Hyperspectral and Multispectral Image Fusion based on a Non-locally Centralized Sparse Model and Adaptive Spatial-Spectral Dictionaries

Kevin Arias[1], Edwin Vargas[2], Henry Arguello[3]
[1]*Dept. of Physics*, [2]*Dept. of Electrical Engineering*, [3]*Dept. of Computer Science*
Universidad Industrial de Santander, Bucaramanga, Colombia

***Abstract*—Hyperspectral (HS) imaging systems are useful in a diverse range of applications that involve detection and classification tasks. However, the low spatial resolution of hyperspectral images may limit the performance of the involved tasks in such applications. In the last years, fusing the information of a HS image with high spatial resolution multispectral (MS) or panchromatic (PAN) images has been widely studied to enhance the spatial resolution. Image fusion has been formulated as an inverse problem whose solution is a HS image which assumed to be sparse in an analytic or learned dictionary. This work proposes a non-local centralized sparse representation model on a set of learned dictionaries in order to regularize the conventional fusion problem. The dictionaries are learned from the observed data taking advantage of the high spectral correlation within the HS image and the non-local self-similarity over the spatial domain of the MS image. Then, conditionally on these dictionaries, the fusion problem is solved by an alternating iterative numerical algorithm. Experimental results with real data show that the proposed method outperforms the state-of-the-art methods under different quantitative assessments.**

I. Introduction

Hyperspectral (HS) imaging consists of acquiring a scene in several hundreds of contiguous spectral bands, each one captured at a specific wavelength. HS images are characterized by a high spectral resolution which allows accurate identification of the different materials contained in the scene of interest. Analyzing the spectral information of HS images has allowed the development of many applications in the fields of remote sensing [1], medical imaging [2] or astronomy [3]. However, due to technological reasons, HS images are limited by their relatively low spatial resolution [4]. For instance, the Hyperion imaging spectrometer has about 220 spectral bands, which extend from the visible region (0.4 to 0.7 μm) through the SWIR (about 2.5 μm), with a spatial resolution of 30 m by pixel [5] that can be insufficient for some practical applications.

To overcome the spatial resolution limitation, a common trend is to fuse images with different spectral and spatial resolutions. A typical example studied in this work is the fusion of HS images (having high spectral resolution) with multispectral (MS) images (having high spatial resolution) [6], [7]. Another example is HS pansharpening, which addresses the fusion of panchromatic and HS images [8].

This work was supported by UIS-ECOPETROL through the grant titled "Acuerdo de Cooperacin No. 27 derivado del CONVENIO MARCO No. 5222395".

Due to the ill-posed nature of the inverse problem behind the image fusion problem, it is important to find an appropriate model taking into account the prior knowledge of natural images. Effective regularizers that restrict the solution spaces have been widely used in image restoration (IR) problems obtaining promising results [9], [10], [11], [12]. In particular, total-variation-based modeling showed good performance for the image fusion problem under a super-resolution approach [7]. In addition, sparsity-based modeling has been proven to be an effective regularization method in IR problems such as image super-resolution, image denoising, image deblurring and image fusion [13], [9], [14], [6]. More formally, a sparse representation model implies that a signal $\mathbf{x} \in \mathbb{R}^{N_p}$ can be represented as a linear combination of few atoms from a dictionary $\mathbf{\Phi}$ via $\mathbf{x} = \mathbf{\Phi}\boldsymbol{\alpha}_x$. If the signals of interest are two dimensional images, a common approach is to represent image patches with an over-complete dictionary learned from the data achieving better representations of the image structures compared to analytically designed dictionaries such as wavelet transforms [15].

