Hyperspectral and multispectral image fusion based on a sparse representation

Qi Wei, José M. Bioucas-Dias, Nicolas Dobigeon, Jean-Yves Tourneret

Hyperspectral and Multispectral Image Fusion Based on a Sparse Representation

Qi Wei, *Student Member, IEEE*, José Bioucas-Dias, *Member, IEEE*,
Nicolas Dobigeon, *Senior Member, IEEE*, and Jean-Yves Tourneret, *Senior Member, IEEE*

***Abstract*—This paper presents a variational-based approach for fusing hyperspectral and multispectral images. The fusion problem is formulated as an inverse problem whose solution is the target image assumed to live in a lower dimensional subspace. A sparse regularization term is carefully designed, relying on a decomposition of the scene on a set of dictionaries. The dictionary atoms and the supports of the corresponding active coding coefficients are learned from the observed images. Then, conditionally on these dictionaries and supports, the fusion problem is solved via alternating optimization with respect to the target image (using the alternating direction method of multipliers) and the coding coefficients. Simulation results demonstrate the efficiency of the proposed algorithm when compared with state-of-the-art fusion methods.**

***Index Terms*—Alternating direction method of multipliers (ADMM), dictionary, hyperspectral (HS) image, image fusion, multispectral (MS) image, sparse representation.**

I. Introduction

FUSION of multisensor images has been explored during recent years and is still a very active research area [2]. A popular fusion problem in remote sensing consists of merging a high spatial resolution panchromatic (PAN) image and a low spatial resolution multispectral (MS) image. Many solutions have been proposed in the literature to solve this problem, which is known as *pansharpening* [2]–[5]. More recently, hyperspectral (HS) imaging acquiring a scene in several hundreds of contiguous spectral bands has opened a new range of relevant applications such as target detection [6] and spectral unmixing [7]. However, while HS sensors provide abundant spectral information, their spatial resolution is generally more limited [8]. To obtain images with good spectral and spatial resolutions, the remote sensing community has been devoting increasing research efforts to the problem of fusing HS with MS or PAN images [9], [10]. From an application point of view, this problem is also important, as motivated by recent national programs, e.g., the Japanese next-generation spaceborne HS image suite, which fuses coregistered MS and HS images acquired over the same scene under the same conditions [11].

The fusion of HS and MS differs from pansharpening since both spatial and spectral information is contained in multiband images. Therefore, a lot of pansharpening methods, such as component substitution [12] and relative spectral contribution [13], are inapplicable or inefficient for the fusion of HS and MS images. Since the fusion problem is generally ill posed, Bayesian inference offers a convenient way to regularize the problem by defining an appropriate generic prior for the scene of interest. Following this strategy, Gaussian or ℓ_2-norm priors have been considered to build various estimators, in the image domain [14]–[16] or in a transformed domain [17]. Recently, the fusion of HS and MS images based on spectral unmixing has also been explored [18], [19].

Sparse representations have received a considerable interest in recent years, exploiting the self-similarity properties of natural images [20]–[23]. Using this property, a sparse constraint has been proposed in [24] and [25] to regularize various ill-posed superresolution and/or fusion problems. The linear decomposition of an image using a few atoms of a redundant dictionary learned from this image (instead of a predefined dictionary, e.g., of wavelets) has recently been used for several problems related to low-level image processing tasks such as denoising [26] and classification [27], demonstrating the ability of sparse representations to model natural images. Learning a dictionary from the image of interest is commonly referred to as dictionary learning (DL). Liu and Boufounos recently proposed to solve the pansharpening problem based on DL [5]. DL has also been investigated to restore HS images [28]. More precisely, a Bayesian scheme was introduced in [28] to learn a dictionary from an HS image, which imposes self-consistency of the dictionary by using Beta-Bernoulli processes. This method provided interesting results at the price of a high computational complexity. Fusing multiple images using a sparse regularization based on the decomposition of these images into high- and low-frequency components was considered in [25]. However, the method developed in [25] required a training data set to learn the dictionaries. The references previously mentioned proposed to solve the corresponding sparse coding problem either by using greedy algorithms such as matching pursuit (MP) and orthogonal MP (OMP) [29] or by relaxing the ℓ_0-norm to an ℓ_1-norm to take advantage of the last absolute shrinkage and selection operator [30].

In this paper, we propose to fuse HS and MS images within a constrained optimization framework, by incorporating a sparse regularization using dictionaries learned from the observed images. Knowing the trained dictionaries and the corresponding supports of the codes circumvents the difficulties inherent to the sparse coding step. The optimization problem can be then solved by optimizing alternatively with respect to (w.r.t.) the projected target image and the sparse code. The optimization w.r.t. the image is achieved by the split augmented Lagrangian shrinkage algorithm (SALSA) [31], which is an instance of the alternating direction method of multipliers (ADMM). By a suitable choice of variable splittings, SALSA enables a huge nondiagonalizable quadratic problem to be decomposed into a sequence of convolutions and pixel decoupled problems, which can be solved efficiently. The coding step is performed using a standard least square (LS) algorithm, which is possible because the supports have been fixed *a priori*.

This paper is organized as follows. Section II formulates the fusion problem within a constrained optimization framework. Section III presents the proposed sparse regularization and the method used to learn the dictionaries and the code support. The strategy investigated to solve the resulting optimization problem is detailed in Section IV. Simulation results are presented in Section V, whereas conclusions are reported in Section VI.

II. Problem Formulation

A. Notations and Observation Model

This paper considers the fusion of HS and MS images. The HS image is supposed to be a blurred and downsampled version of the target image, whereas the MS image is a spectrally degraded version of the target image. Both images are contaminated by white Gaussian noises. Instead of resorting to the totally vectorized notations used in [14], [15], and [17], the HS and MS images are reshaped band by band to build $m_\lambda \times m$ and $n_\lambda \times n$ matrices, respectively, where m_λ is the number of HS bands, $n_\lambda < m_\lambda$ is the number of MS bands, n is the number of pixels in each band of the MS image, and m is the number of pixels in each band of the HS image. The resulting observation models associated with the HS and MS images can be written as follows [14], [32], [33]:

$$\begin{aligned}\mathbf{Y}_{\mathrm{H}} &= \mathbf{XBS} + \mathbf{N}_{\mathrm{H}} \\ \mathbf{Y}_{\mathrm{M}} &= \mathbf{RX} + \mathbf{N}_{\mathrm{M}}\end{aligned} \tag{1}$$

where

- $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] \in \mathbb{R}^{m_\lambda \times n}$ is the full resolution target image with m_λ bands and n pixels;
- $\mathbf{Y}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda \times m}$ and $\mathbf{Y}_{\mathrm{M}} \in \mathbb{R}^{n_\lambda \times n}$ are the observed HS and MS images, respectively;
- $\mathbf{B} \in \mathbb{R}^{n \times n}$ is a cyclic convolution operator acting on the bands;
- $\mathbf{S} \in \mathbb{R}^{n \times m}$ is a downsampling matrix (with downsampling factor denoted by d);
- $\mathbf{R} \in \mathbb{R}^{n_\lambda \times m_\lambda}$ is the spectral response of the MS sensor;
- $\mathbf{N}_{\mathrm{H}}$ and $\mathbf{N}_{\mathrm{M}}$ are the HS and MS noises, respectively.

Note that $\mathbf{B}$ is a sparse block Toeplitz matrix for a symmetric convolution kernel, and $m = n/d^2$, where d is an integer standing for the downsampling factor. Each column of the noise matrices $\mathbf{N}_{\mathrm{H}} = [\mathbf{n}_{\mathrm{H},1}, \ldots \mathbf{n}_{\mathrm{H},m}]$ and $\mathbf{N}_{\mathrm{M}} = [\mathbf{n}_{\mathrm{M},1}, \ldots \mathbf{n}_{\mathrm{M},n}]$ is assumed to be a band-dependent Gaussian vector, i.e., $\mathbf{n}_{\mathrm{H},i} \sim \mathcal{N}(\mathbf{0}_{m_\lambda}, \mathbf{\Lambda}_{\mathrm{H}})(i = 1, \ldots, m)$ and $\mathbf{n}_{\mathrm{M},i} \sim \mathcal{N}(\mathbf{0}_{n_\lambda}, \mathbf{\Lambda}_{\mathrm{M}})(i = 1, \ldots, n)$, where $\mathbf{0}_a$ is the $a \times 1$ vector of zeros, and $\mathbf{\Lambda}_{\mathrm{H}} = \mathrm{diag}(s^2_{\mathrm{H},1}, \ldots, s^2_{\mathrm{H},m_\lambda}) \in \mathbb{R}^{m_\lambda \times m_\lambda}$ and $\mathbf{\Lambda}_{\mathrm{M}} = \mathrm{diag}(s^2_{\mathrm{M},1}, \ldots, s^2_{\mathrm{M},n_\lambda}) \in \mathbb{R}^{n_\lambda \times n_\lambda}$ are diagonal matrices. Note that the Gaussian noise assumption used in this paper is quite popular in image processing [34]–[36] as it facilitates the formulation of the likelihood and the associated optimization algorithms. By denoting the Frobenius norm as $\|\cdot\|_F$, the signal-to-noise ratios (SNRs) of each band in the two images (expressed in decibels) are defined as

$$\begin{aligned}\mathrm{SNR}_{\mathrm{H},i} &= 10 \log \left(\frac{\|(\mathbf{XBS})_i\|_F^2}{s^2_{\mathrm{H},i}} \right), \quad i = 1, \ldots, m_\lambda \\ \mathrm{SNR}_{\mathrm{M},j} &= 10 \log \left(\frac{\|(\mathbf{RX})_j\|_F^2}{s^2_{\mathrm{M},j}} \right), \quad j = 1, \ldots, n_\lambda.\end{aligned}$$

