. 9. We can clearly see the advantage of the proposed manifold constraint in complex urban scenes.

Since we have some free parameters for tuning, the performance of the proposed method may depend on careful parameter selection. We change the parameters  $\beta$  and  $\rho_2$  with the others fixed. Fig. 12 shows the CC and RMSE with respect to changing  $\beta$  and  $\rho_2$ , along with the scores from the other methods as reference lines. We see that the accuracy is quite stable and above the competing methods when  $\beta \leq 1$  and  $\rho_2 \geq 10$ .

Table II  
FUSION RESULTS FOR THE GULFPORT DATASET

	CC <sup>c</sup>	SAM	RMSE	ERGAS
Proposed <sup>b</sup>	<b>0.947</b>	<b>5.74</b>	<b>0.038</b>	4.65
GFPCA	0.848	7.79	0.062	7.11
CNMF	0.931	6.38	0.042	4.27
Bayesian Naive	0.783	7.50	0.063	7.34
Bayesian Sparse	0.836	8.27	0.063	6.95
HySure	0.874	8.28	0.061	<b>3.86</b>

<sup>a</sup> the time costs for the 6 methods were 12, 6, 3, 0.6, 175, 41 seconds respectively.

<sup>b</sup> the proposed method was also run without input of the PSF and the SRF and a similar result (0.947, 5.74, 0.038, 4.64) was obtained.

<sup>c</sup> the standard deviation of CC is 0.055, 0.054, 0.073, 0.049, 0.103, 0.119 respectively.

## VI. DISCUSSION AND CONCLUSION

In this work, we proposed a registration algorithm and a fusion algorithm to handle HS and MS images with significant scale difference and nonrigid distortion. The registration algorithm is based on minimizing an objective function with the PSF and the SRF from the fusion literature, while a freeform transformation is applied to the HS image and a rigid transformation is applied to the MS image. The fusion algorithm minimizes a data fidelity term and a manifold-based regularization term that assumes invariance of local geometric properties after spectral degradation.

We also evaluated them in an integrated manner. Specifically, three datasets were investigated, including a Salton Sea dataset from AVIRIS and Google Earth, and two simulations generated by the Pavia University dataset and the Mississippi Gulfport dataset. For registration, we compared the proposed nonrigid version with its rigid variation and two variations based on the MI metric. The results indicate that the proposed nonrigid algorithm has the best accuracy in general, achieving less than 0.15 pixels error for nonrigid distortion of maximum 1 HS pixel (spatial calibration of airborne HS images typically has accuracy within 1 pixel [47]). For fusion, we compared the proposed algorithm with several state-of-the-art methods. The results show that it is capable of achieving state-of-the-art fusion quality.

### A. Model Assumptions

There are two alternatives to the dual geometric transformations for registration. One is to nonrigidly transform the MS image with the HS image fixed, then perform fusion and transform the fused image back. The problem is that according to Fig. 1, we may have the same spectrum mapped to different MS regions. The second alternative is to nonrigidly transform the HS image with the MS image fixed. The problem is that we may end up with too many interpolations that ruin the carefully collected HS spectra. The proposed scheme avoids these two issues. However, it brings an identifiability issue to the objective function which is invariant to a constant shift added to both the transformations. We remedied it in the implementation by terminating the update of rigid parameters early. The transformations involving the PSF and the SRF rely on significant differences in terms of spatial and spectral

