FUSION OF MULTISPECTRAL AND HYPERSPECTRAL IMAGES BASED ON SPARSE REPRESENTATION

Qi Wei[(1)], *José M. Bioucas-Dias*[(2)], *Nicolas Dobigeon*[(1)], *and Jean-Yves Tourneret*[(1)]

[(1)] University of Toulouse, IRIT/INP-ENSEEIHT, 31071 Toulouse cedex 7, France

[(2)] Instituto de Telecomunicações, Instituto Superior Técnico, ULisbon, 1049-1, Portugal

ABSTRACT

This paper presents an algorithm based on sparse representation for fusing hyperspectral and multispectral images. The observed images are assumed to be obtained by spectral or spatial degradations of the high resolution hyperspectral image to be recovered. Based on this forward model, the fusion process is formulated as an inverse problem whose solution is determined by optimizing an appropriate criterion. To incorporate additional spatial information within the objective criterion, a regularization term is carefully designed, relying on a sparse decomposition of the scene on a set of dictionaries. The dictionaries and the corresponding supports of active coding coefficients are learned from the observed images. Then, conditionally on these dictionaries and supports, the fusion problem is solved by iteratively optimizing 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 fusion method when compared with the state-of-the-art.

***Index Terms*—** Image fusion, hyperspectral image, multispectral image, sparse representation, alternating direction method of multipliers (ADMM).

1. INTRODUCTION

Fusion of multi-sensor images has been a very active research topic during recent years [1]. When considering remotely sensed images, an archetypal fusion task is the pansharpening, i.e., fusing a high spatial resolution panchromatic (PAN) image and a low spatial resolution multispectral (MS) image. In recent years, hyperspectral (HS) imaging, acquiring a same scene in several hundreds of contiguous spectral bands, has opened a new range of relevant applications such as spectral unmixing [2] and classification [3]. To exploit the advantages offered by different sensors, how to fuse HS, MS or PAN images has been explored widely [4–6]. Note that the fusion of MS and HS differs from pansharpening since both spatial and spectral information is contained in multi-band images. Therefore, a lot of pansharpening methods, such as component substitution [7] and relative spectral contribution [8] are inapplicable or inefficient for the HS/MS fusion problem. To overcome the ill-posedness of the fusion problem, Bayesian inference provides a convenient way to regularize the inverse problem by defining an appropriate prior distribution for the scene of interest. Following this strategy, various estimators have been implemented in the image domain [9–11] or in a transformed domain [12].

Part of this work has been supported by the Chinese Scholarship Council (CSC, 201206020007), by the Hypanema ANR Project n°ANR-12- BS03-003, by Fundação para a Ciência e Tecnologia (FCT), projects PEst-OE/EEI/0008/2013 and PTDC/EEI-PRO/1470/2012, and by ANR-11-LABX-0040-CIMI in particular during the program ANR-11-IDEX-0002-02 within the thematic trimester on image processing.

Recent progress in sparse representations and dictionary learning (DL) have offered new efficient tools to address the multi-band fusion problem. Indeed, the self-similarity, which is prominent in natural images, implies that the patches extracted from natural images can be effectively represented with very few atoms coming from over-complete dictionaries [13–15]. More specifically, learning the decomposition dictionary from the images themselves, instead of resorting to predefined ones (e.g., wavelets), has recently led to state-of-the-art results for numerous low-level image processing tasks such as denoising. DL has also been investigated to analyze multi-band images [16]. More recently, Liu *et al.* proposed to solve the pansharpening problem based on a DL strategy [17].

In this paper, we propose to fuse the HS and MS images within a constrained optimization framework, by incorporating sparse regularization using dictionaries learned from the observed images. After learning the dictionaries and the corresponding supports of the codes from these observed images, we define an optimization problem which is solved by optimizing alternately with respect to the target image and the sparse code. The optimization with respect to the image is achieved by the split augmented Lagrangian shrinkage algorithm (SALSA) [18], which is an instance of the alternating direction method of multipliers (ADMM). By a suitable choice of variable splittings, SALSA enables us to decompose a huge non-diagonalizable quadratic problem into a sequence of convolutions and pixel decoupled problems, that can be solved efficiently. The estimation of the code is performed using a standard least-square (LS) algorithm which is possible because the support of the code has been fixed a priori. The resulting fusion strategy is summarized in Algorithm 1.