On the other hand, the IR problem of estimating $\mathbf{x}$ from an observed degraded image $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$ using sparsity constraints over $\mathbf{x}$ is a challenging problem due to the degradation process. The estimated sparse code $\boldsymbol{\alpha}_y$ using traditional algorithms such as basis pursuit, Lasso, iterative shrinkage/thresholding, and Bayesian frameworks [16], [17], [18], [19], [20] may not be so close to the sparse code $\boldsymbol{\alpha}_x$ of the original image. Therefore, the reconstruction $\hat{\mathbf{x}} = \mathbf{\Phi}\boldsymbol{\alpha}_y$ may not lead to an accurate estimation [21], [22]. As an alternative, properties of non-local redundancies existing in natural images are exploited to improve the accuracy of the sparse model. For instance, different approaches have been proposed for combining learned dictionaries with the non-local self similarities of natural images leading to state-of-the-art algorithms in tasks such as denoising or superresolution [21], [23], [24], [25]. In the same direction, a non-locally centralized sparse representation (NCSR) model has been proposed in [22] to further improve the sparse representation-based image restoration methods. In detail, the model aims to improve the quality of the reconstructed image $\hat{\mathbf{x}}$ by centralizing the estimated sparse code $\boldsymbol{\alpha}_y$ to a close estimation of $\boldsymbol{\alpha}_x$ reducing the error $\boldsymbol{\nu}_\alpha = \boldsymbol{\alpha}_y - \boldsymbol{\alpha}_x$ referred to as sparse coding noise (SCN). In practice, a good estimate of $\boldsymbol{\alpha}_x$ is obtained by exploiting a large amount of non-local redundancies present

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

in the image $\mathbf{x}$.

This work proposes a hyperspectral and multispectral image fusion method using a NCSR model of the high-resolution (HR) HS image in a set of learned dictionaries. Specifically, the non-local self-similarities of natural images are exploited by forming groups of similar cubic patches of the HR HS image. Then, each cluster is sparsely represented on a unique adaptive spatial-spectral dictionary whose sparse codes $\boldsymbol{\alpha}_y$ are centralized to a closer estimation of the sparse code $\boldsymbol{\alpha}_x$. This estimation is obtained by also exploiting the non-local redundancy of natural images. The dictionaries for each cluster are built from the data taking advantage of the high spatial resolution of the MS image and the high spectral correlation of the HS image. The resulting image fusion problem is solved by an alternating and iterative shrinkage algorithm which allows to divide the problem in two parts, a quadratic problem and a shrinkage operator. Experimental results with real data show that the proposed method achieves and outperforms the state-of-the-art performance by exhibiting a gain of up to 1[dB] in terms of peak signal-to-noise ratio (PSNR).

II. Problem Statement

It is very common to assume that HS and MS images result from the application of linear spatial and linear spectral degradation to a higher resolution image which is represented by the vector $\mathbf{x} \in \mathbb{R}^{N_p}$ with $N_p = NML$ [26], [27], [7]. The observed HS image represented by the vector $\mathbf{y}_H \in \mathbb{R}^{N_H M_H L}$ is supposed to be a blurred and downsampled version of the target full-resolution image $\mathbf{x}$. On the other hand, the MS image $\mathbf{y}_M \in \mathbb{R}^{NML_M}$ is a spectrally degraded version of the target image [6]. Thus, the resulting acquisition models for the HS and MS images can be written in vector form as

$$\begin{aligned} \mathbf{y}_H &= \mathbf{PBx} + \mathbf{n}_H \\ \mathbf{y}_M &= \mathbf{Rx} + \mathbf{n}_M, \end{aligned} \tag{1}$$

where $\mathbf{P} \in \mathbb{R}^{N_H M_H L \times N_p}$ is the downsampling matrix, $\mathbf{B} \in \mathbb{R}^{N_p \times N_p}$ is the blurring matrix acting as a cyclic convolution operator, $\mathbf{R} \in \mathbb{R}^{NML_M \times N_p}$ is the spectral response of the MS sensor, and the pair of vectors $\mathbf{n}_H \in \mathbb{R}^{N_H M_H L}$ and $\mathbf{n}_M \in \mathbb{R}^{NML_M}$ are additive noise terms for the HS and MS images, respectively. Note that any spatial-spectral dimensional source $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{N \times M \times L}$ ($N \times M$ is used for the spatial dimensions and L is the spectral dimension) is represented by the vector $\mathbf{x} = [\bar{\mathbf{x}}_1^T, ..., \bar{\mathbf{x}}_L^T]^T \in \mathbb{R}^{NML}$, where $\bar{\mathbf{x}}_i \in \mathbb{R}^{NM}$ contains all the image intensities associated with the i-th spectral band. Thus, taking the observation models in (1) into account, the image restoration problem considered in this paper consists of estimating the high-resolution (HR) image $\mathbf{x}$ from the observed measurements $\mathbf{y}_H$ and $\mathbf{y}_M$. This problem is strongly ill-posed and we propose to consider a non-local centralized sparse model in order to regularize its solution.