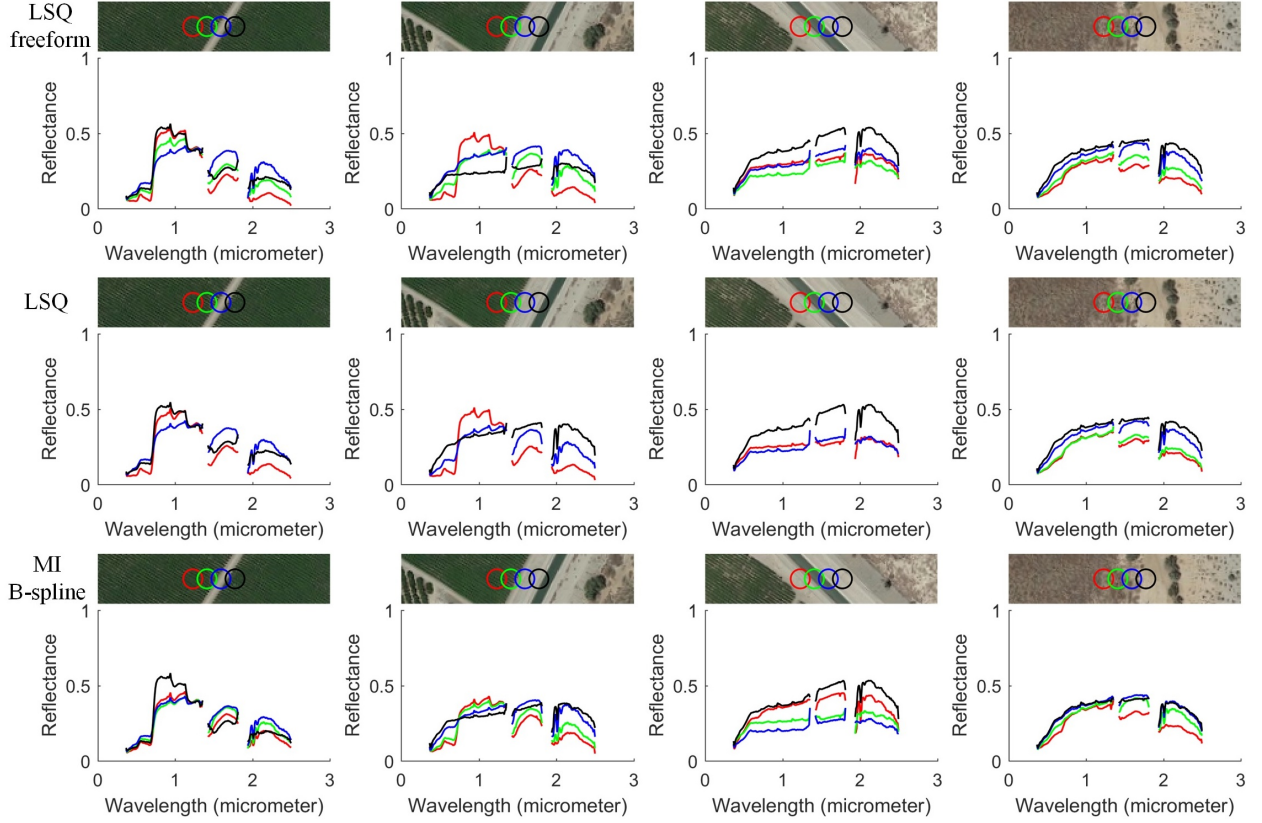


Figure 7. Qualitative region-spectra correspondence for the 4 locations in Fig. 6. The three rows are the results for LSQ freeform, LSQ, and MI B-spline respectively. For each location magnification, we select the contiguous 4 pixels (in the HS image) and denote the PSF covered regions (in the color image) by 4 different colors. The plot below shows the spectra corresponding to the 4 regions after registration (the first column of LSQ has the green and blue spectra coincide together). We expect to see continuous spectra transition according to the transition of materials.

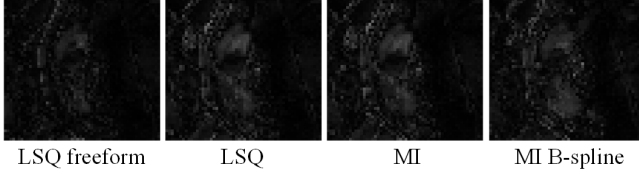


Figure 8. Residual images from reconstruction. The average errors are 5.62, 7.52, 7.62, 7.48 for the four methods respectively. The correlation coefficients for the transformed images are 0.981, 0.965, 0.964, 0.967.

resolution, which is exactly the case we aim for. The PSF transformation is valid when the HS and the MS sensors operate nearby at the same time. When this condition is not satisfied, the performance will be affected by shadowing and multiple scattering from nonflat terrain surface.