The paper is organized as follows. Section 2 formulates the fusion problem within a constrained optimization framework. Section 3 introduces the proposed sparse regularization and the method used to learn the dictionary and the code support. The optimization scheme proposed to solve the resulting optimization problem is detailed in Section 4. Simulation results are presented in Section 5 whereas conclusions are reported in Section 6.

2. PROBLEM FORMULATION

In this paper, we consider the fusion of HS and MS images. The HS image is supposed to be a spatially blurred and down-sampled version of $\mathbf{X}$ corrupted by additive Gaussian noise whereas the MS image is a spectrally degraded noisy version of $\mathbf{X}$. As a consequence, the observation models associated with the HS and MS images can be written as follows [9, 19]

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

where

- $\mathbf{X} \in \mathbb{R}^{m_\lambda \times n}$ is the full resolution unknown 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 HS and MS images,
- $\mathbf{B} \in \mathbb{R}^{n \times n}$ 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 as d),
- $\mathbf{R} \in \mathbb{R}^{n_\lambda \times m_\lambda}$ stands for the spectral response of the MS sensor,
- $\mathbf{N}_\mathrm{H} \in \mathbb{R}^{m_\lambda \times m}$ and $\mathbf{N}_\mathrm{M} \in \mathbb{R}^{n_\lambda \times n}$ are the HS and MS noises.

Note that $\mathbf{B}$ is a sparse symmetric Toeplitz matrix for a symmetric convolution kernel and $m = n/d^2$. In this work, we assume that $\mathbf{B}$, $\mathbf{S}$ and $\mathbf{R}$ are known. The elements of the matrices $\mathbf{N}_\mathrm{H}$ and $\mathbf{N}_\mathrm{M}$ are assumed to be independent zero-mean white Gaussian noises with variances s_h^2 and s_m^2 respectively.

The image $\mathbf{X}$ can be decomposed as $\mathbf{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n]$, where $\boldsymbol{x}_i = [x_{i,1}, \cdots, x_{i,m_\lambda}]^T$ is the $m_\lambda \times 1$ vector, also named hyperpixel, corresponding to the ith spatial location (with $i = 1, \cdots, n$). Because the HS bands are usually spectrally correlated, the HS vector $\boldsymbol{x}_i$ usually lives in a subspace whose dimension $\widetilde{m}_\lambda$ is much smaller than m_λ [2]. This property has been extensively exploited when analyzing HS data, in particular to perform spectral unmixing. More precisely, the image can be rewritten as $\mathbf{X} = \mathbf{V}\mathbf{U}$ where $\mathbf{V} \in \mathbb{R}^{m_\lambda \times \widetilde{m}_\lambda}$ has normalized orthogonal columns and $\mathbf{U} \in \mathbb{R}^{\widetilde{m}_\lambda \times n}$ is the projection of $\mathbf{X}$ onto the subspace spanned by the columns of $\mathbf{V}$. Incorporating this decomposition of the HS image $\mathbf{X}$ into the observation model (1) leads to

$$\begin{aligned} \mathbf{Y}_\mathrm{H} &= \mathbf{V}\mathbf{U}\mathbf{B}\mathbf{S} + \mathbf{N}_\mathrm{H} \\ \mathbf{Y}_\mathrm{M} &= \mathbf{R}\mathbf{V}\mathbf{U} + \mathbf{N}_\mathrm{M}. \end{aligned} \tag{2}$$

In this work, we assume that the signal subspace denoted as $\mathrm{span}\{\mathbf{V}\}$ has been previously identified, e.g., obtained from the available a priori knowledge regarding the scene of interest, or from a principal component analysis (PCA) of the HS data. Then, the considered fusion problem is solved in this lower-dimensional subspace, by estimating the projected image $\mathbf{U}$. The estimation of the projected image $\mathbf{U}$ from $\mathbf{Y}_\mathrm{H}$ and $\mathbf{Y}_\mathrm{M}$ is herein addressed by solving the inverse problem