III. Hyperspectral and Multispectral image fusion based on a NCSR model

First of all, consider that the target high-resolution HS image $\mathbf{x}$ is divided into overlapping image cubic patches. This decomposition has been shown to be very effective in many image processing applications [28]. Let $\mathbf{x}_i = \mathbf{W}_i \mathbf{x} \in \mathbb{R}^{S^2 L}$ be the vector representation of a cubic image patch of size $S \times S \times L$ where the linear index i represents the location of the central pixel of the patch, and $\mathbf{W}_i \in \mathbb{R}^{S^2 L \times N_p}$ is the matrix extracting cubic patch $\mathbf{x}_i$ at location i. Furthermore, consider that each cubic patch can be sparsely represented in a given dictionary, i.e. $\mathbf{x}_i = \boldsymbol{\Phi}\boldsymbol{\alpha}_{x,i}$, where $\boldsymbol{\alpha}_{x,i}$ is a sparse vector. The reconstruction of $\mathbf{x}$ from the set of sparse codes $\{\boldsymbol{\alpha}_{x,i}\}$ can be calculated by averaging all overlapping patches leading to

$$\hat{\mathbf{x}} = \left(\sum_i \mathbf{W}_i^T \mathbf{W}_i \right)^{-1} \sum_i \left(\mathbf{W}_i^T \boldsymbol{\Phi} \boldsymbol{\alpha}_{x,i} \right). \tag{2}$$

For notation convenience we denote the estimation $\hat{\mathbf{x}}$ in (2) as $\hat{\mathbf{x}} := \boldsymbol{\Phi} \circ \boldsymbol{\alpha}_x$ where $\boldsymbol{\alpha}_x$ denotes the concatenation of all $\boldsymbol{\alpha}_{x,i}$.

A. Non-local Sparse Representation based Model

Several sparsity-based regularization approaches have been employed in different IR problems to manage its ill-posed nature [13], [6], [14], [29]. However, due to the degradation process of the observed images, the estimated sparse code $\boldsymbol{\alpha}_y$ deviates from the sparse code $\boldsymbol{\alpha}_x$ of the original image compromising the quality of the recovered image $\hat{\mathbf{x}}$. An effective strategy to further improve the sparsity results has been to employ prior knowledge from the non-local similarities of the underlying image [21], [22]. Therefore, it makes sense that a non-local alternative regularization benefits the image fusion problem. Specifically, this work proposes a NCSR [22] of the underlying HR image to regularize the image fusion problem. Based on the NCSR model and the observation models in (1) we consider the following optimization problem

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_y = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \; & \frac{1}{2} \| \mathbf{y}_H - \mathbf{PB} \left(\boldsymbol{\Phi} \circ \boldsymbol{\alpha} \right) \|_2^2 \\ & + \frac{1}{2} \| \mathbf{y}_M - \mathbf{R} \left(\boldsymbol{\Phi} \circ \boldsymbol{\alpha} \right) \|_2^2 + \lambda \sum_i \| \boldsymbol{\alpha}_i - \boldsymbol{\beta}_i \|_1, \end{aligned} \tag{3}$$

where the first two terms denote the data fidelity with respect to the HS and MS images, the last term is associated to the NCSR sparsity constraint, $\boldsymbol{\beta}_i$ is some good estimation of $\boldsymbol{\alpha}_{x,i}$ and λ is a regularization parameter. The problem in (3) induces sparsity on $\boldsymbol{\alpha}_i$ and at same time centralizes it to the estimation $\boldsymbol{\beta}_i$ suppressing the SCN $\boldsymbol{\alpha}_y - \boldsymbol{\alpha}_x$ [22]. Based on the non-local redundancies of natural images, the estimation $\boldsymbol{\beta}_i$ of $\boldsymbol{\alpha}_{x,i}$ using an iterative optimization algorithm can be calculated as a weighted average of some sparse codes $\boldsymbol{\alpha}_{y,p}$ such that $p \in \boldsymbol{\Omega}_i$, $\boldsymbol{\Omega}_i \subseteq \{1, 2, ...\}$. In detail, the sparse codes $\boldsymbol{\alpha}_{y,p}$ are the sparse representation of similar patches to $\hat{\mathbf{x}}_i$ extracted from the image $\hat{\mathbf{x}}$ in the current iteration. Thus $\boldsymbol{\beta}_i$ can be calculated as