B. Subspace Learning

The unknown image is $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$, where $\boldsymbol{x}_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,m_\lambda}]^T$ is the $m_\lambda \times 1$ vector corresponding to the ith spatial location (with $i = 1, \ldots, n$). As the bands of the HS data are generally spectrally dependent, the HS vector $\boldsymbol{x}_i$ usually lives in a subspace whose dimension is much smaller than the number of bands m_λ [37], [38], i.e.,

$$\boldsymbol{x}_i = \mathbf{H}\boldsymbol{u}_i \tag{2}$$

where $\boldsymbol{u}_i$ is the projection of the vector $\boldsymbol{x}_i$ onto the subspace spanned by the columns of $\mathbf{H} \in \mathbb{R}^{m_\lambda \times \widetilde{m}_\lambda}$ ($\mathbf{H}$ is an orthogonal matrix such that $\mathbf{H}^T\mathbf{H} = \mathbf{I}_{\widetilde{m}_\lambda}$). Using the notation $\mathbf{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n]$, we have $\mathbf{X} = \mathbf{HU}$, where $\mathbf{U} \in \mathbb{R}^{\widetilde{m}_\lambda \times n}$. Moreover, $\mathbf{U} = \mathbf{H}^T\mathbf{X}$ since $\mathbf{H}$ is an orthogonal matrix. In this case, the fusion problem (1) can be reformulated as estimating the unknown matrix $\mathbf{U}$ from the following observation equations:

$$\begin{aligned}\mathbf{Y}_{\mathrm{H}} &= \mathbf{HUBS} + \mathbf{N}_{\mathrm{H}} \\ \mathbf{Y}_{\mathrm{M}} &= \mathbf{RHU} + \mathbf{N}_{\mathrm{M}}.\end{aligned} \tag{3}$$

The dimension of the subspace $\widetilde{m}_\lambda$ is generally much smaller than the number of HS bands, i.e., $\widetilde{m}_\lambda \ll m_\lambda$. As a consequence, inferring in the subspace $\mathbb{R}^{\widetilde{m}_\lambda \times 1}$ greatly decreases the computational burden of the fusion algorithm. Another motivation for working in the subspace associated with $\mathbf{U}$ is to bypass the possible matrix singularity caused by the spectral dependence of the HS data. Note that each column of the matrix $\mathbf{H}$ can be interpreted as a basis of the subspace of interest. In this paper, the matrix $\mathbf{H}$ has been determined from a principal component analysis (PCA) of the HS data $\mathbf{Y}_{\mathrm{H}} = [\mathbf{y}_{\mathrm{H},1}, \ldots, \mathbf{y}_{\mathrm{H},m}]$ (see step 7 of Algorithm 1). Note that, instead of modifying the principal components directly as in the substitution-based method [39], [40], the PCA is only employed to learn the subspace where the fusion problem is solved.

Algorithm 1: Fusion of HS and MS based on a sparse representation

Input: $\mathbf{Y}_{\mathrm{H}}$, $\mathbf{Y}_{\mathrm{M}}$, $\mathrm{SNR}_{\mathrm{H}}$, $\mathrm{SNR}_{\mathrm{M}}$, $\widetilde{m}_{\lambda}$, $\mathbf{R}$, $\mathbf{B}$, $\mathbf{S}$, K

% *Propose a rough estimation* $\tilde{\mathbf{U}}$ *of* $\mathbf{U}$

Compute $\tilde{\mathbf{U}} \triangleq \hat{\boldsymbol{\mu}}_{\mathbf{U}|\mathbf{Y}_{\mathrm{M}}}$ following the method in [14] ;

for $i = 1$ **to** $\widetilde{m}_{\lambda}$ **do**

  % *Online dictionary learning*

  $\bar{\mathbf{D}}_i \leftarrow \mathsf{ODL}(\tilde{\mathbf{U}}_i)$;

  % *Sparse coding*

  $\bar{\mathbf{A}}_i \leftarrow \mathsf{OMP}(\bar{\mathbf{D}}_i, \tilde{\mathbf{U}}_i, K)$;

  % *Identify the supports*

  Set $\bar{\boldsymbol{\Omega}}_i = \left\{(j,k) | \bar{\mathbf{A}}_i(j,k) \neq 0\right\}$;

end

% *Identify the hyperspectral subspace*

$\hat{\mathbf{H}} \leftarrow \mathsf{PCA}(\mathbf{Y}_{\mathrm{H}}, \widetilde{m}_{\lambda})$;

% *Start alternate optimization*

for $t = 1, 2, \ldots$ **to** *stopping rule* **do**

  % *Optimize w.r.t.* $\mathbf{U}$ *using SALSA (see Algo. 2)*

  $\hat{\mathbf{U}}^{(t)} \in \{\mathbf{U} | \, L(\mathbf{U}, \hat{\mathbf{A}}^{(t-1)}) \leq L(\hat{\mathbf{U}}^{(t-1)}, \hat{\mathbf{A}}^{(t-1)})\}$;

  % *Optimize w.r.t.* $\mathbf{A}$ *(LS regression)*

  $\hat{\mathbf{A}}^{(t)} \in \{\mathbf{A} | \, L(\hat{\mathbf{U}}^{(t)}, \mathbf{A}) \leq L(\hat{\mathbf{U}}^{(t)}, \hat{\mathbf{A}}^{(t-1)})\}$;

end

Set $\hat{\mathbf{X}} = \hat{\mathbf{H}}\hat{\mathbf{U}}$;

Output: $\hat{\mathbf{X}}$ (high resolution HS image)

III. Proposed Fusion Rule for MS and HS Images

A. Ill-Posed Inverse Problem

As shown in (3), recovering the projected high spectral and high spatial resolution image $\mathbf{U}$ from the observations $\mathbf{Y}_{\mathrm{H}}$ and $\mathbf{Y}_{\mathrm{M}}$ is a linear inverse problem (LIP) [31]. In most single-image restoration problems (using either $\mathbf{Y}_{\mathrm{H}}$ or $\mathbf{Y}_{\mathrm{M}}$), this inverse problem is ill posed or underconstrained [24], which requires regularization or prior information (in the Bayesian terminology). However, for multisource image fusion, the inverse problem can be ill posed or well posed, depending on the dimension of the subspace and the number of spectral bands. If the matrix $\mathbf{RH}$ has full column rank and is well conditioned, which is seldom the case, the estimation of $\mathbf{U}$ according to (3) is an overdetermined problem instead of an underdetermined problem [41]. In this case, it is redundant to introduce regularizations. Conversely, if there are fewer MS bands than the subspace dimension $\widetilde{m}_{\lambda}$ (e.g., the MS image degrades to a PAN image), the matrix $\mathbf{RH}$ cannot have full column rank, which means that the fusion problem is an ill-posed LIP. In this paper, we focus on the underdetermined case. Note, however, that the overdetermined problem can be viewed as a special case with a regularization term set to zero. Another motivation for studying the underdetermined problem is that it includes an archetypal fusion task referred to as pansharpening [2].

Using (3), the distributions of $\mathbf{Y}_{\mathrm{H}}$ and $\mathbf{Y}_{\mathrm{M}}$ are

$$\begin{aligned} \mathbf{Y}_{\mathrm{H}}|\mathbf{U} &\sim \mathcal{MN}_{m_{\lambda},m}(\mathbf{HUBS}, \boldsymbol{\Lambda}_{\mathrm{H}}, \mathbf{I}_m) \\ \mathbf{Y}_{\mathrm{M}}|\mathbf{U} &\sim \mathcal{MN}_{n_{\lambda},n}(\mathbf{RHU}, \boldsymbol{\Lambda}_{\mathrm{M}}, \mathbf{I}_n) \end{aligned} \tag{4}$$

where $\mathcal{MN}$ represents the matrix normal distribution. The probability density function of a matrix normal distribution $\mathcal{MN}(\mathbf{M}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c)$ is defined by

$$p(\mathbf{X}|\mathbf{M}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c) = \frac{\exp\left(-\frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}_c^{-1}(\mathbf{X}-\mathbf{M})^T\boldsymbol{\Sigma}_r^{-1}(\mathbf{X}-\mathbf{M})\right]\right)}{(2\pi)^{np/2}|\boldsymbol{\Sigma}_c|^{n/2}|\boldsymbol{\Sigma}_r|^{p/2}}$$

where $\mathbf{M}$ denotes the mean matrix; and $\boldsymbol{\Sigma}_r$ and $\boldsymbol{\Sigma}_c$ are two matrices denoting row and column covariance matrices.

According to Bayes theorem and using the fact that the noises $\mathbf{N}_H$ and $\mathbf{N}_M$ are independent, the posterior distribution of $\mathbf{U}$ can be written as

$$p\left(\mathbf{U}|\mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}}\right) \propto p\left(\mathbf{Y}_{\mathrm{H}}|\mathbf{U}\right) p\left(\mathbf{Y}_{\mathrm{M}}|\mathbf{U}\right) p\left(\mathbf{U}\right). \tag{5}$$

In this paper, we want to compute the maximum *a posteriori* (MAP) estimator of $\mathbf{U}$ using an optimization framework to solve the fusion problem. Taking the negative logarithm of the posterior distribution, maximizing the posterior distribution w.r.t. $\mathbf{U}$ is equivalent to solving the following minimization problem:

$$\min_{\mathbf{U}} \underbrace{\frac{1}{2}\left\|\boldsymbol{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS})\right\|_F^2}_{\substack{\text{HS data term} \\ \propto \ln p(\mathbf{Y}_{\mathrm{H}}|\mathbf{U})}} + \underbrace{\frac{1}{2}\left\|\boldsymbol{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU})\right\|_F^2}_{\substack{\text{MS data term} \\ \propto \ln p(\mathbf{Y}_{\mathrm{M}}|\mathbf{U})}} + \underbrace{\lambda\phi(\mathbf{U})}_{\substack{\text{regularizer} \\ \propto \ln p(\mathbf{U})}} \tag{6}$$

where the first two terms are associated with the MS and HS images (data fidelity terms), and the last term is a penalty ensuring appropriate regularization. Note that λ is a parameter adjusting the importance of regularization w.r.t. the data fidelity terms. It is also noteworthy that the MAP estimator is equivalent to the minimum mean square error (MMSE) estimator when $\phi(\mathbf{U})$ has a quadratic form, which is the case in our approach.