To validate if the assumption of the regularization term in fusion holds, we put the ground truth  $\mathbf{R}$  into the regularization term  $\text{Tr}(\mathbf{R}^T \mathbf{L} \mathbf{R})$  to quantify how accurately the reverse of (17) holds. Specifically, for the Gulfport dataset, we created multiple MS images with increasing number of bands (decreasing bandwidths) and calculated  $\mathbf{L}$ 's using these MS images to evaluate this quantity. We also ran the proposed fusion algorithm on them to check the accuracy. Fig. 13 shows this quantity and fusion error versus various  $b$ . The figure shows that the validity of this assumption is in line with the

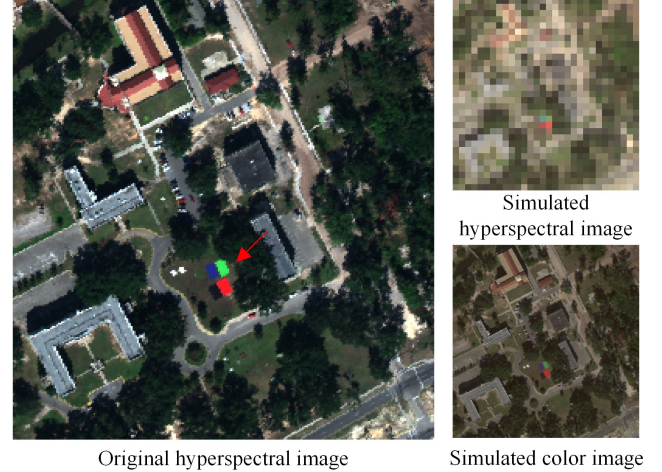


Figure 9. Original HS image and generated test HS and color images. There are 4 cloths in the scene (indicated by a red arrow) which we will inspect for fusion quality.

fusion quality, and that starting from  $b = 4$ , the curve has plateaued, reaching  $3e-4$ , which is possibly the random noise in the original HS image. Even for a small  $b = 3$ , its value is only five times the saturated value. Hence, this verifies our assumption that the coefficients estimated from the MS image can be used to constrain the unknown HS image. This could

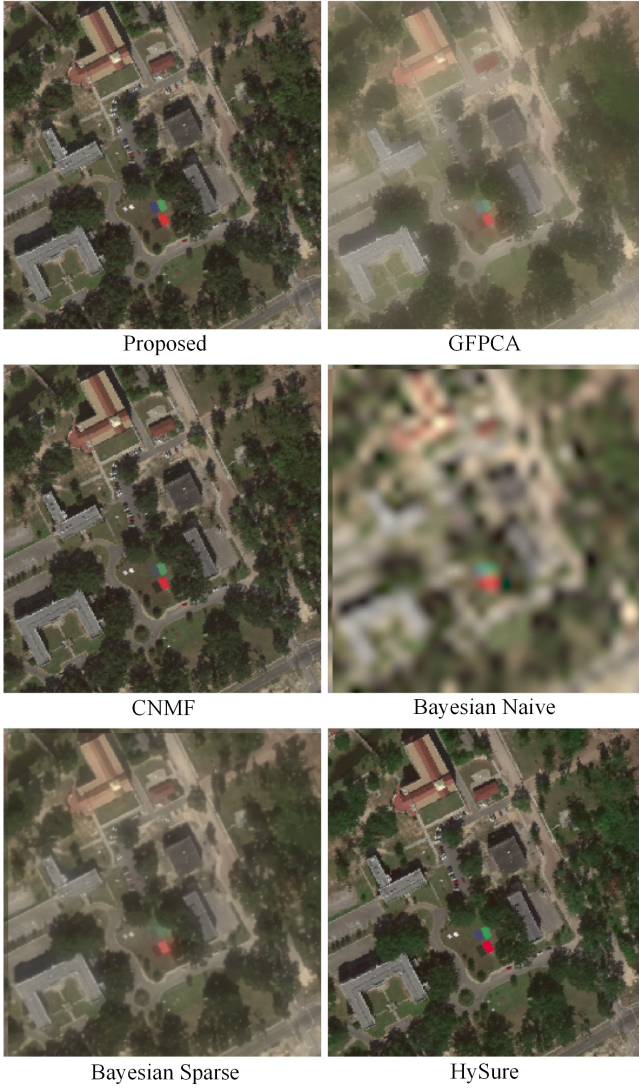


Figure 10. Visible bands from the fusion results for the Gulfport dataset.

also explain the robustness of  $\beta$  in Fig. 12. Since the original fusion problem (15) is undetermined (an infinite number of solutions exist), a slight regularization should work well. Also, since this regularization really captures the structure of the data, it can tolerate a relatively large  $\beta$ .