$$\begin{aligned} \boldsymbol{\beta}_i &= \sum_{p \in \boldsymbol{\Omega}_i} \omega_p \boldsymbol{\alpha}_{y,p}, \\ \omega_p &= \frac{1}{\varphi} exp\left(\frac{-\| \hat{\mathbf{x}}_i - \hat{\mathbf{x}}_p \|_2^2}{h} \right) \end{aligned} \tag{4}$$

where ω_p is the associated weight, h is a predefined scalar and φ is a normalization factor. Note that the indexes i and $p \in \mathbf{\Omega}_i$ are employed to differentiate the current patch and its similar neighboring patches, respectively.

B. Dictionary learning

Studies about the human visual system suggest that natural images can be sparsely represented in some appropriate domain [30]. Sparsifying domains have been described by analytic and learned dictionaries which contain basis functions that represent structural primitives of conventional image. Let each cubic patch $\mathbf{x}_i$ be represented as a matrix $\mathbf{X}_i \in \mathbb{R}^{S^2 \times L}$ where the rows contain all the spectral pixels of the cubic patch. Then, the patch $\mathbf{X}_i$ can be represented by a small number of atoms as follows [23]

$$\mathbf{X}_i = \sum_m \alpha_m \mathbf{\Phi}_m \tag{5}$$

where $\{\mathbf{\Phi}_m\}$ is the set of dictionary elements (matrix atoms) and α_m are the respective scalar coefficients with $m \ll S^2 L$.

One important issue of sparsity-based methods is the selection of the dictionary elements $\mathbf{\Phi}_m$. In this work, we propose to consider spatial-spectral dictionary elements that are constructed using separable spatial and spectral components [31], that is

$$\mathbf{\Phi}_m = \mathbf{v}_d \mathbf{u}_r^T \tag{6}$$

where $\{\mathbf{v}_d\}_{d=1}^{S^2}$ and $\{\mathbf{u}_r\}_{r=1}^{L}$ are orthonormal basis spanning the spatial and spectral spaces, respectively. The main motivation to do that is to exploit the complementary information of the HS and MS images. More precisely, the base for the spectral domain $\mathbf{U}$ is extracted from the high-resolution HS image, and the base for the spatial domain $\mathbf{V}$ from the high spatial resolution MS image allowing to synthesize dictionary elements in (6) with simultaneously high spectral and high spatial resolutions. In the spectral domain, the HS image $\mathbf{y}_H$ is reorganized as a $L \times N_H M_H$ matrix to learn via PCA the spectral basis $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_L]$. Likewise, the spatial basis $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{S^2}]$ is trained using PCA from information extracted along all bands of the MS image $\mathbf{y}_M$.

In the spatial domain, instead of learning a single dictionary for all image patches, an alternative approach is to cluster the training patches and learn a dictionary for each cluster. This strategy has been shown to be a flexible solution leading to sparse representations on any spatial patch [32], [22]. Therefore, we cluster by similarity spatial patches of $\mathbf{y}_M$ pooled along all the bands using the K-means algorithm. This process consists in collecting spatial patches in clusters C_k, with $k = 1, 2, ..., K$, by their high frequency patterns since clustering by intensity has presented discrimination problems [32], [23]. The high frequency patterns $\tilde{\mathbf{y}}_M$ of $\mathbf{y}_M$ can be calculated as

$$\tilde{\mathbf{y}}_M = \mathbf{y}_M - \mathbf{G}\mathbf{y}_M, \tag{7}$$

where $\mathbf{G}$ represents a blurring operator. Here, it is important to note that each cluster C_k has a corresponding spatial basis $\mathbf{V}_k = \left[\mathbf{v}_1^{(k)}, \mathbf{v}_2^{(k)}, ..., \mathbf{v}_{S^2}^{(k)}\right]$ obtained directly from the information of $\mathbf{y}_M$ and the image $\tilde{\mathbf{y}}_M$ is only employed for the clustering process. Combining the PCA spectral and spatial basis following (6), the spatial-spectral dictionary for any patch $\mathbf{x}_i$ in cluster C_k can be expressed as

$$\bar{\mathbf{\Phi}}_k = \mathbf{V}_k \otimes \mathbf{U}, \tag{8}$$

where $\bar{\mathbf{\Phi}}_k = \left[\boldsymbol{\phi}_1^{(k)}, \boldsymbol{\phi}_2^{(k)}, ..., \boldsymbol{\phi}_{S^2 L}^{(k)}\right]$ and $\otimes$ denotes the Kronecker operator. It is worth noting that each cubic patch $\mathbf{x}_i$ needs to be represented in a single spatial-spectral dictionary $\bar{\mathbf{\Phi}}_k$.