B. Sparse Regularization

Based on the self-similarity property of natural images, modeling image patches with a sparse representation has been shown to be very effective in many signal processing applications [24], [42], [43]. Instead of incorporating a simple Gaussian prior or smooth regularization for the fusion of HS and MS images [14], [16], [17], a sparse representation is introduced to regularize the fusion problem. More specifically, image patches of the target image projected into a subspace are represented as a sparse linear combination of elements from an appropriately chosen overcomplete dictionary with columns referred to as atoms. In this paper, the atoms of the dictionary are tuned to the input images, leading to much better results than predefined dictionaries. More specifically, the goal of sparse regularization is to represent the patches of the target image as a weighted linear combination of a few elementary basis vectors or atoms, chosen from a learned overcomplete dictionary. The proposed sparse regularization is defined as

$$\phi(\mathbf{U}) = \frac{1}{2}\sum_{i=1}^{\tilde{m}_{\lambda}} \left\|\mathbf{U}_i - \mathcal{P}(\bar{\mathbf{D}}_i\bar{\mathbf{A}}_i)\right\|_F^2 \tag{7}$$

where

- $\mathbf{U}_i \in \mathbb{R}^n$ is the ith band (or row) of $\mathbf{U} \in \mathbb{R}^{\widetilde{m}_\lambda \times n}$, with $i = 1, \ldots, \widetilde{m}_\lambda$;
- $\mathcal{P}(\cdot) : \mathbb{R}^{n_\mathrm{p} \times n_\mathrm{pat}} \mapsto \mathbb{R}^{n \times 1}$ is a linear operator that averages the overlapping patches[1] of each band;
- $\bar{\mathbf{D}}_i \in \mathbb{R}^{n_\mathrm{p} \times n_\mathrm{at}}$ is an overcomplete dictionary whose columns are basis elements of size n_p (corresponding to the size of a patch);
- $\bar{\mathbf{A}}_i \in \mathbb{R}^{n_\mathrm{at} \times n_\mathrm{pat}}$ is the ith band code (n_at is the number of atoms, and n_pat is the number of patches associated with the ith band).

Note that there are $\widetilde{m}_\lambda$ vectors $\mathbf{U}_i \in \mathbb{R}^n$ since the dimension of the HS subspace in which the observed vectors $\boldsymbol{x}_i$ have been projected is $\widetilde{m}_\lambda$. The operation decomposing each band into overlapping patches of size $\sqrt{n_\mathrm{p}} \times \sqrt{n_\mathrm{p}}$ is denoted by $\mathcal{P}^*(\cdot) : \mathbb{R}^{n \times 1} \mapsto \mathbb{R}^{n_\mathrm{p} \times n_\mathrm{pat}}$, which is the adjoint operation of $\mathcal{P}(\cdot)$, i.e., $\mathcal{P}\left[\mathcal{P}^*(\mathbf{X})\right] = \mathbf{X}$.

C. Dictionary Learning

The DL strategy advocated in this paper consists of learning the dictionaries $\bar{\mathbf{D}}_i$ and an associated sparse code $\bar{\mathbf{A}}_i$ for each band of a rough estimation of $\mathbf{U}$ using the observed HS and MS images. A rough estimation of $\mathbf{U}$, referred as $\tilde{\mathbf{U}}$, is constructed using the MS image $\mathbf{Y}_\mathrm{M}$ and the HS image $\mathbf{Y}_\mathrm{H}$, following the strategy initially studied in [14] (see [45] for more details). Note that other estimation methods might also be used to compute a rough estimation of $\mathbf{U}$ (see step 1 in Algorithm 1). Then, each band $\tilde{\mathbf{U}}_i$ of $\tilde{\mathbf{U}}$ is decomposed into n_pat overlapping patches of size $\sqrt{n_\mathrm{p}} \times \sqrt{n_\mathrm{p}}$ forming a patch matrix $\mathcal{P}^*(\tilde{\mathbf{U}}_i) \in \mathbb{R}^{n_\mathrm{p} \times n_\mathrm{pat}}$.

Many DL methods have been studied in the recent literature. These methods are, for instance, based on K-SVD [46], online DL (ODL) [22], or Bayesian learning [28]. In this paper, we propose to learn the set $\bar{\mathbf{D}} \triangleq [\bar{\mathbf{D}}_1, \ldots, \bar{\mathbf{D}}_{\widetilde{m}_\lambda}]$ of overcomplete dictionaries using ODL since it is effective from a computational point of view and has empirically demonstrated to provide more relevant representations. More specifically, the dictionary $\mathbf{D}_i$ associated with the band $\mathbf{U}_i$ is trained by solving the following optimization problem (see step 3 in Algorithm 1):

$$\{\bar{\mathbf{D}}_i, \tilde{\mathbf{A}}_i\} = \underset{\mathbf{D}_i, \mathbf{A}_i}{\arg\min} \frac{1}{2} \left[\left\| \mathcal{P}^*(\tilde{\mathbf{U}}_i) - \mathbf{D}_i \mathbf{A}_i \right\|_F^2 + \mu \|\mathbf{A}_i\|_1 \right]. \tag{8}$$

Then, to provide a more compact representation, we propose to reestimate the sparse code

$$\bar{\mathbf{A}}_i = \underset{\mathbf{A}_i}{\arg\min} \frac{1}{2} \left\| \mathcal{P}^*(\tilde{\mathbf{U}}_i) - \bar{\mathbf{D}}_i \mathbf{A}_i \right\|_F^2, \text{ s.t. } \|\mathbf{A}_i\|_0 \leq K \tag{9}$$

where s.t. stands for "subject to" and K is a given maximum number of atoms, for each patch of $\mathbf{U}_i$. This ℓ_0-norm constrained regression problem can be addressed using greedy algorithms, e.g., OMP. Generally, the maximum number of atoms K is set much smaller than the number of atoms in the dictionary, i.e., $K \ll n_\mathrm{at}$. The positions of the nonzero elements of the code $\bar{\mathbf{A}}_i$, namely, the supports denoted by $\bar{\boldsymbol{\Omega}}_i \triangleq \{(j,k) | \bar{\mathbf{A}}_i(j,k) \neq 0\}$, are also identified (see steps 4 and 5 in Algorithm 1).

[1]Note that the overlapping decomposition adopted here is to prevent block artifacts [44].

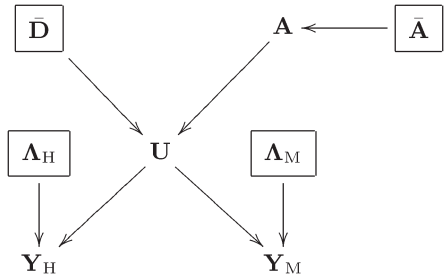

Fig. 1. Directed acyclic graph for the data, parameters, and hyperparameters (the fixed parameters appear in boxes).

D. Including the Sparse Code Into the Estimation Framework

Since the regularization term (7) exhibits separable terms w.r.t. each image $\mathbf{U}_i$ in band i, it can be easily interpreted in a Bayesian framework as the joint prior distribution of the images $\mathbf{U}_i$ ($i = 1, \ldots, \widetilde{m}_\lambda$) assumed to be *a priori* independent, where each marginal prior $p(\mathbf{U}_i)$ is a Gaussian distribution with mean $\mathcal{P}(\bar{\mathbf{D}}_i \bar{\mathbf{A}}_i)$. More formally, by denoting $\bar{\mathbf{A}} \triangleq [\bar{\mathbf{A}}_1, \ldots, \bar{\mathbf{A}}_{\widetilde{m}_\lambda}]$, the prior distribution for $\mathbf{U}$ associated with the regularization (7) can be written as

$$p\left(\mathbf{U} | \bar{\mathbf{D}}, \bar{\mathbf{A}}\right) = \prod_{i=1}^{\widetilde{m}_\lambda} p\left(\mathbf{U}_i | \bar{\mathbf{D}}_i, \bar{\mathbf{A}}_i\right). \tag{10}$$

In a standard approach, the hyperparameters $\bar{\mathbf{D}}$ and $\bar{\mathbf{A}}$ can be *a priori* fixed, e.g., based on the DL step detailed in the previous section. However, this choice can drastically impact the accuracy of the representation and therefore the relevance of the regularization term. Inspired by hierarchical models frequently encountered in Bayesian inference [47], we propose to add a second level in the Bayesian paradigm by fixing the dictionaries $\bar{\mathbf{D}}$ and the set of supports $\bar{\boldsymbol{\Omega}} \triangleq \{\bar{\boldsymbol{\Omega}}_1, \ldots, \bar{\boldsymbol{\Omega}}_{\widetilde{m}_\lambda}\}$, but by including the code $\mathbf{A}$ within the estimation process. The associated joint prior can be written as follows:

$$p\left(\mathbf{U}, \mathbf{A} | \bar{\mathbf{D}}, \bar{\mathbf{A}}\right) = \prod_{i=1}^{\widetilde{m}_\lambda} p\left(\mathbf{U}_i | \bar{\mathbf{D}}_i, \mathbf{A}_i\right) p\left(\mathbf{A}_i | \bar{\mathbf{A}}_i\right) \tag{11}$$

where $\bar{\boldsymbol{\Omega}}$ is derived from $\bar{\mathbf{A}}$. Therefore, the regularization term (7) reduces to

$$\begin{aligned} &\phi(\mathbf{U}, \mathbf{A}) = \frac{1}{2} \sum_{i=1}^{\widetilde{m}_\lambda} \left\| \mathbf{U}_i - \mathcal{P}(\bar{\mathbf{D}}_i \mathbf{A}_i) \right\|_F^2 = \frac{1}{2} \|\mathbf{U} - \bar{\mathbf{U}}\|_F^2 \\ &\text{s.t. } \{\mathbf{A}_{i, \setminus \bar{\boldsymbol{\Omega}}_i} = 0\}_{i=1}^{\widetilde{m}_\lambda} \end{aligned} \tag{12}$$

where $\bar{\mathbf{U}} \triangleq [\mathcal{P}(\bar{\mathbf{D}}_1 \mathbf{A}_1), \ldots, \mathcal{P}(\bar{\mathbf{D}}_{\widetilde{m}_\lambda} \mathbf{A}_{\widetilde{m}_\lambda})]$, and $\mathbf{A}_{i, \setminus \bar{\boldsymbol{\Omega}}_i} = \{\mathbf{A}_i(j,k) \mid (j,k) \notin \bar{\boldsymbol{\Omega}}_i\}$. It is worthy to note that 1) the regularization term in (12) is still separable w.r.t. each band $\mathbf{U}_i$ and 2) the optimization of (12) w.r.t. $\mathbf{A}_i$ reduces to an ℓ_2-norm optimization task w.r.t. the nonzero elements in $\mathbf{A}_i$, which can be solved easily. The hierarchical structure of the observed data, parameters, and hyperparameters is summarized in Fig. 1.