### B. Parameter Selection

Here we discuss the strategy to select the registration and fusion parameters. In registration,  $\alpha$  controls the smoothness of the nonrigid deformation. Considering the interpolation issue illustrated in Fig. 1, it should be set to be small, e.g.  $\alpha = 0.05$ .  $\lambda$  controls the smoothness of the SRF. We visually checked the obtained SRF from the Salton Sea dataset to determine  $\lambda = 10^{-3}N$ .  $\Delta t$  controls the convergence rate of the PDE. In the experiments, we found that  $\Delta t = 1$  was small enough to guarantee that the difference scheme converges. In fusion,  $\gamma$  balances the two data fidelity terms. Clearly,  $\gamma = 0.5$  is an ideal choice. We set  $K = 3$  by considering the number of bands in the MS image and the computational complexity.  $\beta$  should be less than or equal to one according to Fig. 12.

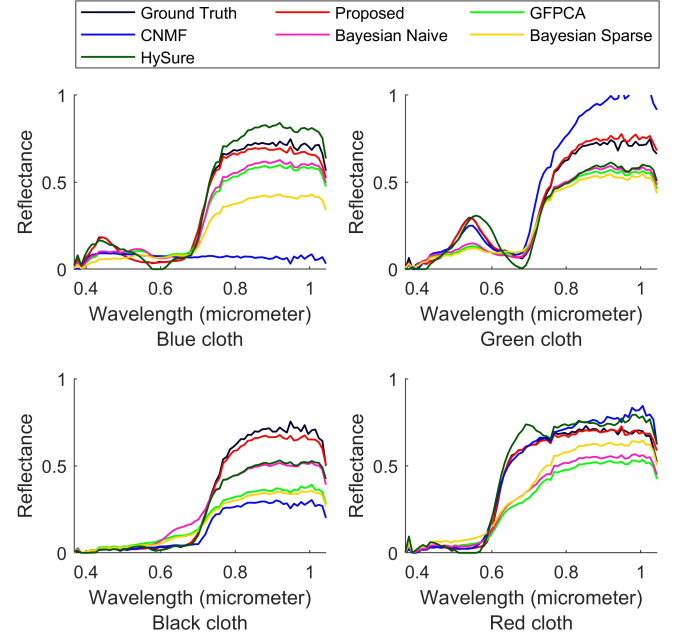


Figure 11. Reconstructed spectra at the center pixel of each of the 4 cloths in Fig. 9. Compared to the LMM based methods, more spectral details are preserved by our method.

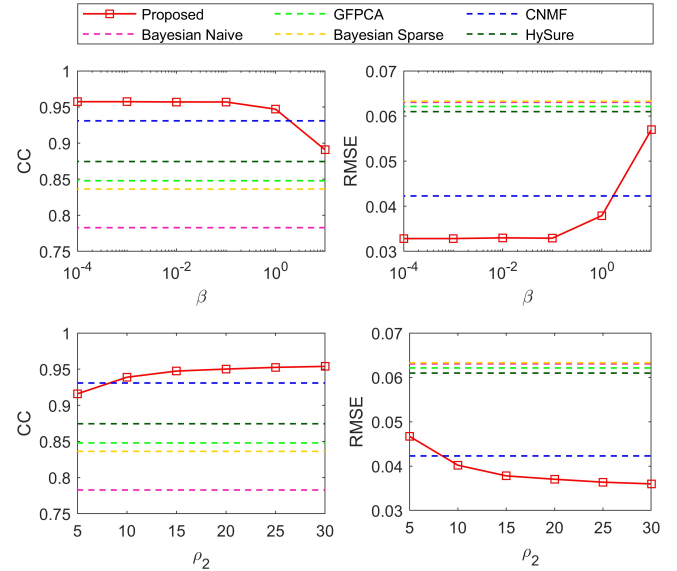


Figure 12. Parameter analysis of the proposed method on the Gulfport dataset. The accuracy is stable and above the other methods when  $\beta \leq 1$  and  $\rho_2 \geq 10$ .