IV. NCSR-based Fusion Algorithm

The HS and MS image fusion problem based on a NCSR model in (3) is solved by an iterative shrinkage algorithm [22] described in Algorithm 1. Two main loops are implemented to perform the necessary updates. The outer loop indexed by t iteratively updates the spatial-spectral dictionary $\bar{\mathbf{\Phi}}_k$. Then, for fixed $\bar{\mathbf{\Phi}}_k$ and $\boldsymbol{\beta}$ the inner loop indexed by q finds an estimation $\hat{\mathbf{x}}$ by using an alternating two-step procedure. First, the solution of the quadratic problem related to the image fidelity is solved by a gradient descent algorithm. Second, a shrinkage operator [33], [22] is used to estimate the non-local centralized sparse codes $\hat{\boldsymbol{\alpha}}_y$. Thus, in the $(q+1)$-th iteration, $\hat{\boldsymbol{\alpha}}_y$ is computed as

$$\hat{\boldsymbol{\alpha}}_{y,i}^{(q+1)} = \mathcal{H}_\tau \left(\hat{\boldsymbol{\alpha}}_{y,i}^{(q)} - \boldsymbol{\beta}_i\right) + \boldsymbol{\beta}_i \tag{9}$$

where $\mathcal{H}_\tau(\cdot)$ is the soft-thresholding operator with a threshold τ.

Algorithm 1 NCSR-Based HS and MS Image Fusion

Input : $\mathbf{y}_H$ and $\mathbf{y}_M$
Output : High-resolution HS image $\hat{\mathbf{x}}$
Initialization: $\hat{\mathbf{x}} = bicubic(\mathbf{y}_H)$, $\mathbf{x}_M^{(0)} = \mathbf{y}_M$, $\mathbf{x}_H^{(0)} = \mathbf{y}_H$

for $t = 1, 2, ..., T$ **do**
  Update spatial dictionaries $\mathbf{V}_k^{(t)}$ via K-means and PCA from $\mathbf{x}_M^{(t)}$
  Update spectral dictionary $\mathbf{U}^{(t)}$ via PCA from $\mathbf{x}_H^{(t)}$
  Update adaptive spatial-spectral dictionary $\bar{\mathbf{\Phi}}_k^{(t)}$ via (8)
  for $q = 1, 2, ..., Q$ **do**
    $\hat{\mathbf{x}}^{(q+1/2)} = \hat{\mathbf{x}}^{(q)} + \delta \mathbf{B}^T \left(\mathbf{y}_H - \mathbf{B}\hat{\mathbf{x}}^{(q)}\right) + \delta \mathbf{R}^T \left(\mathbf{y}_M - \mathbf{R}\hat{\mathbf{x}}^{(q)}\right)$ where δ is a predefined constant
    $\hat{\boldsymbol{\alpha}}_y^{(q)} = \left[\bar{\mathbf{\Phi}}_{k,1}^{(t)} \mathbf{W}_1 \hat{\mathbf{x}}^{(q+1/2)}, ..., \bar{\mathbf{\Phi}}_{k,npat}^{(t)} \mathbf{W}_{npat} \hat{\mathbf{x}}^{(q+1/2)}\right]$ where $\bar{\mathbf{\Phi}}_{k,m}^{(t)}$ is the dictionary associated to the patch $\mathbf{W}_m \hat{\mathbf{x}}^{q+1/2}$ and $npat$ denote the number of patches extracted from image
    Compute $\hat{\boldsymbol{\alpha}}_{y,i}^{(q+1)}$ using (9)
    if $mod(q, J)$ **then**
      Update $\boldsymbol{\beta}_i$ using (4)
    end if
    Update image: $\hat{\mathbf{x}}^{(q+1)} = \bar{\mathbf{\Phi}}_k \circ \boldsymbol{\alpha}_y^{(q+1)}$ using (2)
  end for
  Update HS and MS images from $\hat{\mathbf{x}}^{(t)} = \hat{\mathbf{x}}^{(q+1)}$: $\mathbf{x}_H^{(t+1)} = \mathbf{P}\mathbf{B}\hat{\mathbf{x}}^{(t)}$, $\mathbf{x}_M^{(t+1)} = \mathbf{R}\hat{\mathbf{x}}^{(t)}$
end for