Finally, substituting (12) into (6), the optimization problem to be solved can be expressed as follows:

$$\min_{\mathbf{U},\mathbf{A}} L(\mathbf{U},\mathbf{A}) \triangleq \frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}}-\mathbf{HUBS})\right\|_F^2 + \frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}}-\mathbf{RHU})\right\|_F^2+\frac{\lambda}{2}\|\mathbf{U}-\bar{\mathbf{U}}\|_F^2$$
$$\text{s.t. } \left\{\mathbf{A}_{i,\backslash\bar{\mathbf{\Omega}}_i}=0\right\}_{i=1}^{\widetilde{m}_\lambda}. \quad (13)$$

Note that the set of constraints $\left\{\mathbf{A}_{i,\backslash\bar{\mathbf{\Omega}}_i}=0\right\}_{i=1}^{\widetilde{m}_\lambda}$ could have been removed. In this case, to ensure a sparse representation of $\mathbf{U}_i$ ($i=1,\ldots,\widetilde{m}_\lambda$), sparse constraints on the codes $\mathbf{A}_i$ ($i=1,\ldots,\widetilde{m}_\lambda$), such as $\{\|\mathbf{A}_i\|_0<K\}_{i=1}^{\widetilde{m}_\lambda}$ or sparsity promoting penalties, e.g., $\sum_{i=1}^{\widetilde{m}_\lambda}\|\mathbf{A}_i\|_1$, should have been included into the object function (13). This would have resulted in a much more computationally intensive algorithm.

IV. Alternate Optimization

Once $\bar{\mathbf{D}}$, $\bar{\mathbf{\Omega}}$, and $\mathbf{H}$ have been learned from the observed data, (13) reduces to a standard constrained quadratic optimization problem w.r.t. $\mathbf{U}$ and $\mathbf{A}$. However, this problem is difficult to solve due to its large dimension and the fact that the operators $\mathbf{H}(\cdot)\mathbf{BD}$ and $\mathcal{P}(\cdot)$ cannot be easily diagonalized. To cope with this difficulty, we propose an optimization technique that alternates optimization w.r.t. $\mathbf{U}$ and $\mathbf{A}$, which is a simple version of a block coordinate descent algorithm.

The optimization w.r.t. $\mathbf{U}$ conditional on $\mathbf{A}$ (or equivalent on $\bar{\mathbf{U}}$) can be achieved efficiently with the ADMM [48] whose convergence has been proved in the convex case. The optimization w.r.t. $\mathbf{A}$ with the support constraint $\mathbf{A}_{i,\backslash\bar{\mathbf{\Omega}}_i}=0$ ($i=1,2,\ldots,\widetilde{m}_\lambda$) conditional on $\mathbf{U}$ is an LS regression problem for the nonzero elements of $\mathbf{A}$, which can be solved easily. The resulting scheme, including learning $\bar{\mathbf{D}}$, $\bar{\mathbf{\Omega}}$, and $\mathbf{H}$, is detailed in Algorithm 1. The alternating ADMM and LS steps are detailed in what follows.

A. ADMM Step

The function to be minimized w.r.t. $\mathbf{U}$ conditionally on $\mathbf{A}$ (or $\bar{\mathbf{U}}$) is

$$\frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}}-\mathbf{HUBS})\right\|_F^2 + \frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}}-\mathbf{RHU})\right\|_F^2+\frac{\lambda}{2}\left\|\mathbf{U}-\bar{\mathbf{U}}\right\|_F^2. \quad (14)$$

By introducing the splittings $\mathbf{V}_1=\mathbf{UB}$, $\mathbf{V}_2=\mathbf{U}$, and $\mathbf{V}_3=\mathbf{U}$ and the respective scaled Lagrange multipliers $\mathbf{G}_1,\mathbf{G}_2$, and $\mathbf{G}_3$, the augmented Lagrangian associated with the optimization of $\mathbf{U}$ can be written as

$$\begin{aligned}&L(\mathbf{U},\mathbf{V}_1,\mathbf{V}_2,\mathbf{V}_3,\mathbf{G}_1,\mathbf{G}_2,\mathbf{G}_3)\\ &=\frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}}-\mathbf{HV}_1\mathbf{S})\right\|_F^2+\frac{\mu}{2}\|\mathbf{UB}-\mathbf{V}_1-\mathbf{G}_1\|_F^2\\ &+\frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}}-\mathbf{RHV}_2)\right\|_F^2+\frac{\mu}{2}\|\mathbf{U}-\mathbf{V}_2-\mathbf{G}_2\|_F^2\\ &+\frac{1}{2}\|\bar{\mathbf{U}}-\mathbf{V}_3\|_F^2+\frac{\mu}{2}\|\mathbf{U}-\mathbf{V}_3-\mathbf{G}_3\|_F^2.\end{aligned}$$

Algorithm 2: SALSA sub-iterations

Input: $\hat{\mathbf{U}}^{(t)}$, $\bar{\mathbf{D}}$, $\hat{\mathbf{A}}^{(t)}$, $\mathbf{Y}_{\mathrm{H}}$, $\mathbf{Y}_{\mathrm{M}}$, $\mathrm{SNR}_{\mathrm{H}}$, $\mathrm{SNR}_{\mathrm{M}}$, $\mathbf{H}$, $\mathbf{R}$, $\mathbf{B}$, $\mathbf{S}$, λ and μ (SALSA parameter)

1 Set $\bar{\mathbf{U}}=\left[\mathcal{P}\left(\bar{\mathbf{D}}_1\hat{\mathbf{A}}_1^{(t)}\right),\ldots,\mathcal{P}\left(\bar{\mathbf{D}}_{\widetilde{m}_\lambda}\hat{\mathbf{A}}_{\widetilde{m}_\lambda}^{(t)}\right)\right]$;

2 Set $\boldsymbol{\delta}\in\{0,1\}^n$ such that $\boldsymbol{\delta}(i)=\begin{cases}1 & \text{if pixel } i \text{ is sampled,}\\ 0 & \text{otherwise;}\end{cases}$

3 **Initialization**: $\mathbf{V}_1^{(0)}$,$\mathbf{V}_2^{(0)}$,$\mathbf{V}_3^{(0)}$, $\mathbf{G}_1^{(0)}$, $\mathbf{G}_2^{(0)}$,$\mathbf{G}_3^{(0)}$;

4 **for** $k=0$ **to** n_{it} **do**

  % *Optimize w.r.t* $\mathbf{U}$

5 $\hat{\mathbf{U}}^{(t,k+1)}\leftarrow \left[(\mathbf{V}_1^{(k)}+\mathbf{G}_1^{(k)})\mathbf{B}^T+(\mathbf{V}_2^{(k)}+\mathbf{G}_2^{(k)})+(\mathbf{V}_3^{(k)}+\mathbf{G}_3^{(k)})\right]\left(\mathbf{BB}^T+2\mathbf{I}\right)^{-1}$;

  % *Optimize* $\mathbf{V}_1$ *(according to down-sampling)*

6 $\boldsymbol{\nu}_1\leftarrow(\hat{\mathbf{U}}^{(t,k+1)}\mathbf{B}-\mathbf{G}_1^{(k)})$;

7 $\mathbf{V}_1^{(k+1)}(:,\boldsymbol{\delta})\leftarrow \left(\mathbf{H}^T\mathbf{\Lambda}_{\mathrm{H}}^{-1}\mathbf{H}+\mu\mathbf{I}\right)^{-1}(\mathbf{H}^T\mathbf{\Lambda}_{\mathrm{H}}^{-1}\mathbf{Y}_{\mathrm{H}}+\boldsymbol{\nu}_1(:,\boldsymbol{\delta}))$;

8 $\mathbf{V}_1^{(k+1)}(:,1-\boldsymbol{\delta})\leftarrow\boldsymbol{\nu}_1(:,1-\boldsymbol{\delta})$;

  % *Optimize w.r.t* $\mathbf{V}_2$

9 $\boldsymbol{\nu}_2\leftarrow(\hat{\mathbf{U}}^{(t,k+1)}-\mathbf{G}_2^{(k)})$;

10 $\mathbf{V}_2^{(k+1)}\leftarrow\left(\mathbf{H}^T\mathbf{R}^T\mathbf{\Lambda}_{\mathrm{M}}^{-1}\mathbf{RH}+\mu\mathbf{I}\right)^{-1}(\mathbf{H}^T\mathbf{R}^T\mathbf{\Lambda}_{\mathrm{M}}^{-1}\mathbf{Y}_{\mathrm{M}}+\mu\boldsymbol{\nu}_2)$;

  % *Optimize w.r.t* $\mathbf{V}_3$

11 $\boldsymbol{\nu}_3\leftarrow(\hat{\mathbf{U}}^{(t,k+1)}-\mathbf{G}_3^{(k)})$;

12 $\mathbf{V}_3^{(k+1)}\leftarrow(\lambda+\mu)^{-1}\left(\lambda\bar{\mathbf{U}}+\mu\boldsymbol{\nu}_3\right)$;

  % *Update Lagrange multipliers*

13 $\mathbf{G}_1^{(k+1)}\leftarrow(-\boldsymbol{\nu}_1+\mathbf{V}_1^{(k+1)})$;

14 $\mathbf{G}_2^{(k+1)}\leftarrow(-\boldsymbol{\nu}_2+\mathbf{V}_2^{(k+1)})$;

15 $\mathbf{G}_3^{(k+1)}\leftarrow(-\boldsymbol{\nu}_3+\mathbf{V}_3^{(k+1)})$;

16 **end**

17 Set $\hat{\mathbf{U}}^{(t+1)}=\hat{\mathbf{U}}^{(t,n_{\mathrm{it}})}$;

Output: $\hat{\mathbf{U}}^{(t+1)}$

The updates of $\mathbf{U},\mathbf{V}_1,\mathbf{V}_2,\mathbf{V}_3,\mathbf{G}_1,\mathbf{G}_2$, and $\mathbf{G}_3$ are obtained with the SALSA algorithm [31], [49], which is an instance of the ADMM algorithm with guaranteed convergence. The SALSA scheme is summarized in Algorithm 2. Note that the optimization w.r.t. $\mathbf{U}$ (step 5) can be efficiently solved in the Fourier domain.

