

Fast Fusion of Multi-Band Images Based on Solving a Sylvester Equation

- Complementary results and supporting materials

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

TECHNICAL REPORT – 2015, February

University of Toulouse, IRIT/INP-ENSEEIH

2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France

Abstract

This paper proposes a fast multi-band image fusion algorithm, which combines a high-spatial low-spectral resolution image and a low-spatial high-spectral resolution image. The well admitted forward model is explored to form the likelihoods of the observations. Maximizing the likelihoods leads to solving a Sylvester equation. By exploiting the properties of the circulant and downsampling matrices associated with the fusion problem, a closed-form solution for the corresponding Sylvester equation is obtained explicitly, getting rid of any iterative update step. Coupled with the alternating direction method of multipliers and the block coordinate descent method, the proposed algorithm can be easily generalized to incorporate prior information for the fusion problem, allowing a Bayesian estimator. Simulation results show that the proposed algorithm achieves the same performance as existing algorithms with the advantage of significantly decreasing the computational complexity of these algorithms.

Index Terms

Multi-band image fusion, Bayesian estimation, circulant matrix, Sylvester equation, alternating direction method of multipliers, block coordinate descent.

Part of this work has been supported by the Chinese Scholarship Council, the Hypanema ANR Project n° ANR-12-BS03-003 and by ANR-11-LABX-0040-CIMI within the program ANR-11-IDEX-0002-02 within the thematic trimester on image processing.

Qi Wei, Nicolas Dobigeon and Jean-Yves Tourneret are with University of Toulouse, IRIT/INP-ENSEEIH, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France (e-mail: {qi.wei, nicolas.dobigeon, jean-yves.tourneret}@enseeiht.fr).

I. INTRODUCTION

A. Background

In general, a multi-band image can be represented as a three-dimensional data cube indexed by three exploratory variables (x, y, λ) , where x and y are the two spatial dimensions of the scene, and λ is the spectral dimension (covering a range of wavelengths). Typical examples of multi-band images include hyperspectral (HS) images [1], multi-spectral (MS) images [2], integral field spectrographs [3], magnetic resonance spectroscopy images etc. However, multi-band imaging generally suffers from the limited spatial resolution of the data acquisition devices, mainly due to an unsurpassable tradeoff between spatial and spectral sensitivities [4]. For example, HS images benefit from excellent spectroscopic properties with hundreds of bands but are limited by their relatively low spatial resolution compared with MS and panchromatic (PAN) images (which are acquired in much fewer bands). As a consequence, reconstructing a high-spatial and high-spectral multi-band image from two degraded and complementary observed images is a challenging but crucial issue that has been addressed in various scenarios. In particular, fusing a high-spatial low-spectral resolution image and a low-spatial high-spectral image is an archetypal instance of multi-band image reconstruction, such as pansharpening (MS+PAN) [5] or hyperspectral pansharpening (HS+PAN) [6]. Generally, the linear degradations applied to the observed images with respect to (w.r.t.) the target high-spatial and high-spectral image reduce to spatial and spectral transformations. Thus, the multi-band image fusion problem can be interpreted as restoring a three dimensional data-cube from two degraded data-cubes. A more precise description of the problem formulation is provided in the following paragraph.

B. Problem Statement

To better distinguish spectral and spatial degradations, the pixels of the target multi-band image, which is of high-spatial and high-spectral resolution, can be rearranged to build an $m_\lambda \times n$ matrix \mathbf{X} , where m_λ is the number of spectral bands and $n = n_r \times n_c$ is the number of pixels in each band (n_r and n_c represents the number of rows and columns respectively). In other words, each column of the matrix \mathbf{X} consists of a m_λ -valued pixel and each row gathers

all the pixel values in a given spectral band. Based on this pixel ordering, any linear operation applied on the left (resp. right) side of \mathbf{X} describes a spectral (resp. spatial) degradation.