V. Experimental Results

This section studies the performance of the proposed image fusion method for two different data sets with available ground truth. We conduct experiments on the images: Pavia data set with dimensions $128 \times 128 \times 93$ 1and the Moffett data set with dimensions $128 \times 128 \times 103$. Each target image was degraded to generate the HS and MS images. The HS image is estimated from the target image by applying a 5×5 Gaussian spatial filter on each image band followed by a downsampling bidirectional with a scale factor of $d = 4$. The high spatial and low spectral resolution MS image with $L_M = 4$ was generated for both data set by a filtering using the LANDSAT-like spectral responses as in [6]. Additionally, the HS and MS images are contaminated with zero-mean additive Gaussian noises. The last 50 bands of both data set are perturbed by a SNR$= 30$[dB] and the bands remaining by a SNR$= 35$[dB]. The basic parameters were established as follows: patch size $= 6 \times 6$, number of clusters $K = 60$, number of neighboring cubic patches to estimate β_i (i.e., $|\mathbf{\Omega}_i|$) is equal to 12, search area size of similar patches 50×50, thresholding parameter $\tau = 1.841$ and parameter δ initially set in 1.25 and then adaptively adjusted by reducing its value along the iterations.The proposed method is compared against two fusion approaches presented in [6] and [7] referred to as (HMIF-SR) and (HySure), respectively.

The HS and MS distorted images and the fusion results for the Pavia and Moffett data set are shown in Figs. 1 and 2 respectively. The reconstructed images using the proposed fusion strategy are visually very close to the original image. Furthermore, the qualitative comparison between the evaluated fusion methods shows a better quality for fused images from the proposed method. Note that, the fused images from adversary methods exhibit further spatial degradation of the original image for both Pavia data set and Moffett data set.

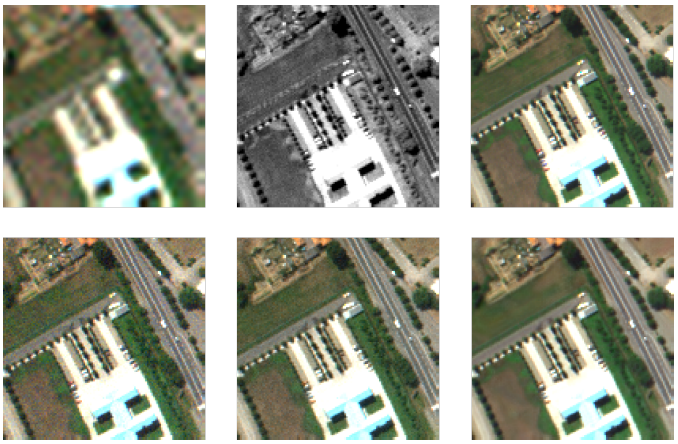

Fig. 1. Fusion results (Pavia data set). (Top 1) HS Pavia image. (Top 2) MS Pavia image. (Top 3) Original Pavia image. (Bottom 1) HySure. (Bottom 2) HMIF-SR. (Bottom 3) Proposed method.

Quantitative results reported in Table I and II show a gain in favor of the proposed method in therms of the RMSE (Root-Mean-Square error), UIQI (Universal Image Quality Index), SAM (Spectral Angle Mapper), ERGAS (Relative Dimensionless Global Error in Synthesis), PSNR (Peak Signal-to-Noise Ratio) and DD (Degree of Distortion) image fusion metrics (see [6] for more details).