$$\min_{\mathbf{U}} \frac{1}{2}\left\|\mathbf{Y}_\mathrm{H} - \mathbf{V}\mathbf{U}\mathbf{B}\mathbf{S}\right\|_F^2 + \frac{\lambda_m}{2}\left\|\mathbf{Y}_\mathrm{M} - \mathbf{R}\mathbf{V}\mathbf{U}\right\|_F^2 + \lambda_d \phi(\mathbf{U}), \tag{3}$$

where the two first terms are linked with the MS and HS images (data fidelity terms) and the last term is a penalty ensuring appropriate regularization. The parameter λ_m is equal to the ratio of the noise variances s_h^2/s_m^2 that is supposed to be a priori known and λ_d is regularization parameter. Various regularizations relying on ℓ_1, ℓ_2 or total variation [20] norms have been widely used to tackle this ill-posed problem. In this work, we derive an appropriate regularization term exploiting a sparse representation of the target image on a dictionary. More details are given in the next section.

3. DICTIONARY-BASED REGULARIZATION

The regularization proposed in this paper relies on the assumption that the target image $\mathbf{U}$ can be sparsely approximated on a given dictionary. Based on the self-similarity property of natural images, modeling images with a sparse representation has been shown to be very effective in many signal processing applications [13]. Based on these works, we propose to define the regularization term of (3) as

$$\phi(\mathbf{U}) = \frac{1}{2}\left\|\mathbf{U} - \bar{\mathbf{U}}\left(\mathbf{D}, \mathbf{A}\right)\right\|_F^2 \tag{4}$$

where $\mathbf{D}$ is the dictionary, $\mathbf{A}$ is the sparse code, and $\bar{\mathbf{U}}$ is the approximation of $\mathbf{U}$ derived from the dictionary and the code. Generally, an over-complete dictionary is proposed as a basis for the image patches. In many applications, the dictionary $\mathbf{D}$ is fixed a priori, and corresponds to various types of bases constructed using atoms such as wavelets [21] or discrete cosine transform coefficients [22]. However, these bases are not necessarily well matched to natural or remote sensing images since they do not necessarily adapt to the nature of the observed images. As a consequence, learning the dictionary from the observed images instead of using predefined bases generally improves signal representation [23]. More precisely, the strategy advocated in this paper consists of learning a dictionary $\mathbf{D}$ from the high resolution MS image to capture most of the spatial information contained in this image. To learn a dictionary from a multi-band image, a popular method consists of searching for a dictionary whose columns (or atoms) result from the lexicographically vectorization of the HS 3D patches [16, 24]. However, this strategy cannot be followed here since the dictionary is learned on the MS image $\mathbf{Y}_m \in \mathbb{R}^{n_\lambda \times n}$ composed of n_λ bands to approximate the target image $\mathbf{U}$ composed of $\widetilde{m}_\lambda$ spectral bands. Conversely, to capture most of the spatial details contained in each band of the MS image, we propose to approximate each band of the target image $\mathbf{U}$ by a sparse decomposition on a dedicated dictionary. In this case, the regularization term (4) can be written as

$$\phi(\mathbf{U}) = \frac{1}{2}\sum_{i=1}^{\widetilde{m}_\lambda}\left\|\mathbf{U}_i - \mathbf{L}\left(\mathbf{D}_i\mathbf{A}_i\right)\right\|_F^2 \tag{5}$$

where

- $\mathbf{U}_i \in \mathbb{R}^n$ is the ith band (or row) of $\mathbf{U} \in \mathbb{R}^{\widetilde{m}_\lambda \times n}$,
- $\mathbf{D}_i \in \mathbb{R}^{n_\mathrm{p} \times n_\mathrm{at}}$ is the dictionary dedicated to the ith band of $\mathbf{U}$ (n_p is the patch size and n_at is the number of atoms) and $\mathbf{D} = [\mathbf{D}_1, \cdots, \mathbf{D}_{\widetilde{m}_\lambda}]$,
- $\mathbf{A}_i \in \mathbb{R}^{n_\mathrm{at} \times n_\mathrm{pat}}$ is the ith band's code (n_pat is the number of patches associated with the ith band) and $\mathbf{A} = [\mathbf{A}_1, \cdots, \mathbf{A}_{\widetilde{m}_\lambda}]$,
- $\mathbf{L}(\cdot)$ is a linear operator that averages the overlapping patches of each band to restore the target image.