B. Patchwise Sparse Coding

The optimization w.r.t. $\mathbf{A}$ conditional on $\mathbf{U}$ is

$$\hat{\mathbf{A}}_i=\arg\min_{\mathbf{A}_i}\left\|\mathbf{U}_i-\mathcal{P}(\bar{\mathbf{D}}_i\mathbf{A}_i)\right\|_F^2,\quad \text{s.t. } \mathbf{A}_{i,\backslash\bar{\mathbf{\Omega}}_i}=0 \quad (15)$$

where $i=1,\ldots,\widetilde{m}_\lambda$. Since the operator $\mathcal{P}(\cdot)$ is a linear mapping from patches to images and $\mathcal{P}[\mathcal{P}^*(\mathbf{X})]=\mathbf{X}$, the problem (15) can be rewritten as

$$\hat{\mathbf{A}}_i=\arg\min_{\mathbf{A}_i}\left\|\mathcal{P}\left(\mathcal{P}^*(\mathbf{U}_i)-\bar{\mathbf{D}}_i\mathbf{A}_i\right)\right\|_F^2,\ \text{s.t. } \mathbf{A}_{i,\backslash\bar{\mathbf{\Omega}}_i}=0. \quad (16)$$

The solution of (16) can be approximated by solving

$$\hat{\mathbf{A}}_i=\arg\min_{\mathbf{A}_i}\left\|\mathcal{P}^*(\mathbf{U}_i)-\bar{\mathbf{D}}_i\mathbf{A}_i\right\|_F^2,\ \text{s.t. } \mathbf{A}_{i,\backslash\bar{\mathbf{\Omega}}_i}=0. \quad (17)$$

Note that using the suboptimal solution instead of the optimal solution does not affect the convergence of the alternating optimization. Tackling the support constraint consists of only updating the nonzero elements of each column of $\mathbf{A}_i$. The jth

vectorized column of $\mathcal{P}^*(\mathbf{U}_i)$ is denoted by $\mathbf{p}_{i,j}$, the vector composed of the K nonzero elements of the jth column of $\mathbf{A}_i$ is denoted by $\mathbf{a}_{\bar{\mathbf{\Omega}}_i^j}$, and the corresponding column of $\bar{\mathbf{D}}_i$ is denoted by $\bar{\mathbf{D}}_{\bar{\mathbf{\Omega}}_i^j}$. Then, the $\widetilde{m}_\lambda$ problems in (17) reduce to $\widetilde{m}_\lambda \times n_{\text{pat}}$ subproblems

$$\begin{aligned}\hat{\mathbf{a}}_{\bar{\mathbf{\Omega}}_i^j} &= \arg\min_{\mathbf{a}_{\bar{\mathbf{\Omega}}_i^j}} \left\| \mathbf{p}_{i,j} - \bar{\mathbf{D}}_{\bar{\mathbf{\Omega}}_i^j} \mathbf{a}_{\bar{\mathbf{\Omega}}_i^j} \right\|_F^2 \\ &= (\bar{\mathbf{D}}_{\bar{\mathbf{\Omega}}_i^j}^T \bar{\mathbf{D}}_{\bar{\mathbf{\Omega}}_i^j})^{-1} \bar{\mathbf{D}}_{\bar{\mathbf{\Omega}}_i^j}^T \mathbf{p}_{i,j} \\ \text{for} \quad & i = 1, \ldots, \widetilde{m}_\lambda; j = 1, \ldots, n_{\text{pat}} \end{aligned} \tag{18}$$

which can be computed in parallel. The corresponding patch estimate is

$$\begin{aligned}\hat{\mathbf{p}}_{i,j} &\triangleq \mathbf{T}_{i,j} \mathbf{p}_{i,j} \\ \mathbf{T}_{i,j} &= \bar{\mathbf{D}}_{\bar{\mathbf{\Omega}}_i^j} (\bar{\mathbf{D}}_{\bar{\mathbf{\Omega}}_i^j}^T \bar{\mathbf{D}}_{\bar{\mathbf{\Omega}}_i^j})^{-1} \bar{\mathbf{D}}_{\bar{\mathbf{\Omega}}_i^j}^T. \end{aligned}$$

These patches are used to build $\bar{\mathbf{U}}$ (i.e., equivalently, $\mathcal{P}\left(\bar{\mathbf{D}}_i \mathbf{A}_i\right)$) required in the optimization w.r.t. $\mathbf{U}$ (see Section IV-A). Note that $\mathbf{T}_{i,j}$ is a projection operator, and hence is symmetric ($\mathbf{T}_{i,j}^T = \mathbf{T}_{i,j}$) and idempotent ($\mathbf{T}_{i,j}^2 = \mathbf{T}_{i,j}$). Note also that $\mathbf{T}_{i,j}$ needs to be calculated only once, given the learned dictionaries and associated supports.

C. Complexity Analysis

The SALSA algorithm has the order of complexity $\mathcal{O}\left(n_{\text{it}} \widetilde{m}_\lambda n \log(\widetilde{m}_\lambda n)\right)$ [31], where n_{it} is the number of SALSA iterations. The computational complexity of the patch-wise sparse coding is $\mathcal{O}\left(K n_{\text{p}} n_{\text{pat}} \widetilde{m}_\lambda\right)$. Conducting the fusion in a subspace of dimension $\widetilde{m}_\lambda$ instead of working with the initial space of dimension m_λ greatly decreases the complexity of both SALSA and sparse coding steps.

V. Simulation Results on Synthetic Data

This section studies the performance of the proposed sparse representation-based fusion algorithm. The reference image considered here as the high spatial and high spectral image is a $128 \times 128 \times 93$ HS image with a spatial resolution of 1.3 m acquired by the reflective optics system imaging spectrometer optical sensor over the urban area of the University of Pavia, Italy. The flight was operated by the Deutsches Zentrum für Luft- und Raumfahrt (DLR, the German Aerospace Agency) in the framework of the HySens project, managed and sponsored by the European Union. This image was initially composed of 115 bands, which have been reduced to 93 bands after removing the water vapor absorption bands (with spectral range from 0.43 to 0.86 m). It has received a lot of attention in the remote sensing literature [50]–[52]. A composite color image, formed by selecting the red, green, and blue bands of the reference image, is shown in the left panel of Fig. 2.

A. Simulation Scenario

We propose to reconstruct the reference HS image from two lower resolved images. A high spectral low spatial resolution HS image has been constructed by applying a 5×5 Gaussian spatial filter on each band of the reference image and down-sampling every four pixels in both horizontal and vertical directions. In a second step, we have generated a four-band MS image by filtering the reference image with the IKONOS-like reflectance spectral responses depicted in Fig. 3. The HS and MS images are both contaminated by zero-mean additive Gaussian noises. Our simulations have been conducted with $\text{SNR}_{1,\cdot} = 35$ dB for the first 43 bands and $\text{SNR}_{1,\cdot} = 30$ dB for the remaining 50 bands of the HS image. For the MS image, $\text{SNR}_{2,\cdot}$ is 30 dB for all bands. The noise-contaminated HS and MS images are depicted in the middle and right panels in Fig. 2 (the HS image has been interpolated for better visualization).

Fig. 2. (Left) Reference image. (Middle) HS image. (Right) MS image.

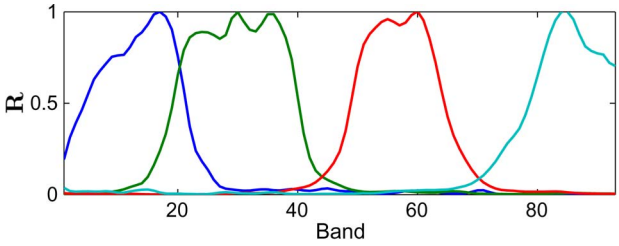

Fig. 3. IKONOS-like spectral responses.

B. Learning the Subspace, Dictionaries, and Code Supports

1) Subspace: To learn the transform matrix $\mathbf{H}$, we used the PCA as in [16]. Note that PCA is a classical dimensionality reduction technique used in HS imagery. The empirical correlation matrix $\mathbf{\Upsilon} = \mathrm{E}[\boldsymbol{x}_i \boldsymbol{x}_i^T]$ of the HS pixel vectors is diagonalized, leading to

$$\mathbf{W}^T \mathbf{\Upsilon} \mathbf{W} = \mathbf{\Gamma} \tag{19}$$

where $\mathbf{W}$ is an $m_\lambda \times m_\lambda$ unitary matrix ($\mathbf{W}^T = \mathbf{W}^{-1}$), and $\mathbf{\Gamma}$ is a diagonal matrix whose diagonal elements are the ordered eigenvalues of $\mathbf{\Upsilon}$ denoted by $d_1 \geq d_2 \geq \cdots \geq d_{m_\lambda}$. The top $\widetilde{m}_\lambda$ components are selected, and the matrix $\mathbf{H}$ is constructed as the eigenvectors associated with the $\widetilde{m}_\lambda$ largest eigenvalues of $\mathbf{\Upsilon}$. In practice, the selection of the number of principal components $\widetilde{m}_\lambda$ depends on how many materials (or endmembers) the target image contains. If the number of truncated principal components is smaller than the dimension of the subspace spanned by the target image vectors, the projection will lead to a loss of information. On the contrary, if the number of principal components is larger than the real dimension, the overfitting problem may arise, leading to a degradation of the fusion performance. As an illustration, the eigenvalues of $\mathbf{\Upsilon}$ for the Pavia image are displayed in Fig. 4. For this example, the $\widetilde{m}_\lambda = 5$ eigenvectors contain 99.9% of the information and have been chosen to build the subspace of interest. A more detailed discussion can be found in [45] with regard to the choice of parameter $\widetilde{m}_\lambda$.