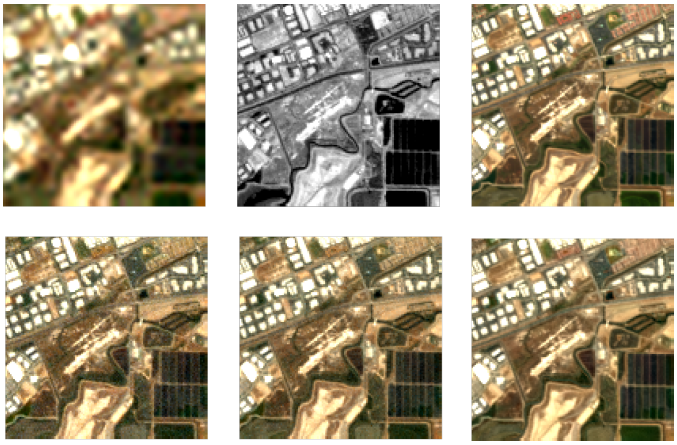

Fig. 2. Fusion results (Moffett data set). (Top 1) HS Moffett image. (Top 2) MS Moffett image. (Top 3) Original Moffett image. (Bottom 1) HySure. (Bottom 2) HMIF-SR. (Bottom 3) Proposed method.

TABLE I
Performance of the HySure [7], HMIF-SR [6] and proposed HS and MS image fusion methods on the Pavia data set: RMSE(10^{-2}), UIQI, SAM [degrees], ERGAS, PSNR [dB] , DD (10^{-3})

Methods	RMSE	UIQI	SAM	ERGAS	PSNR	DD
HySure	1.511	0.978	2.682	1.313	36.214	11.131
HMIF-SR	0.947	0.991	1.495	0.847	39.684	7.011
Proposed	**0.863**	**0.992**	**1.345**	**0.764**	**40.632**	**6.105**

TABLE II
Performance of the HySure [7], HMIF-SR [6] and proposed HS and MS image fusion methods on the Moffett data set: RMSE(10^{-2}), UIQI, SAM [degrees], ERGAS, PSNR [dB] , DD (10^{-3})

Methods	RMSE	UIQI	SAM	ERGAS	PSNR	DD
HySure	1.246	0.987	3.232	1.486	37.994	9.082
HMIF-SR	0.867	0.993	2.103	1.037	40.578	6.429
Proposed	**0.852**	**0.994**	**1.948**	**0.994**	**41.4**	**6.191**

To better evidence the difference between the fusion results, four spectral signatures depicted in Fig. 3 on the Pavia and Moffett dat set were extracted and compared with the results obtained by the HySure, HMIF-SR and the proposed method. For all cases, the resultant spectral signatures using the proposed method were closer to the ground truth which indicates a better estimation of the spectral distribution of the image when compared with the HySure and HMIF-SR methods.

VI. Conclusion

In this work, we propose an HS and MS image fusion method based on a non-local centralized sparse representation. This model allows us to include the non-local redundancy of natural images in the HS image fusion problem and improve the performance of the sparsity-based fusion approaches. Furthermore, an adaptive spatial-spectral dictionary is constructed

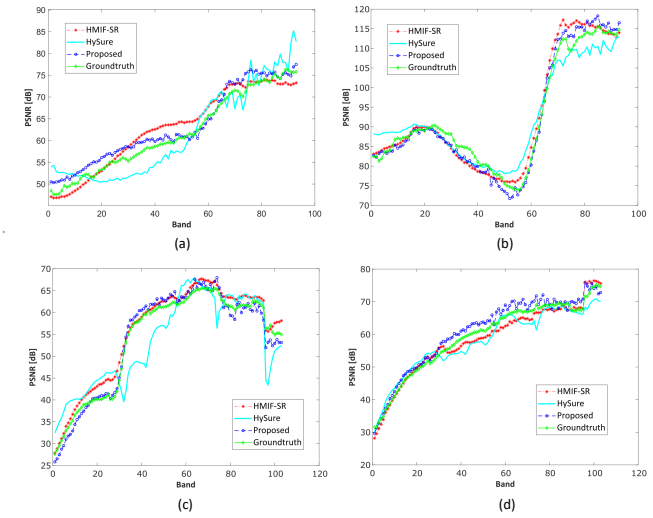

Fig. 3. Spectral responses : (a) spectral signature $(15, 20)$, (b) spectral signature $(120, 105)$ on the Pavia data set. (c) spectral signature $(14, 110)$, (d) spectral signature $(25, 72)$ on the Moffett data set.

exploiting the complementary information in the HS and MS images. This dictionary is composed of K sub-dictionaries associated to spatial clusters of the HS image and thus each cubic patch of the image is sparsely represented in a compact way by choosing the appropriate dictionary. An alternating numerical algorithm was implemented by including two steps, a shrinkage operator to solve the ℓ_1 regularization and a gradient descent algorithm to solve the quadratic problem and preserve the fidelity of the data. Experimental results show that the proposed image fusion model outperforms state-of-the-art sparsity-based methods in terms of the RMSE, UIQI, SAM, ERGAS, PSNR, DD metrics for image fusion.