Note that each column of $\mathbf{D}_i$ is a basis element of size n_p (corresponding to the size of a patch). The dictionary is supposed to be fixed before addressing the fusion problem. The learning procedure used to estimate the dictionary is detailed in the following paragraph.

3.1. Dictionary learning and sparse coding

We propose to learn the set of dictionaries $\mathbf{D}_i$ from a rough estimation of $\mathbf{U}$, constructed from the MS image $\mathbf{Y}_\mathrm{M}$ and HS image $\mathbf{Y}_\mathrm{H}$, following the strategy used by Hardie *et al.* [9] and Zhang *et al.* [12]. More precisely, assuming that the hyperpixels of the target image $\mathbf{U}$ and MS data are jointly Gaussian distributed, the probability density function (pdf) of $\mathbf{U}$ conditionally upon $\mathbf{Y}_\mathrm{M}$ is also Gaussian

$$\begin{aligned} p(\mathbf{U}|\mathbf{Y}_\mathrm{M}) &= \prod_{i=1}^{n}\left[(2\pi)^{\widetilde{m}_\lambda}\left|\mathbf{C}_{\boldsymbol{u}_i|\boldsymbol{y}_{m,i}}\right|\right]^{-1/2} \\ &\times \exp\left\{-\frac{1}{2}\left(\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i|\boldsymbol{y}_{m,i}}\right)^T \mathbf{C}_{\boldsymbol{u}_i|\boldsymbol{y}_{m,i}}^{-1}\left(\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i|\boldsymbol{y}_{m,i}}\right)\right\} \end{aligned}$$

where $\mathbf{Y}_\mathrm{M} = \left[\boldsymbol{y}_{m,1}, \cdots, \boldsymbol{y}_{m,n}\right]$ and $\mathbf{U} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n]$. The conditional mean $\boldsymbol{\mu}_{\mathbf{U}|\mathbf{Y}_\mathrm{M}} = \mathrm{E}\left[\mathbf{U}|\mathbf{Y}_\mathrm{M}\right] = \left[\boldsymbol{\mu}_{\boldsymbol{u}_1|\boldsymbol{y}_{m,1}}, \cdots, \boldsymbol{\mu}_{\boldsymbol{u}_n|\boldsymbol{y}_{m,n}}\right]$ can be computed using joint pdf $p\left(\mathbf{U}, \mathbf{Y}_\mathrm{M}\right)$ and approximated as

in [9]. It provides a first approximation of the target image $\mathbf{U}$ to be restored. We propose to estimate the dictionaries $\mathbf{D}_i$ introduced in (5) by applying a DL algorithm on the patches of $\boldsymbol{\mu}_{\mathbf{U}|\mathbf{Y}_\mathrm{M}}$. Many DL methods have been studied in the recent literature. These methods are for instance based on K-SVD [14], online dictionary learning (ODL) [15] or Bayesian learning [16]. In this study, we have considered the ODL method to learn the set of over-complete dictionaries $\mathbf{D} = [\mathbf{D}_1, \cdots, \mathbf{D}_{\widetilde{m}_\lambda}]$. Once the dictionaries are learned, the orthogonal matching pursuit (OMP) is adopted to estimate the sparse code $\mathbf{A}_i$ for each band of $\mathbf{U}_i$. A maximum number of atoms, denoted as $n_{\max}$, is assumed to represent each patch of $\mathbf{U}_i$. Generally, the maximum number of atoms is much lower than the number of atoms in the dictionary, i.e., $n_{\max} \ll n_{\mathrm{at}}$. The positions of the non-zero elements of the code $\mathbf{A}_i$, namely the support $\boldsymbol{\Omega}_i \subset \mathbb{N}^2, i = 1, \cdots, \widetilde{m}_\lambda$ are also detected.