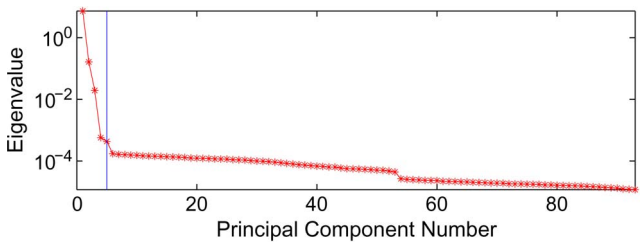

Fig. 4. Eigenvalues of $\boldsymbol{\Upsilon}$ for the Pavia HS image.

2) Dictionaries: As explained before, the target high-resolution image is assumed to live in a lower dimensional subspace. First, a rough estimation of the projected image is obtained with the method proposed in [14]. In a second step, $\widetilde{m}_\lambda = 5$ dictionaries are learned from the rough estimation of the projected image using the ODL method. As $n_{\mathrm{at}} \gg n_{\mathrm{p}}$, the dictionary is overcomplete. There is no unique rule to select the dictionary size n_{p} and the number of atoms n_{at}. However, two limiting cases can be identified.

- The patch reduces to a single pixel, which means $n_{\mathrm{p}} = 1$. In this case, the sparsity is not necessary to be introduced since only one 1-D dictionary atom (which is a constant) is enough to represent any target patch.
- The patch is as large as the whole image, which means that only one atom is needed to represent the image. In this case, the atom is too "specialized" to describe any other image.

More generally, the smaller the patches, the more objects the atoms can approximate. However, too small patches are not efficient to properly capture the textures, edges, etc. With larger patch size, a larger number of atoms are required to guarantee the overcompleteness (which requires larger computation cost). In general, the size of patches is empirically selected. For the ODL algorithm used in this paper, this size has been fixed to $n_{\mathrm{p}} = 6 \times 6$, and the number of atoms is $n_{\mathrm{at}} = 256$. The learned dictionaries for the first three bands of $\tilde{\mathbf{U}}$ are displayed in Fig. 5. This figure shows that the spatial properties of the target image have been captured by the atoms of the dictionaries.

3) Code Supports: Based on the dictionaries learned following the strategy presented in Section V-B2, the codes are re estimated by solving (9) with OMP. Note that the target sparsity K represents the maximum number of atoms used to represent one patch, which also determines the number of nonzero elements of $\mathbf{A}$ estimated jointly with the projected image $\mathbf{U}$. If K is too large, the optimization w.r.t. $\mathbf{U}$ and $\mathbf{A}$ leads to overfitting, which means there are too many parameters to estimate while the sample size is too small. The training supports for the first three bands are displayed in the right column in Fig. 5. The number of rows is 256, which represents the number of atoms in each dictionary $\tilde{\mathbf{D}}_i$ ($i = 1, \ldots, \widetilde{m}_\lambda$). The white dots in the jth column indicate which atoms are used for reconstructing the jth patch ($j = 1, \ldots, n_{\mathrm{pat}}$). The sparsity is clearly observed in this figure. Note that some atoms are frequently used whereas some others are not. The most popular atoms represent spatial details that are quite common in images. The other atoms represent details that are characteristics of specific patches.

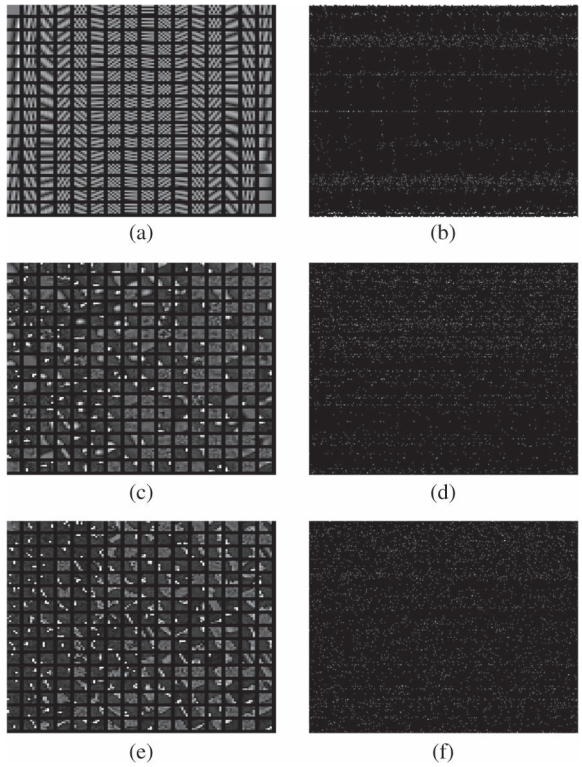

Fig. 5. (Left) Learned dictionaries and (right) corresponding supports. (a) Dictionary for band 1. (b) Support for band 1 (for some patches). (c) Dictionary for band 2. (d) Support for band 2 (for some patches). (e) Dictionary for band 3. (f) Support for band 3 (for some patches).

C. Fusion Quality Metrics

To evaluate the quality of the proposed fusion strategy, several image quality measures have been employed. Referring to [17], we propose to use RMSE, SAM, UIQI, ERGAS, and DD that are defined below.

1) RMSE: The root-mean-square error (RMSE) is a similarity measure between the target image $\mathbf{X}$ and the fused image $\hat{\mathbf{X}}$ defined as

$$\mathrm{RMSE}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{n m_\lambda} \|\mathbf{X} - \hat{\mathbf{X}}\|_F^2.$$

The smaller the RMSE, the better the fusion quality.

2) SAM: The spectral angle mapper (SAM) measures the spectral distortion between the actual and estimated images. The SAM of two spectral vectors $\boldsymbol{x}_n$ and $\hat{\boldsymbol{x}}_n$ is defined as

$$\mathrm{SAM}(\boldsymbol{x}_n, \hat{\boldsymbol{x}}_n) = \arccos\left(\frac{\langle \boldsymbol{x}_n, \hat{\boldsymbol{x}}_n \rangle}{\|\boldsymbol{x}_n\|_2 \|\hat{\boldsymbol{x}}_n\|_2}\right).$$

The overall SAM is finally obtained by averaging the SAMs computed for all image pixels. Note that the value of SAM is expressed in degrees and thus belongs to $(-90, 90]$. The smaller the absolute value of SAM, the less important the spectral distortion.

3) UIQI: The universal image quality index (UIQI) was proposed in [53] for evaluating the similarity between two single-band images. It is related to the correlation, luminance distortion, and contrast distortion of the estimated image w.r.t.

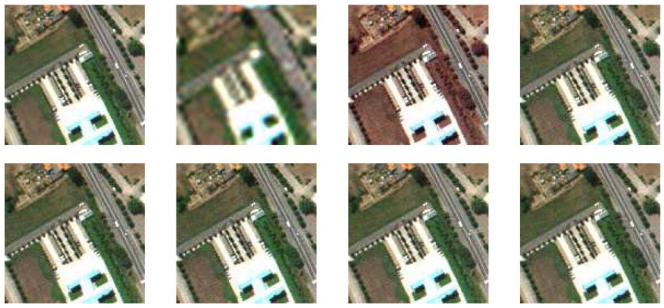

Fig. 6. Pavia data set. (Top 1) Reference. (Top 2) HS. (Top 3) MS. (Top 4) MAP [14]. (Bottom 1) Wavelet MAP [17]. (Bottom 2) Coupled nonnegative matrix factorization (CNMF) fusion [18]. (Bottom 3) MMSE estimator [16]. (Bottom 4) Proposed method.

the reference image. The UIQI between two single-band images $\mathbf{a} = [a_1, a_2, \ldots, a_N]$ and $\hat{\mathbf{a}} = [\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_N]$ is defined as

$$\text{UIQI}(\mathbf{a}, \hat{\mathbf{a}}) = \frac{4\sigma^2_{a\hat{a}}\mu_a\mu_{\hat{a}}}{(\sigma^2_a + \sigma^2_{\hat{a}})\,(\mu^2_a + \mu^2_{\hat{a}})}$$

where $(\mu_a, \mu_{\hat{a}}, \sigma^2_a, \sigma^2_{\hat{a}})$ are the sample means and variances of a and $\hat{a}$, and $\sigma^2_{a\hat{a}}$ is the sample covariance of $(a, \hat{a})$. The range of UIQI is $[-1, 1]$, and $\text{UIQI}(\mathbf{a}, \hat{\mathbf{a}}) = 1$ when $\mathbf{a} = \hat{\mathbf{a}}$. For multiband images, the overall UIQI can be computed by averaging the UIQI computed band by band.

4) ERGAS: The relative dimensionless global error in synthesis (ERGAS) calculates the amount of spectral distortion in the image [54]. This measure of fusion quality is defined as

$$\text{ERGAS} = 100 \times \frac{m}{n}\sqrt{\frac{1}{m_\lambda}\sum_{i=1}^{m_\lambda}\left(\frac{\text{RMSE}(i)}{\mu_i}\right)^2}$$

where m/n is the ratio between the pixel sizes of the MS and HS images, μ_i is the mean of the ith band of the HS image, and m_λ is the number of HS bands. The smaller the ERGAS, the smaller the spectral distortion.

5) DD: The degree of distortion (DD) between two images $\mathbf{X}$ and $\hat{\mathbf{X}}$ is defined as

$$\text{DD}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{nm_\lambda}\left\|\text{vec}(\mathbf{X}) - \text{vec}(\hat{\mathbf{X}})\right\|_1.$$

Note that $\text{vec}(\mathbf{X})$ represents the vectorization of matrix $\mathbf{X}$. The smaller the DD, the better the fusion.

D. Comparison With Other Fusion Methods

This section compares the proposed fusion method with four other state-of-the-art fusion algorithms for MS and HS images [14], [16]–[18]. The parameters used for the proposed fusion algorithm have been specified as follows.