In this work, we assume that two complementary images of high-spectral or high-spatial resolutions, respectively, are available to reconstruct the target high-spectral and high-spatial resolution target image. These images result from linear spectral and spatial degradations of the full resolution image \mathbf{X} , according to the well-admitted model

$$\begin{aligned}\mathbf{Y}_L &= \mathbf{L}\mathbf{X} + \mathbf{N}_L \\ \mathbf{Y}_R &= \mathbf{X}\mathbf{R} + \mathbf{N}_R\end{aligned}\tag{1}$$

where

- $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}^{m_\lambda \times n}$ is the full resolution target image,
- $\mathbf{Y}_L \in \mathbb{R}^{n_\lambda \times n}$ and $\mathbf{Y}_R \in \mathbb{R}^{m_\lambda \times m}$ are the observed spectrally degraded and spatially degraded images,
- $m = m_r \times m_c$ is the number of pixels of the high-spectral resolution image,
- n_λ is the number of bands of the high-spatial resolution image,
- \mathbf{N}_L and \mathbf{N}_R are additive terms that include both modeling errors and sensor noises.

The noise matrices are assumed to be distributed according to the following matrix normal distributions¹

$$\begin{aligned}\mathbf{N}_L &\sim \mathcal{MN}_{m_\lambda, m}(\mathbf{0}_{m_\lambda, m}, \mathbf{\Lambda}_L, \mathbf{I}_m) \\ \mathbf{N}_R &\sim \mathcal{MN}_{n_\lambda, n}(\mathbf{0}_{n_\lambda, n}, \mathbf{\Lambda}_R, \mathbf{I}_n).\end{aligned}$$

Note that no particular structure is assumed for the row covariance matrices $\mathbf{\Lambda}_L$ and $\mathbf{\Lambda}_R$ except that they are both positive definite, which allows for considering spectrally colored noises. Conversely, the column covariance matrices are assumed to be the identity matrix to reflect the fact that the noise is pixel-independent.

In most practical scenarios, the spectral degradation $\mathbf{L} \in \mathbb{R}^{n_\lambda \times m_\lambda}$ only depends on the spectral response of the sensor, which can be *a priori* known or estimated by cross-calibration

¹The probability density function $p(\mathbf{X}|\mathbf{M}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c)$ of a matrix normal distribution $\mathcal{MN}_{r,c}(\mathbf{M}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c)$ is defined by

$$p(\mathbf{X}|\mathbf{M}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c) = \frac{\exp\left(-\frac{1}{2}\text{tr}\left[\mathbf{\Sigma}_c^{-1}(\mathbf{X} - \mathbf{M})^T \mathbf{\Sigma}_r^{-1}(\mathbf{X} - \mathbf{M})\right]\right)}{(2\pi)^{rc/2} |\mathbf{\Sigma}_c|^{r/2} |\mathbf{\Sigma}_r|^{c/2}}$$

where $\mathbf{M} \in \mathbb{R}^{r \times c}$ is the mean matrix, $\mathbf{\Sigma}_r \in \mathbb{R}^{r \times r}$ is the row covariance matrix and $\mathbf{\Sigma}_c \in \mathbb{R}^{c \times c}$ is the column covariance matrix.

[7]. The spatial degradation \mathbf{R} includes warp, translation, blurring, decimation, etc. As the warp and translation can be attributed to the image co-registration problem and mitigated by precorrection, only blurring and decimation degradations, denoted \mathbf{B} and \mathbf{S} are considered in this work. If the spatial blurring is assumed to be space-invariant, $\mathbf{B} \in \mathbb{R}^{n \times n}$ owns the specific property of being a cyclic convolution operator acting on the bands. The matrix $\mathbf{S} \in \mathbb{R}^{n \times m}$ is a $d = d_r \times d_c$ uniform downsampling operator, which has $m = n/d$ ones on the block diagonal and zeros elsewhere, and such that $\mathbf{S}^T \mathbf{S} = \mathbf{I}_m$. Note that multiplying by \mathbf{S}^T represents zero-interpolation to increase the number of pixels from m to n . Therefore, assuming \mathbf{R} can be decomposed as $\mathbf{R} = \mathbf{B}\mathbf{S} \in \mathbb{R}^{n \times m}$, the fusion model (1) can be rewritten as

$$\begin{aligned} \mathbf{Y}_L &= \mathbf{L}\mathbf{X} + \mathbf{N}_L \\ \mathbf{Y}_R &= \mathbf{X}\mathbf{B}\mathbf{S} + \mathbf{N}_R \end{aligned} \tag{2}$$

where all matrix dimensions and their respective relations are summarized in Table I.