REFERENCES

[1] M. E. Schaepman, S. L. Ustin, A. J. Plaza, T. H. Painter, J. Verrelst, and S. Liang, "Earth system science related imaging spectroscopy-An assessment," *Remote Sens. Environ.*, vol. 113, no. 123–137, 2009.

[2] S. V. Panasyuk, S. Yang, D. V. Faller, D. Ngo, R. A. Lew, J. E. Freeman, and A. E. Rogers, "Medical hyperspectral imaging to facilitate residual tumor identification during surgery," *Cancer biology & therapy*, vol. 6, no. 3, pp. 439–446, 2007.

[3] N. F. Schreiber, R. Genzel, N. Bouché, G. Cresci, R. Davies, P. Buschkamp, K. Shapiro, L. Tacconi, E. Hicks, S. Genel *et al.*, "The SINS survey: SINFONI integral field spectroscopy of z 2 star-forming galaxies," *The Astrophysical Journal*, vol. 706, no. 2, p. 1364, 2009.

[4] G. A. Shaw and H.-H. K. Burke, "Spectral imaging for remote sensing," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 3–28, 2003.

[5] E. M. Middleton, S. G. Ungar, D. J. Mandl, L. Ong, S. W. Frye, P. E. Campbell, D. R. Landis, J. P. Young, and N. H. Pollack, "The Earth Observing One (EO-1) satellite mission: Over a decade in space," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 6, no. 2, pp. 243–256, 2013.

[6] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens*, vol. 53, no. 7, pp. 3658–3668, 2015.

[7] M. Simões, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, 2015.

[8] 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 Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, 2015.

[9] W. Dong, L. Zhang, and G. Shi, "Centralized sparse representation for image restoration," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 1259–1266.

[10] T. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent developments in total variation image restoration," *Mathematical Models of Computer Vision*, vol. 17, no. 2, 2005.

[11] J. P. Oliveira, J. M. Bioucas-Dias, and M. A. Figueiredo, "Adaptive total variation image deblurring: a majorization–minimization approach," *Signal processing*, vol. 89, no. 9, pp. 1683–1693, 2009.

[12] E. Vargas, Ó. Espitia, H. Arguello, and J.-Y. Tourneret, "Spectral image fusion from compressive measurements," *IEEE Trans. on Image Process.*, vol. 28, no. 5, pp. 2271–2282, 2019.

[13] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE transactions on image processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[14] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1838–1857, 2011.

[15] S. Mallat, *A wavelet tour of signal processing*. Elsevier, 1999.

[16] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[17] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[18] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[19] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, 2010.

[20] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM review*, vol. 51, no. 1, pp. 34–81, 2009.

[21] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *2009 IEEE 12th International Conference on Computer Vision (ICCV)*. IEEE, 2009, pp. 2272–2279.

[22] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1620–1630, 2013.

[23] Y. Fu, A. Lam, I. Sato, and Y. Sato, "Adaptive spatial-spectral dictionary learning for hyperspectral image restoration," *International Journal of Computer Vision*, vol. 122, no. 2, pp. 228–245, 2017.

[24] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[25] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5344–5353.

[26] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tourneret, M. Chen, and S. Godsill, "Multiband image fusion based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens*, vol. 54, no. 12, pp. 7236–7249, 2016.

[27] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multi-band images," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117–1127, 2015.

[28] I. Tosic and P. Frossard, "Dictionary learning," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 27–38, 2011.

[29] H. Yin, S. Li, and L. Fang, "Simultaneous image fusion and super-resolution using sparse representation," *Information Fusion*, vol. 14, no. 3, pp. 229–240, 2013.

[30] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, p. 607, 1996.

[31] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *CVPR 2011*. IEEE, 2011, pp. 193–200.

[32] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *CVPR 2011*. IEEE, 2011, pp. 457–464.

[33] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 57, no. 11, pp. 1413–1457, 2004.