3.2. Re-estimation of the sparse code

Once the dictionaries $\mathbf{D}$ and codes $\mathbf{A}$ have been learned following the procedure detailed in the previous paragraph, it can be interesting to make the approximation in (5) more flexible for the fusion task. Interpreting the minimization problem in (3) as a standard maximum a posteriori estimation in a Bayesian framework, the regularization term (5) can be interpreted as a Gaussian prior distribution for the target image $\mathbf{U}$, with hyperparameters $\mathbf{D}$ and $\mathbf{A}$. Inspired by hierarchical models frequently encountered in Bayesian inference, we propose to include the code $\mathbf{A}$ within the estimation process. One strategy would consist of defining a new regularization term

$$\phi(\mathbf{U}, \mathbf{A}) = \frac{1}{2} \sum_{i=1}^{\widetilde{m}_\lambda} \left\| \mathbf{U}_i - \mathbf{L} \left(\mathbf{D}_i \mathbf{A}_i \right) \right\|_F^2 + \mu_a \left\| \mathbf{A}_i \right\|_0 \qquad (6)$$

where $\|.\|_0$ is the ℓ_0 counting function (or ℓ_0 norm) and μ_a is a regularization parameter. The ℓ_0-norm of code $\mathbf{A}$ is naturally chosen to enforce the sparsity of the code $\mathbf{A}_i \in \mathbb{R}^{n_{\mathrm{at}} \times n_{\mathrm{pat}}}$. However, the resulting optimization problem would become NP-hard. Conversely, in this work, we propose to fix the supports $\boldsymbol{\Omega}_i$ to the values coming from the sparse coding step detailed in the previous paragraph. Therefore, the ℓ_0 norm becomes a constant and the final regularization term (5) reduces to

$$\phi(\mathbf{U}, \mathbf{A}) = \frac{1}{2} \sum_{i=1}^{\widetilde{m}_\lambda} \left\| \mathbf{U}_i - \mathbf{L} \left(\mathbf{D}_i \mathbf{A}_i \right) \right\|_F^2 \text{ s.t. } \mathbf{A}_{i,\setminus\boldsymbol{\Omega}_i} = 0, \qquad (7)$$

where $\mathbf{A}_{i,\setminus\boldsymbol{\Omega}_i} = \{\mathbf{A}_i(l,k) \mid (l,k) \notin \boldsymbol{\Omega}_i\}$. The resulting objective criterion, which combines (7) with (3), is minimized using an alternate optimization procedure introduced in the following section.

4. ALTERNATE OPTIMIZATION

With known $\mathbf{D}$, $\boldsymbol{\Omega}$ and $\mathbf{V}$ learned from the HS and MS data, the problem (3) is a constrained quadratic optimization problem with respect to $\mathbf{U}$ and $\mathbf{A}$. However, this problem is difficult to solve due to the large dimensionality of $\mathbf{U}$ and due to the fact that the linear operators $\mathbf{V}(\cdot)\mathbf{B}\mathbf{D}$ and $\mathbf{L}(\cdot)$ cannot be easily diagonalized. To cope with this difficulty, we propose an optimization technique that alternates optimization with respect to $\mathbf{U}$ and $\mathbf{A}$.

Conditional on $\mathbf{A}$, the optimization with respect to $\mathbf{U}$ can be achieved efficiently with the SALSA algorithm [18]. Conditional on $\mathbf{U}$, the optimization with respect to $\mathbf{A}$ under the support constraint is an LS problem for the non-zero elements of $\mathbf{A}$, which can be easily solved. The overall resulting scheme that includes learning $\mathbf{D}$, $\boldsymbol{\Omega}$ and $\mathbf{V}$ is detailed in Algorithm 1. The alternate SALSA and LS steps are detailed below.

Algorithm 1: Alternate Optimization

Input: $\mathbf{Y}_\mathrm{H}$, $\mathbf{Y}_\mathrm{M}$, SNR_h, SNR_m, $\widetilde{m}_\lambda$ (HS subspace dimension), $\mathbf{R}$, $n_{\max}$ (number of maximum atoms for the support of each image patch)
Output: $\hat{\mathbf{X}}$ (high resolution HS image)

```
/* Estimate the conditional mean */
Approximate μ_{U|Y_M} using Y_M and Y_H following the method of [9]
/* Online dictionary learning */
D̂ ← ODL(μ̂_{U|Y_M})
/* Sparse image coding */
Â ← OMP(D̂, μ̂_{U|Y_M}, n_max)
/* Computing support */
Ω̂ ← Â ≠ 0
/* Computing subspace transform matrix */
V̂ ← PCA(Y_H, m̃_λ)
/* Start alternate optimization */
for t = 1, 2, ... to stopping rule do
    Û_t ∈ {U : L(U, Â_{t-1}) ≤ L(Û_{t-1}, Â_{t-1})} ;
    /* solved with SALSA */
    Â_t ∈ {A : L(Û_t, A) ≤ L(Û_t, Â_{t-1})} ;
    /* solved with LS */
end
X̂ = V̂Û
```