Considering the registration errors in Fig. 5, a slight large value in this range was chosen.  $\rho_2$  should be as large as possible since the manifold should only depend on the spectral information (considering the actual time cost, we set  $\rho_2 = 15$ ). This is also verified in Fig. 12.  $\epsilon$  is introduced to speed up the computation and avoid identifiability issues (since the neighbors in (16) have similar spectra to the current pixel). A small  $\epsilon = 1e-4$  is sufficient for these two purposes.

In summary, we would suggest keeping  $\Delta t = 1$ ,  $\gamma = 0.5$ ,  $\epsilon = 1e-4$  fixed for all the datasets and noise levels. For the other parameters, they may be tuned based on the datasets and noise levels. For example, the scale of the dataset (e.g.

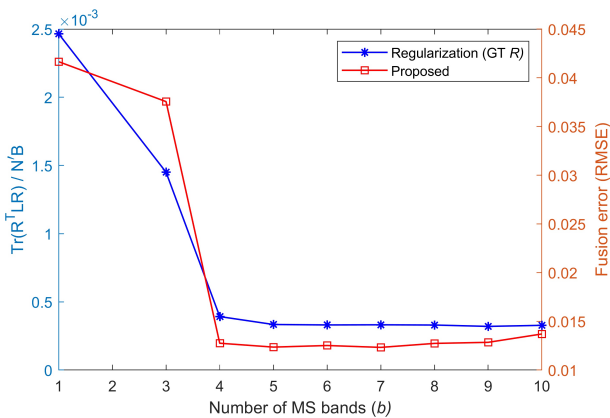


Figure 13. Regularization fitting error (blue) and final fusion error (red) versus number of bands for the Gulfport dataset. The y-axis has two scales corresponding to  $\text{Tr}(\mathbf{R}^T \mathbf{L} \mathbf{R}) / N'B$  (blue) and RMSE (red) respectively.

number of pixels) may influence the parameters, though the algorithm already calibrated it to some extent (e.g. step 1 in Algorithm 1). When we apply the proposed approach to rural scenes where large homogeneous areas are present, a wider range of parameters may be accepted (since urban imagery with complex scenes poses more difficulties in fusion, e.g. the fusion accuracy on the Pavia and Gulfport datasets in this paper is slightly lower than the rural scene in the pansharpening review [3]). For the noise level, a large random noise will lead to large fitting errors of the least-squares terms. Hence, we may need a large  $\alpha$ ,  $\lambda$ ,  $\beta$  to balance the regularization terms against the data fidelity terms.

### C. Limitation and Future Work

There are several limitations to this research. First, the proposed registration algorithm relies on pixel level adjustment, which may be too expensive for large remote sensing images. Second, the application of the proposed fusion algorithm to pansharpening (i.e. the MS image becomes a panchromatic image with only 1 band) is less desirable. A direct application to sample images from [3] leads to a CC of 0.973, compared to 0.935 from GFPCA, 0.971 from CNMF, 0.977 from Bayesian Naive, 0.980 from Bayesian Sparse, 0.979 from HySure. This is expected as the less bands, the less likely that the reverse of (17) is true (Fig. 13). Third, for the real dataset, we do not have ground truth and the smallest time interval we can find is one year which is too large. Though most previous research relies on simulated datasets, we still tried our algorithm on a real dataset, which poses several problems (Fig. 1) not considered before. Future work could extend the validation dataset from airborne (AVIRIS) and space (Google Earth) images to airborne and ground images. Moreover, impact of the terrain surface may be considered in the registration and fusion process. Future work could also include using the proposed integrated approach to build ground truth for spectral unmixing.

### REFERENCES

[1] F. D. Van der Meer, H. M. Van der Werff, F. J. van Ruitenbeek, C. A. Hecker, W. H. Bakker, M. F. Noomen, M. van der Meijde, E. J. M.

Carranza, J. B. de Smeth, and T. Woldai, "Multi-and hyperspectral geologic remote sensing: A review," *International Journal of Applied Earth Observation and Geoinformation*, vol. 14, no. 1, pp. 112–128, 2012.