This matrix equation (1) has been widely advocated in the pansharpening and HS pansharpening problems, which consist of fusing a PAN image with an MS or an HS image [6], [8], [9]. Similarly, most of the techniques developed to fuse MS and HS images also rely on a similar linear model [10]–[16]. From an applicative point of view, this problem is also important as motivated by recent national programs, e.g., the Japanese next-generation space-borne HS image suite (HISUI), which fuses co-registered MS and HS images acquired over the same scene under the same conditions [17].

To summarize, the problem of fusing high-spectral and high-spatial resolution images can be formulated as estimating the unknown matrix \mathbf{X} from (2). There are two main statistical estimation methods that can be used to solve this problem. These methods are based on maximum likelihood (ML) or on Bayesian inference. ML estimation is purely data-driven while Bayesian estimation relies on prior information, which can be regarded as a regularization (or a penalization) for the fusion problem. Various priors have been already advocated to regularize the multi-band image fusion problem, such as Gaussian priors [18], [19], sparse representations [16] or total variation (TV) [20] priors. The choice of the prior usually depends on the information resulting from previous experiments or from a subjective view of constraints affecting the unknown model parameters [21], [22].

Computing the ML or the Bayesian estimators (whatever the form chosen for the prior) is a challenging task, mainly due to the large size of \mathbf{X} and to the presence of the downsampling operator \mathbf{S} , which prevents any direct use of the Fourier transform to diagonalize the blurring operator \mathbf{B} . To overcome this difficulty, several computational strategies have been designed to approximate the estimators. Based on a Gaussian prior modeling, a Markov chain Monte Carlo (MCMC) algorithm has been implemented in [18] to generate a collection of samples asymptotically distributed according to the posterior distribution of \mathbf{X} . The Bayesian estimators of \mathbf{X} can then be approximated using these samples. Despite this formal appeal, MCMC-based methods have the major drawback of being computationally expensive, which prevents their effective use when processing images of large size. Relying on exactly the same prior model, the strategy developed in [19] exploits an alternating direction method of multipliers (ADMM) embedded in a block coordinate descent method (BCD) to compute the maximum a posteriori (MAP) estimator of \mathbf{X} . This optimization strategy allows the numerical complexity to be greatly decreased when compared to its MCMC counterpart. Based on a prior built from a sparse representation, the fusion problem is solved in [16], [20] with the split augmented Lagrangian shrinkage algorithm (SALSA) [23], which is an instance of ADMM.

In this paper, contrary to the algorithms described above, a much more efficient method is proposed to solve explicitly an underlying Sylvester equation (SE) associated with the fusion problem derived from (2). This solution can be implemented *per se* to compute the ML estimator in a computationally efficient manner. The proposed solution has also the great advantage of being easily generalizable within a Bayesian framework when considering various priors. The MAP estimators associated with a Gaussian prior similar to [18], [19] can be directly computed thanks to the proposed strategy. When handling more complex priors such as [16], [20], the SE solution can be conveniently embedded within a conventional ADMM or a BCD algorithm.

C. Paper organization

The remaining of this paper is organized as follows. Section II studies the optimization problem to be addressed in absence of any regularization, i.e., in an ML framework. The

proposed fast fusion method is presented in Section III and generalized to Bayesian estimators associated with various priors in Section IV. Section V presents experimental results assessing the accuracy and the numerical efficiency of the proposed fusion method. Conclusions are finally reported in Section VI.

II. PROBLEM FORMULATION

Using the statistical properties of the noise matrices \mathbf{N}_L and \mathbf{N}_R , \mathbf{Y}_L and \mathbf{Y}_R have matrix Gaussian distributions, i.e.,

$$\begin{aligned} p(\mathbf{Y}_L|\mathbf{X}) &= \mathcal{MN}_{n_\lambda, n}(\mathbf{L}\mathbf{X}, \mathbf{\Lambda}_L, \mathbf{I}_n) \\ p(\mathbf{Y}_R|\mathbf{X}) &= \mathcal{MN}_{m_\lambda, m}(\mathbf{XBS}, \mathbf{\Lambda}_R, \mathbf{I}_m). \end{aligned} \quad (3)$$