4.1. SALSA Step

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

$$\begin{aligned} L(\mathbf{U}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3) = \\ \frac{1}{2} \left\| \mathbf{Y}_\mathrm{H} - \mathbf{V}\mathbf{V}_1\mathbf{S} \right\|_F^2 + \frac{\mu}{2} \left\| \mathbf{U}\mathbf{B} - \mathbf{V}_1 - \mathbf{G}_1 \right\|_F^2 + \\ \frac{\lambda_m}{2} \left\| \mathbf{Y}_\mathrm{M} - \mathbf{R}\mathbf{V}\mathbf{V}_2 \right\|_F^2 + \frac{\mu}{2} \left\| \mathbf{U} - \mathbf{V}_2 - \mathbf{G}_2 \right\|_F^2 + \\ \frac{\lambda_d}{2} \left\| \bar{\mathbf{U}}(\mathbf{D}, \mathbf{A}) - \mathbf{V}_3 \right\|_F^2 + \frac{\mu}{2} \left\| \mathbf{U} - \mathbf{V}_3 - \mathbf{G}_3 \right\|_F^2. \end{aligned}$$

The update of $\mathbf{U}$ is achieved with the SALSA algorithm [18], which has a $\mathcal{O}\left(n_{it} \widetilde{m}_\lambda n \log\left(\widetilde{m}_\lambda n\right)\right)$ computational complexity, where n_{it} is the number of iterations for SALSA.

4.2. LS step

The objective of this step is to solve the following optimization problem with respect to $\mathbf{A}_i$ ($i = 1, \cdots, \widetilde{m}_\lambda$) conditional on $\mathbf{U}_i$

$$\min_{\mathbf{A}_i} \left\| \mathbf{U}_i - \mathbf{L}(\mathbf{D}_i \mathbf{A}_i) \right\|_F^2 \text{ s.t. } \mathbf{A}_{i,\setminus\boldsymbol{\Omega}_i} = 0.$$

It is a standard LS problem, which can be solved analytically. To tackle the support constraint efficiently, the optimization with respect to $\mathbf{A}_i$ considers only the non-zero elements of $\mathbf{A}_i$, denoted as $\mathbf{A}_{i,\boldsymbol{\Omega}_i} = \{\mathbf{A}_i(l,k) \mid (l,k) \in \boldsymbol{\Omega}_i\}$, which allows the computational complexity of the algorithm to be generally reduced to $\mathcal{O}\left(n_{\max} n_\mathrm{p} n_{\mathrm{pat}}\right)$.

5. SIMULATION RESULTS

This section studies the performance of the proposed sparse representation based fusion algorithm. The reference image considered here as the high spectral and high spectral image is an HS image acquired over Moffett field, CA, in 1994 by the JPL/NASA airborne visible/infrared imaging spectrometer (AVIRIS) [25]. This image is of size 128×128 and was composed of 224 bands that have been reduced to 177 bands after removing the water vapor absorption and noisy bands.

5.1. Simulation Scenario

We propose to reconstruct the reference hyperspectral image from two lower resolved images. First, we have generated a high-spectral low-spatial resolution HS image by applying a 5×5 Gaussian low-pass filter on each band of the reference image and downsampling every 4 pixels in both horizontal and vertical directions. In a second step, we have generated a 4-band MS image by filtering the reference image with the LANDSAT reflectance spectral responses [26]. The HS and MS images are both contaminated by zero-mean additive Gaussian noises with the signal to noise ratios (expressed in decibels) $\text{SNR}_h = 10 \log \left(\frac{\|\mathbf{XBS}\|_F^2}{\|\mathbf{N}_h\|_F^2} \right) = 30\text{dB}$ (HS image) and $\text{SNR}_m = 10 \log \left(\frac{\|\mathbf{RX}\|_F^2}{\|\mathbf{N}_m\|_F^2} \right) = 30\text{dB}$ (MS image). A composite color image, formed by selecting the red, green and blue bands of the reference image is shown in the bottom right of Fig. 1. The noise-contaminated HS and MS images are depicted in the top left and top right figures. (Note that the HS image has been interpolated for better visualization and that the MS image has been displayed using an arbitrary color composition).