[2] 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 Sensing of Environment*, vol. 44, no. 2-3, pp. 127–143, 1993.

[3] L. Loncan, L. B. de Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes *et al.*, "Hyperspectral pansharpening: a review," *IEEE Geoscience and remote sensing magazine*, vol. 3, no. 3, pp. 27–46, 2015.

[4] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 2, pp. 29–56, 2017.

[5] D. A. Roberts, M. Gardner, R. Church, S. Ustin, G. Scheer, and R. Green, "Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models," *Remote Sensing of Environment*, vol. 65, no. 3, pp. 267–279, 1998.

[6] R. L. Powell, D. A. Roberts, P. E. Dennison, and L. L. Hess, "Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil," *Remote Sensing of Environment*, vol. 106, no. 2, pp. 253–267, 2007.

[7] E. B. Wetherley, D. A. Roberts, and J. P. McFadden, "Mapping spectrally similar urban materials at sub-pixel scales," *Remote Sensing of Environment*, vol. 195, pp. 170–183, 2017.

[8] Y. Zhou, E. B. Wetherley, and P. D. Gader, "Unmixing urban hyperspectral imagery with a Gaussian mixture model on endmember variability," *arXiv preprint arXiv:1801.08513*, 2018.

[9] P. Blanc, L. Wald, and T. Ranchin, "Importance and effect of co-registration quality in an example of pixel to pixel fusion process," in *2nd International Conference Fusion of Earth Data: merging point measurements, raster maps and remotely sensed images*. SEE/URISCA, Nice, France, 1998, pp. 67–74.

[10] N. Yokoya, N. Mayumi, and A. Iwasaki, "Cross-calibration for data fusion of EO-1/Hyperion and Terra/ASTER," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 419–426, 2013.

[11] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[12] W. Lu, M.-L. Chen, G. H. Olivera, K. J. Ruchala, and T. R. Mackie, "Fast free-form deformable registration via calculus of variations," *Physics in Medicine and Biology*, vol. 49, no. 14, p. 3067, 2004.

[13] P. Thévenaz and M. Unser, "A pyramid approach to sub-pixel image fusion based on mutual information," in *Image Processing, 1996. Proceedings., International Conference on*, vol. 1. IEEE, 1996, pp. 265–268.

[14] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *International journal of computer vision*, vol. 24, no. 2, pp. 137–154, 1997.

[15] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE transactions on image processing*, vol. 5, no. 8, pp. 1266–1271, 1996.

[16] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE transactions on image processing*, vol. 11, no. 3, pp. 188–200, 2002.

[17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[18] C. Harris and M. Stephens, "A combined corner and edge detector," in *Alvey vision conference*, vol. 15. Citeseer, 1988, p. 50.

[19] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Computer Vision and Image Understanding*, vol. 89, no. 2, pp. 114–141, 2003.

[20] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.

[21] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, and J. Tian, "Robust feature matching for remote sensing image registration via locally linear transforming," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 12, pp. 6469–6481, 2015.

[22] Y. Bentoutou, N. Taleb, K. Kpalma, and J. Ronsin, "An automatic image registration for applications in remote sensing," *IEEE transactions on geoscience and remote sensing*, vol. 43, no. 9, pp. 2127–2137, 2005.