- The regularization parameter used in the SALSA method is $\mu = 0.05/\|\mathbf{N}_\text{H}\|_F$. The selection of this parameter μ is still an open issue even if there are some strategies to tune it to accelerate convergence [31]. According to the convergence theory [55], for any $\mu > 0$, if the minimization of (14) has a solution, for example, $\mathbf{U}^\star$, then the sequence $\{\mathbf{U}^{(t,k)}\}_{k=1}^{\infty}$ converges to $\mathbf{U}^\star$. If the minimization of (14) has no solution, then at least one of the sequences $\{\mathbf{U}^{(t,k)}\}_{k=1}^{\infty}$ or $\{\mathbf{G}^{(t,k)}\}_{k=1}^{\infty}$ diverges. Simulations have shown that the choice of μ does not significantly affect the fusion performance as long as μ is positive.
- The regularization coefficient is $\lambda = 25$. The choice of this parameter will be discussed in Section V-E.

TABLE I
PERFORMANCE OF DIFFERENT MS + HS FUSION METHODS (PAVIA DATA SET): RMSE (IN 10^{-2}), UIQI, SAM (IN DEGREES), ERGAS, DD (IN 10^{-3}), AND TIME (IN SECONDS)

Methods	RMSE	UIQI	SAM	ERGAS	DD	Time
MAP [14]	1.148	0.9875	1.962	1.029	8.666	**3**
Wavelet MAP [17]	1.099	0.9885	1.849	0.994	8.349	75
CNMF [18]	1.119	0.9857	2.039	1.089	9.007	14
HMC [16]	1.011	0.9903	1.653	0.911	7.598	6003
Rough $\tilde{\mathbf{U}}$	1.136	0.9878	1.939	1.019	8.586	\
Proposed	**0.947**	**0.9913**	**1.492**	**0.850**	**7.010**	282

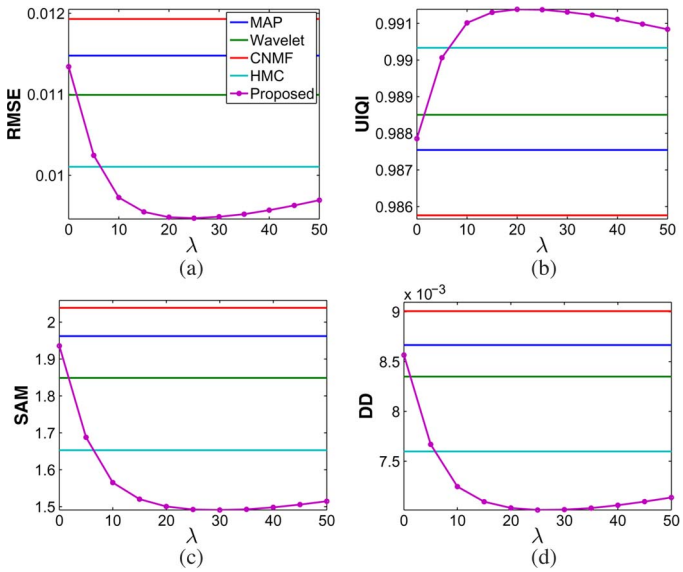

Fig. 7. Performance of the proposed fusion algorithm versus λ. (a) RMSE. (b) UIQI. (c) SAM. (d) DD.

All the algorithms have been implemented using MATLAB R2013A on a computer with an Intel Core i7-2600 central processing unit at 3.40 GHz and 8-GB random access memory. The fusion results obtained with the different algorithms are depicted in Fig. 6. Visually, the proposed method performs competitively with other state-of-the-art methods. To better illustrate the difference of the fusion results, quantitative results are reported in Table I, which shows the RMSE, UIQI, SAM, ERGAS, and DD for all methods. It can be seen that the proposed method always provides the best results.

E. Selection of the Regularization Parameter λ

To select an appropriate value of λ, the performance of the proposed algorithm has been evaluated as a function of λ. The results are displayed in Fig. 7, showing that there is no optimal value of λ for all the quality measures. In the simulation in Section V-D, we have chosen $\lambda = 25$, which provides the best fusion results in terms of RMSE. Note that for a wide range

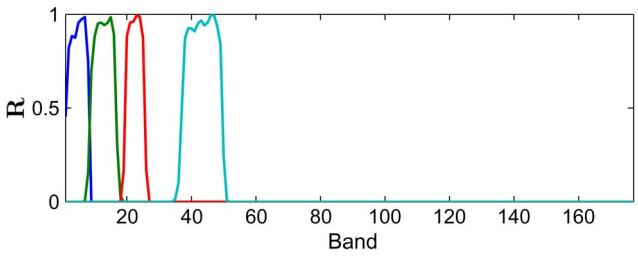

Fig. 8. LANDSAT spectral responses.

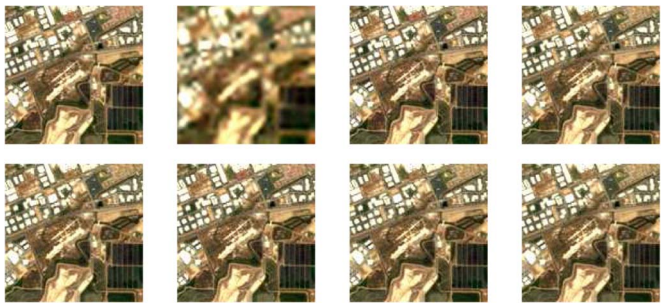

Fig. 9. Moffett data set. (Top 1) Reference. (Top 2) HS. (Top 3) MS. (Top 4) MAP [14]. (Bottom 1) Wavelet MAP [17]. (Bottom 2) CNMF fusion [18]. (Bottom 3) MMSE estimator [16]. (Bottom 4) Proposed method.

of λ, the proposed method always outperforms the other four methods.

F. Test With Other Data Sets

1) Fusion of AVIRIS Data and MS Data: The proposed fusion method has been tested with another data set. The reference image is a $128 \times 128 \times 176$ HS image acquired over Moffett Field, CA, USA, in 1994 by the Jet Propulsion Laboratory/ National Aeronautics and Space Administration Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [56]. The blurring kernel $\mathbf{B}$, downsampling operator $\mathbf{S}$, and SNRs for the two images are the same as in Section V-B. The reference image is filtered using the LANDSAT-like spectral responses depicted in Fig. 8, to obtain a four-band MS image. For the dictionaries and supports, the number and size of atoms and the sparsity of the code are the same as in Section V-B. The proposed fusion method has been applied to the observed HS and MS images with a subspace of dimension $\widetilde{m}_\lambda = 10$. The regularization parameter has been selected by cross-validation to get the best performance in terms of RMSE. The images (reference, HS, and MS) and the fusion results obtained with the different methods are shown in Fig. 9. More quantitative results are reported in Table II. These results are in good agreement with those we obtained with the previous image, proving that the proposed sparse-representation-based fusion algorithms improve the fusion quality. Note that other simulation results are available in [45] and confirm this improvement.

2) Pansharpening of AVIRIS Data: The only difference with Section V-F1 is that the MS image is replaced with a PAN image obtained by averaging all the bands of the reference image (contaminated by Gaussian noise with SNR $= 30$ dB). The quantitative results are given in Table III and are again in favor of the proposed fusion method.

TABLE II
PERFORMANCE OF DIFFERENT MS + HS FUSION METHODS (MOFFETT FIELD): RMSE (IN 10^{-2}), UIQI, SAM (IN DEGREES), ERGAS, DD (IN 10^{-2}), AND TIME (IN SECONDS)

Methods	RMSE	UIQI	SAM	ERGAS	DD	Time
MAP [14]	2.583	0.9580	4.586	3.106	1.814	**3**
Wavelet MAP [17]	2.150	0.9695	3.815	2.584	1.509	68
CNMF [18]	2.218	0.9669	3.885	2.663	1.529	17
HMC [16]	1.813	0.9772	3.201	2.221	1.239	12708
Rough $\tilde{\mathbf{U}}$	2.585	0.9579	4.575	3.107	1.814	\
Proposed	**1.745**	**0.9790**	**3.182**	**2.117**	**1.222**	280

TABLE III
PERFORMANCE OF DIFFERENT PANSHARPENING (HS + PAN) METHODS (MOFFETT FIELD): RMSE (IN 10^{-2}), UIQI, SAM (IN DEGREES), DD (IN 10^{-2}), AND TIME (IN SECONDS)

Methods	RMSE	UIQI	SAM	ERGAS	DD	Time
MAP [14]	1.857	0.9690	4.162	2.380	1.356	**2**
Wavelet MAP [17]	1.848	0.9697	4.191	2.354	1.360	55
CNMF [18]	1.964	0.9669	4.569	2.467	1.450	5
HMC [16]	1.748	0.9730	3.996	2.234	1.288	7828
Rough $\tilde{\mathbf{U}}$	1.853	0.9691	4.158	2.375	1.355	\
Proposed	**1.745**	**0.9731**	**3.948**	**2.231**	**1.281**	252

VI. CONCLUSION

In this paper, we proposed a novel method for HS and MS image fusion based on a sparse representation. The sparse representation ensured that the target image was well represented by atoms of dictionaries *a priori* learned from the observations. Identifying the supports jointly with the dictionaries circumvented the difficulty inherent to sparse coding. An alternate optimization algorithm, consisting of an ADMM and an LS regression, was designed to minimize the target function. Compared with other state-of-the-art fusion methods, the proposed fusion method offered smaller spatial error and smaller spectral distortion with a reasonable computation complexity. This improvement was attributed to the specific sparse prior designed to regularize the resulting inverse problem. Future works include the estimation of the regularization parameter λ within the fusion scheme. Updating the dictionary jointly with the target image would also deserve some attention.

ACKNOWLEDGMENT

The authors would like to thank Dr. P. Scheunders and Dr. Y. Zhang for sharing the codes of [17], Dr. N. Yokoya for sharing the codes of [18], Prof. P. Gamba for providing the ROSIS data over Pavia, and J. Inglada, from Centre National d'Études Spatiales, for providing the LANDSAT spectral responses used in the experiments.