The parameters used for the proposed fusion algorithm have been specified as follows

- The ODL algorithm has been run with patches of size 6×6, and with a maximum number of atoms $n_{\max} = 4$. These parameters have been selected by cross-validation.
- The regularization parameter used in the ADMM method is $\mu = 0.05$. Simulations have shown that the choice of μ does not affect significantly the fusion performance as long as the two optimization steps have converged.
- The regularization coefficient is $\lambda_d = 34 s_h^2$. The choice of this parameter will be discussed in Sec. 5.3 and has been tuned by cross-validation.

5.2. Comparison with other fusion methods

This section compares the proposed method with two other state-of-the-art algorithms studied in [9] and [12] for the fusion of HS and MS images. To evaluate the quality of the proposed fusion strategy, different image quality measures are investigated. Referring to [12], we propose to use RMSE (root mean square error), SAM (spectral angle mapper), UIQI (universal image quality index) and DD (degree of distortion) as quantitative measures. The definition of these indexes can be found in [1, 27]. Larger UIQI and smaller RMSE, SAM and DD indicate better fusion results. Fig. 1 shows that the proposed method offers competitive results comparing with the other two methods. Quantitative results are reported in Table 1 which shows the RMSE, UIQI, SAM and DD for all methods. It can be seen that the proposed method always provides the best results for the considered quality measures (at the price of a higher computational complexity).

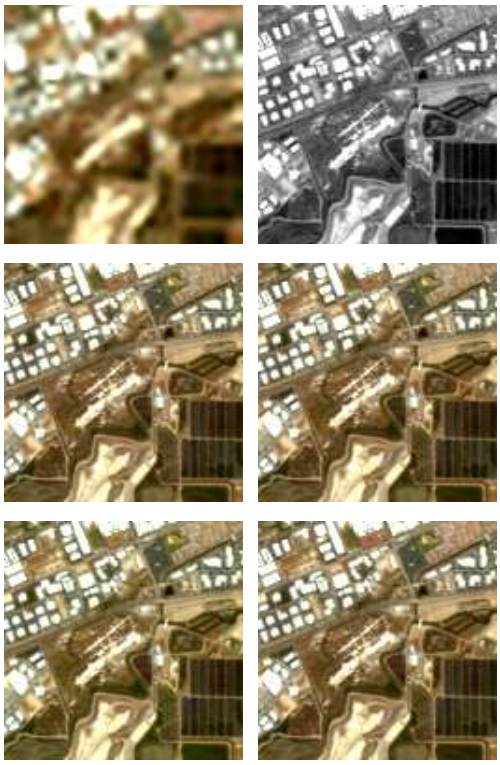

Fig. 1. Fusion results. (Top left) HS image. (Top right) MS image. (Middle left) MAP estimator [9]. (Middle right) Wavelet MAP estimator [12]. (Bottom left) Proposed DL-based fusion method. (Bottom right) Reference image.

Table 1. Performance of different MS + HS fusion methods: RMSE (in 10^{-2}), UIQI, SAM (in degree), DD (in 10^{-2}) and Time (in second)).

Methods	RMSE	UIQI	SAM	DD	Time
Hardie	15.416	0.9770	8.1158	9.9937	**3.2**
Zhang	13.892	0.9807	7.2929	8.9801	74.4
Proposed	**12.632**	**0.9848**	**6.8994**	**8.189**	747.0

5.3. Selection of the regularization parameter λ_d

In order to select an appropriate value of λ_d, we have tested the performance of the proposed algorithm when this parameter varies. The results are displayed in Fig. 2. Obviously, when λ_d is approaching 0 (no regularization), the performance is relatively poor. Each quality measure is convex with respect to λ_d. However, there is not a unique optimal value of λ_d for all the quality measures. In terms of RMSE, $\lambda_d = 38 s_h^2$ provides the best fusion results. The value of λ_d that has been used for all simulations presented before is selected as $\lambda_d = 34 s_h^2$, which is not too far from the 'optimal' point in the sense of RMSE.