- [23] X. Fan, H. Rhody, and E. Saber, "A spatial-feature-enhanced mmi algorithm for multimodal airborne image registration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2580–2589, 2010.
- [24] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013.
- [25] W. Liao, X. Huang, F. Van Coillie, S. Gautama, A. Pižurica, W. Philips, H. Liu, T. Zhu, M. Shimoni, G. Moser *et al.*, "Processing of multiresolution thermal hyperspectral and digital color data: Outcome of the 2014 ieee grss data fusion contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2984–2996, 2015.
- [26] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. D. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, 2012.
- [27] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel, "Unmixing-based multisensor multiresolution image fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1212–1226, 1999.
- [28] N. Yokoya, T. Yairi, and 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.
- [29] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1174–1184, 2004.
- [30] M. T. Eismann and R. C. Hardie, "Hyperspectral resolution enhancement using high-resolution multispectral imagery with arbitrary response functions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 455–465, 2005.
- [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 Processing*, vol. 24, no. 11, pp. 4109–4121, 2015.
- [32] 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, 2015.
- [33] M. Simões, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3373–3388, 2015.
- [34] R. Heylen, M. Parente, and P. D. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 1844–1868, 2014.
- [35] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 95–104, 2014.
- [36] M. A. Veganzones, M. Simoes, G. Licciardi, N. Yokoya, J. M. Bioucas-Dias, and J. Chanussot, "Hyperspectral super-resolution of locally low rank images from complementary multisource data," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 274–288, 2016.
- [37] L. Zhang, W. Wei, C. Bai, Y. Gao, and Y. Zhang, "Exploiting clustering manifold structure for hyperspectral imagery super-resolution," *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 5969–5982, 2018.
- [38] 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.
- [39] Y. Zhou, A. Rangarajan, and P. D. Gader, "Nonrigid registration of hyperspectral and color images with vastly different spatial and spectral resolutions for spectral unmixing and pansharpening," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, July 2017, pp. 1571–1579.
- [40] C. Huang, J. R. Townshend, S. Liang, S. N. Kalluri, and R. S. DeFries, "Impact of sensor's point spread function on land cover characterization: assessment and deconvolution," *Remote Sensing of Environment*, vol. 80, no. 2, pp. 203–212, 2002.
- [41] K. A. Kalpoma and J.-i. Kudoh, "Image fusion processing for IKONOS 1-m color imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3075–3086, 2007.
- [42] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.
- [43] D. P. Bertsekas, *Nonlinear programming*. Athena Scientific, 1999.
- [44] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [45] D. C. Heinz and C. "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE transactions on geoscience and remote sensing*, vol. 39, no. 3, pp. 529–545, 2001.
- [46] A. El Guennouni, K. Jbilou, and A. Riquet, "Block Krylov subspace methods for solving large Sylvester equations," *Numerical Algorithms*, vol. 29, no. 1-3, pp. 75–96, 2002.
- [47] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote sensing of environment*, vol. 65, no. 3, pp. 227–248, 1998.
- [48] P. Gader, A. Zare, R. Close, J. Aitken, and G. Tuell, "MUUFL Gulfport hyperspectral and LiDAR airborne data set," Univ. Florida, Gainesville, FL, USA, Tech. Rep. REP-2013-570, 2013.



computer vision and machine learning.



**Yuan Zhou** received the Ph.D. degree in Computer Science (2018) from the Department of CISE, University of Florida, Gainesville, FL, USA. Prior to that, he received the B.E degree in Software Engineering (2008), the M.E. degree in Computer Application Technology (2011), both from Huazhong University of Science and Technology, Wuhan, Hubei, China. Since 2018, he has been a postdoctoral associate in the Department of Radiology and Biomedical Imaging, Yale University, New Haven, CT, USA. His research interests include image processing,

**Anand Rangarajan** is a Professor in the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, USA. His research interests are machine learning, computer vision, hyperspectral and medical imaging and the scientific study of consciousness.



**Paul Gader** (M'86–SM'09–F'11) received the Ph.D. degree in mathematics for image-processing-related research from the University of Florida, Gainesville, FL, USA, in 1986. He was a Senior Research Scientist with Honeywell, a Research Engineer and a Manager with the Environmental Research Institute of Michigan, Ann Arbor, MI, USA, and a Faculty Member with the University of Wisconsin, Oshkosh, WI, USA, the University of Missouri, Columbia, MO, USA, and the University of Florida, FL, USA, where he is currently a Professor of Computer and Information Science and Engineering. He performed his first research in image processing in 1984 working on algorithms for the detection of bridges in forward-looking infrared imagery as a Summer Student Fellow at Eglin Air Force Base. He has since worked on a wide variety of theoretical and applied research problems including fast computing with linear algebra, mathematical morphology, fuzzy sets, Bayesian methods, handwriting recognition, automatic target recognition, biomedical image analysis, landmine detection, human geography, and hyperspectral and light detection, and ranging image analysis projects. He has authored/co-authored hundreds of refereed journal and conference papers.