REFERENCES

[1] Q. Wei, J. Bioucas-Dia, N. Dobigeon, and J.-Y. Tourneret, "Fusion of multispectral and hyperspectral images based on sparse representation," in *Proc. EUSIPCO*, Lisbon, Portugal, 2014, pp. 1–5.

[2] I. Amro, J. Mateos, M. Vega, R. Molina, and A. K. Katsaggelos, "A survey of classical methods and new trends in pansharpening of multispectral images," *EURASIP J. Adv. Signal Process.*, vol. 2011, no. 1, pp. 1–22, Sep. 2011.

[3] M. González-Audícana, J. L. Saleta, R. G. Catalán, and R. García, "Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1291–1299, Jun. 2004.

[4] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011.

[5] D. Liu and P. T. Boufounos, "Dictionary learning based pan-sharpening," in *Proc. IEEE ICASSP*, Kyoto, Japan, Mar. 2012, pp. 2397–2400.

[6] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.

[7] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[8] C.-I. Chang, *Hyperspectral Data Exploitation: Theory and Application*. Hoboken, NJ, USA: Wiley, 2007.

[9] G. Chen, S.-E. Qian, J.-P. Ardouin, and W. Xie, "Super-resolution of hyperspectral imagery using complex ridgelet transform," *Int. J. Wavelets Multiresolution Inf. Process.*, vol. 10, no. 3, pp. 1–22, May 2012.

[10] X. He, L. Condat, J. Bioucas-Dias, J. Chanussot, and J. Xia, "A new pansharpening method based on spatial and spectral sparsity priors," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4160–4174, Sep. 2014.

[11] N. Yokoya and A. Iwasaki, "Hyperspectral and multispectral data fusion mission on hyperspectral imager suite (HISUI)," in *Proc. IEEE IGARSS*, Melbourne, Vic., Australia, Jul. 2013, pp. 4086–4089.

[12] V. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogramm. Eng. Remote Sens.*, vol. 58, no. 5, pp. 561–567, 1992.

[13] J. Zhou, D. Civco, and J. Silander, "A wavelet transform method to merge Landsat TM and SPOT panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, Mar. 1998.

[14] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1174–1184, Sep. 2004.

[15] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multi-band images," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2014, pp. 3176–3180.

[16] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of hyperspectral and multispectral images," in *Proc. IEEE ICASSP*, Florence, Italy, May 2014.

[17] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3834–3843, Nov. 2009.

[18] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.

[19] N. Yokoya, J. Chanussot, and A. Iwasaki, "Hyperspectral and multispectral data fusion based on nonlinear unmixing," in *Proc. 4th WHISPERS*, Shanghai, China, Jun. 2012, pp. 1–4.

[20] E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," in *Proc. IEEE CVPR*, Minnesota, MN, USA, 2007, pp. 1–8.

[21] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process.*, vol. 17, no. 1, pp. 53–69, Jan. 2008.

[22] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. 26th Annu. ICML*, Montreal, QC, Canada, 2009, pp. 689–696.

[23] T. Deselaers and V. Ferrari, "Global and efficient self-similarity for object classification and detection," in *Proc. IEEE CVPR*, San Francisco, CA, USA, 2010, pp. 1633–1640.

[24] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

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

[26] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[27] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. IEEE CVPR*, San Francisco, CA, USA, 2010, pp. 3501–3508.

[28] Z. Xing, M. Zhou, A. Castrodad, G. Sapiro, and L. Carin, "Dictionary learning for noisy and incomplete hyperspectral images," *SIAM J. Imag. Sci.*, vol. 5, no. 1, pp. 33–56, Jan. 2012.

[29] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[30] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Stat. Soc. Ser. B*, vol. 73, no. 3, pp. 267–288, Jun. 1996.

[31] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–95, Mar. 2011.

[32] R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 231–246, Feb. 1999.

[33] R. Molina, M. Vega, J. Mateos, and A. K. Katsaggelos, "Variational posterior distribution approximation in Bayesian super resolution reconstruction of multispectral images," *Appl. Comput. Harmonic Anal.*, vol. 24, no. 2, pp. 251–267, Mar. 2008.

[34] A. Jalobeanu, L. Blanc-Feraud, and J. Zerubia, "An adaptive Gaussian model for satellite image deblurring," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 613–621, Apr. 2004.

[35] A. Duijster, P. Scheunders, and S. De Backer, "Wavelet-based em algorithm for multispectral-image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3892–3898, Nov. 2009.

[36] M. Xu, H. Chen, and P. K. Varshney, "An image fusion approach based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 5116–5127, Dec. 2011.

[37] C.-I. Chang, X.-L. Zhao, M. L. Althouse, and J. J. Pan, "Least squares subspace projection approach to mixed pixel classification for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 898–912, May 1998.

[38] J. M. Bioucas-Dias and J. M. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[39] T. Tu, S. Su, H. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inf. Fusion*, vol. 2, no. 3, pp. 177–186, Sep. 2001.

[40] V. P. Shah, N. H. Younan, and R. L. King, "An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1323–1335, May 2008.

[41] G. Dahlquist and Å. Björck, *Numerical Methods in Scientific Computing*, vol. 1, Philadelphia, PA, USA: SIAM, ser. Numerical Methods in Scientific Computing, 2008.

[42] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[43] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[44] O. G. Guleryuz, "Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising—Part I: Theory," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 539–554, Mar. 2006.

[45] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," IRIT-ENSEEIHT, Tech. Report, Univ. Toulouse, Toulouse, France, 2014. [Online]. Available: http://wei.perso.enseeiht.fr/papers/2014-TR-SparseFusion-WEI.pdf

[46] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[47] A. Gelman *et al.*, *Bayesian Data Analysis*, 3rd ed. Boca Raton, FL, USA: CRC Press, 2013.

[48] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *J. Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[49] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.

[50] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, Supplement 1, pp. S110–S122, Sep. 2009.

[51] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, Oct. 2010.

[52] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, Feb. 2013.
[53] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[54] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in *Proc. Int. Conf. Fusion Earth Data*, Nice, France, Jan. 2000, pp. 99–103.
[55] J. Eckstein and D. P. Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Programm.*, vol. 55, no. 1–3, pp. 293–318, Apr. 1992.
[56] R. O. Green *et al.*, "Imaging spectroscopy and the Airborne Visible/ Infrared Imaging Spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, Sep. 1998.

Qi Wei (S'13) was born in Shanxi, China, in 1989. He received the B.Sc. degree in electrical engineering from Beihang University (BUAA), Beijing, China, in 2010. He is currently working toward the Ph.D. degree with the National Polytechnic Institute of Toulouse (INP-ENSEEIHT, University of Toulouse), Toulouse, France.

He is also currently with the Signal and Communications Group, IRIT Laboratory, University of Toulouse. From February to August 2012, he was an Exchange Master Student with the Signal Processing and Communications Group, Department of Signal Theory and Communications (TSC), Universitat Politècnica de Catalunya, Barcelona, Spain. His research has been focused on statistical signal processing, particularly on inverse problems in image processing.

José Bioucas-Dias (S'87–M'95) received the E.E., M.Sc., Ph.D., and "Agregado" degrees from Instituto Superior Técnico (IST), Technical University of Lisbon (TULisbon, now University of Lisbon), Portugal, in 1985, 1991, 1995, and 2007, respectively, all in electrical and computer engineering.

Since 1995, he has been with the Department of Electrical and Computer Engineering, Instituto Superior Tico, Universidade de Lisboa, where he was an Assistant Professor from 1995 to 2007 and has been an Associate Professor since 2007. Since 1993, he has been also a Senior Researcher with the Pattern and Image Analysis Group, Instituto de Telecomunicações, which is a private nonprofit research institution. His research interests include inverse problems, signal and image processing, pattern recognition, optimization, and remote sensing. He has authored or coauthored over 250 scientific publications, including over 70 journal papers (48 of which were published in IEEE journals) and 180 peer-reviewed international conference papers and book chapters.

Nicolas Dobigeon (S'05–M'08–SM'13) was born in Angoulême, France, in 1981. He received the Bachelor's degree in electrical engineering from ENSEEIHT, Toulouse, France, in 2004 and the M.Sc. degree in signal processing, the Ph.D. degree, and the Habilitation Diriger des Recherches degree in signal processing from the National Polytechnic Institute of Toulouse (INP), Toulouse, France, in 2004, 2007, and 2012, respectively.

From 2007 to 2008, he was a Postdoctoral Research Associate with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor. MI, USA. Since 2008, he has been with INP Toulouse, University of Toulouse, Toulouse, where he is currently an Associate Professor. He conducts his research within the Signal and Communications Group, IRIT Laboratory. He is also an Affiliated Faculty Member of the TeSA Laboratory. His recent research activities have been focused on statistical signal and image processing, with a particular interest in Bayesian inverse problems, with applications to remote sensing, biomedical imaging, and genomics.

Jean-Yves Tourneret (SM'08) received the Ingénieur degree in electrical engineering from ENSEEIHT, Toulouse, France, in 1989 and the Ph.D. degree from the National Polytechnic Institute of Toulouse (INP), Toulouse, in 1992.

He is currently a Professor with the ENSEEIHT, University of Toulouse, and a member of the IRIT Laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal and image processing with a particular interest to Bayesian and Markov chain Monte Carlo methods.

Prof. Tourneret has been involved in the organization of several conferences, including the European Conference on Signal Processing, 2002 (Program Chair); the international conference ICASSP06 (Plenaries); the Statistical Signal Processing Workshop (SSP), 2012 (International Liaison); the International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2013 (local arrangements); SSP'14 (special sessions); and the Workshop on Machine Learning for Signal Processing, 2014 (special sessions). He has been the General Chair of the Centre International de Mathmatiques et Informatique de Toulouse (CIMI) workshop on optimization and statistics in image processing held in Toulouse in 2013 (with F. Malgouyres and D. Kouamé) and of CAMSAP 2015 (with P. Djuric). He has been a member of different technical committees, including the Signal Processing Theory and Methods Committee of the IEEE Signal Processing Society (2001–2007, 2010 to present). He has been serving as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2008–2011) and of the *EURASIP Journal on Signal Processing* (since July 2013).