6. CONCLUSIONS

This paper proposed a new dictionary learning based fusion method for the fusion of multispectral and hyperspectral images. A sparse regularization was introduced by considering that the image patches

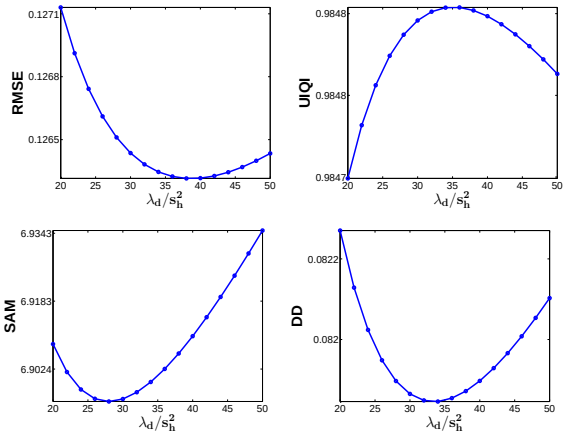

Fig. 2. Performance of the proposed fusion algorithm versus λ_d (from left to right): RMSE, UIQI, SAM and DD.

of the target image can be represented by the atoms learned from the observed images. The resulting cost function was simplified assuming that the code support has been estimated a priori by sparse coding. The target image and the values of the code were then determined by an alternate optimization technique. The alternating direction method of multipliers was finally investigated to solve the optimization with respect to the unknown image projected onto a lower dimensional subspace. Numerical experiments showed that the proposed method is always competitive with other state-of-the-art fusion methods. Future work includes the estimation of the HS and MS degradation operators and the validation of the proposed method on other datasets including real multispectral and hyperspectral images. Including the estimation of the regularization parameter into the optimization algorithm would also be interesting.

REFERENCES

[1] 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, 2011.

[2] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observations and Remote Sens.*, vol. 5, no. 2, pp. 354–379, 2012.

[3] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. Atli Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan 2014.

[4] M. Cetin and N. Musaoglu, "Merging hyperspectral and panchromatic image data: qualitative and quantitative analysis," *Int. J. Remote Sens.*, vol. 30, no. 7, pp. 1779–1804, 2009.

[5] C. Chisense, J. Engels, M. Hahn, and E. Gülch, "Pansharpening of hyperspectral images in urban areas," in *Proc. XXII Congr. of the Int. Society for Photogrammetry, Remote Sens.*, Melbourne, Australia, 2012.

[6] 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. 03, 2012.

[7] 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.

[8] 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, 1998.

[9] 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, Sept. 2004.

[10] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multi-band images," *arXiv preprint arXiv:1307.5996*, 2013.

[11] ——, "Bayesian fusion of hyperspectral and multispectral images," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Florence, Italy, May 2014.

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

[13] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1998.

[14] 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, 2006.

[15] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. Int. Conf. Machine Learning (ICML)*. ACM, 2009, pp. 689–696.

[16] Z. Xing, M. Zhou, A. Castrodad, G. Sapiro, and L. Carin, "Dictionary learning for noisy and incomplete hyperspectral images," *SIAM Journal on Imaging Sciences*, vol. 5, no. 1, pp. 33–56, 2012.

[17] D. Liu and P. T. Boufounos, "Dictionary learning based pan-sharpening," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, March 2012, pp. 2397–2400.

[18] M. Afonso, J. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, 2011.

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

[20] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 460–489, 2005.

[21] S. Mallat, *A wavelet tour of signal processing*. New York: Academic Press, 1999.

[22] N. Ahmed, T. Natarajan, and K. Rao, "Discrete cosine transform," *IEEE Trans. Computers*, vol. C-23, no. 1, pp. 90–93, 1974.

[23] 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, 2006.

[24] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Trans. Geosci. and Remote Sens.*, vol. 51, no. 9, pp. 4779–4789, Sept 2013.

[25] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. of Environment*, vol. 65, no. 3, pp. 227–248, 1998.

[26] D. Fleming, "Effect of relative spectral response on multi-spectral measurements and NDVI from different remote sensing systems," Ph.D. dissertation, University of Maryland, 2006.

[27] X. X. Zhu, X. Wang, and R. Bamler, "Compressive sensing for image fusion-with application to pan-sharpening," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Vancouver, Canada, July 2011, pp. 2793–2796.