As the collected measurements \mathbf{Y}_L and \mathbf{Y}_R have been acquired by different (possibly heterogeneous) sensors, the noise matrices \mathbf{N}_L and \mathbf{N}_R are sensor-dependent and can be generally assumed to be statistically independent. Therefore, \mathbf{Y}_L and \mathbf{Y}_R are independent conditionally upon the unobserved scene \mathbf{X} . As a consequence, the joint likelihood function of the observed data is

$$p(\mathbf{Y}_L, \mathbf{Y}_R|\mathbf{X}) = p(\mathbf{Y}_L|\mathbf{X}) p(\mathbf{Y}_R|\mathbf{X}). \quad (4)$$

Since adjacent HS bands are known to be highly correlated, the HS vector \mathbf{x}_i usually lives in a subspace whose dimension is much smaller than the number of bands m_λ [24], [25], i.e., $\mathbf{X} = \mathbf{H}\mathbf{U}$ where \mathbf{H} is an orthogonal matrix such that $\mathbf{H}^T\mathbf{H} = \mathbf{I}_{\tilde{m}_\lambda}$ and $\mathbf{U} \in \mathbb{R}^{\tilde{m}_\lambda \times n}$ is the projection of \mathbf{X} onto the subspace spanned by the columns of $\mathbf{H} \in \mathbb{R}^{m_\lambda \times \tilde{m}_\lambda}$.

Defining $\Psi = \{\mathbf{Y}_L, \mathbf{Y}_R\}$ as the set of the observed images, the negative logarithm of the likelihood is

$$\begin{aligned} -\log p(\Psi|\mathbf{U}) &= -\log p(\mathbf{Y}_L|\mathbf{U}) - \log p(\mathbf{Y}_R|\mathbf{U}) + C \\ &= \frac{1}{2} \text{tr} \left((\mathbf{Y}_R - \mathbf{HUBS})^T \mathbf{\Lambda}_R^{-1} (\mathbf{Y}_R - \mathbf{HUBS}) \right) + \\ &\quad \frac{1}{2} \text{tr} \left((\mathbf{Y}_L - \mathbf{LHU})^T \mathbf{\Lambda}_L^{-1} (\mathbf{Y}_L - \mathbf{LHU}) \right) + C \end{aligned}$$

where C is a constant. Thus, calculating the ML estimator of \mathbf{U} from the observed images

Ψ , i.e., maximizing the likelihood can be achieved by solving the following problem

$$\arg \min_{\mathbf{U}} L(\mathbf{U}) \quad (5)$$

where

$$L(\mathbf{U}) = \text{tr} \left((\mathbf{Y}_R - \mathbf{HUBS})^T \mathbf{\Lambda}_R^{-1} (\mathbf{Y}_R - \mathbf{HUBS}) \right) + \text{tr} \left((\mathbf{Y}_L - \mathbf{LHU})^T \mathbf{\Lambda}_L^{-1} (\mathbf{Y}_L - \mathbf{LHU}) \right).$$

Note that it is also obvious to formulate the optimization problem (5) from the linear model (2) directly in the least-squares (LS) sense [26]. However, specifying the distributions of the noises $\mathbf{\Lambda}_L$ and $\mathbf{\Lambda}_R$ allows us to consider the case of colored noises (band-dependent) more easily by introducing the covariance matrices $\mathbf{\Lambda}_R$ and $\mathbf{\Lambda}_L$, leading to the weighting LS problem (5).

In this paper, we prove that the minimization of (5) w.r.t. the target image \mathbf{U} can be solved analytically, without any iterative optimization scheme or Monte Carlo based method. The resulting closed-form solution to the optimization problem is presented in Section III. Furthermore, it is shown in Section IV that the proposed method can be easily generalized to Bayesian fusion methods with appropriate prior distributions.

III. FAST FUSION SCHEME

A. Sylvester equation

Minimizing (5) w.r.t. \mathbf{U} is equivalent to force the derivative of $L(\mathbf{U})$ to be zero, i.e., $dL(\mathbf{U})/d\mathbf{U} = 0$, leading to the following matrix equation

$$\begin{aligned} & \mathbf{H}^H \mathbf{\Lambda}_R^{-1} \mathbf{HUBS} (\mathbf{BS})^H + \left((\mathbf{LH})^H \mathbf{\Lambda}_L^{-1} \mathbf{LH} \right) \mathbf{U} \\ & = \mathbf{H}^H \mathbf{\Lambda}_R^{-1} \mathbf{Y}_R (\mathbf{BS})^H + (\mathbf{LH})^H \mathbf{\Lambda}_L^{-1} \mathbf{Y}_L. \end{aligned} \quad (6)$$

As mentioned in Section I-B, the difficulty for solving (6) results from the high dimensionality of \mathbf{U} and the presence of the downsampling matrix \mathbf{S} . In this work, we will show that Eq. (6) can be solved analytically with some assumptions summarized below.

Assumption 1. *The blurring matrix \mathbf{B} is a circulant matrix.*

A consequence of this assumption is that \mathbf{B} can be decomposed as $\mathbf{B} = \mathbf{F}\mathbf{D}\mathbf{F}^H$ and $\mathbf{B}^H = \mathbf{F}\mathbf{D}^*\mathbf{F}^H$, where $\mathbf{F} \in \mathbb{R}^{n \times n}$ is the discrete Fourier transform (DFT) matrix ($\mathbf{F}\mathbf{F}^H = \mathbf{F}^H\mathbf{F} = \mathbf{I}_n$), $\mathbf{D} \in \mathbb{R}^{n \times n}$ is a diagonal matrix and $*$ represents the conjugate operator.

Assumption 2. *The decimation matrix \mathbf{S} corresponds to downsampling the original signal and its conjugate transpose \mathbf{S}^H interpolates the decimated signal with zeros.*

A decimation matrix satisfies the property $\mathbf{S}^H\mathbf{S} = \mathbf{I}_m$. Moreover, the matrix $\underline{\mathbf{S}} \triangleq \mathbf{S}\mathbf{S}^H \in \mathbb{R}^{n \times n}$ is symmetric and idempotent, i.e., $\underline{\mathbf{S}} = \underline{\mathbf{S}}^H$ and $\underline{\mathbf{S}}\underline{\mathbf{S}}^H = \underline{\mathbf{S}}^2 = \underline{\mathbf{S}}$. For a practical implementation, multiplying an image by $\underline{\mathbf{S}}$ can be achieved by doing entry-wise multiplication with an $n \times n$ mask matrix with ones in the sampled position and zeros elsewhere.

Note that the assumptions w.r.t. the blurring matrix \mathbf{B} and the decimation matrix \mathbf{S} have been widely used in image processing application, like super-resolution [27], [28], fusion [10], [20], etc.

After multiplying (6) on both sides by $(\mathbf{H}^H\mathbf{\Lambda}_R^{-1}\mathbf{H})^{-1/2}$ we obtain

$$\mathbf{C}_1\mathbf{U} + \mathbf{U}\mathbf{C}_2 = \mathbf{C}_3 \quad (7)$$

where

$$\begin{aligned} \mathbf{C}_1 &= (\mathbf{H}^H\mathbf{\Lambda}_R^{-1}\mathbf{H})^{-1} \left((\mathbf{LH})^H\mathbf{\Lambda}_L^{-1}\mathbf{LH} \right) \\ \mathbf{C}_2 &= \mathbf{B}\underline{\mathbf{S}}\mathbf{B}^H \\ \mathbf{C}_3 &= (\mathbf{H}^H\mathbf{\Lambda}_R^{-1}\mathbf{H})^{-1} \left(\mathbf{H}^H\mathbf{\Lambda}_R^{-1}\mathbf{Y}_R(\mathbf{BS})^H + (\mathbf{LH})^H\mathbf{\Lambda}_L^{-1}\mathbf{Y}_L \right). \end{aligned}$$

Eq. (7) is a generalized Sylvester matrix equation [29]. It is well known that a SE has a unique solution if and only if an arbitrary sum of the eigenvalues of \mathbf{C}_1 and \mathbf{C}_2 is not equal to zero [29].

B. Existence of a solution

In this section, we study the eigenvalues of \mathbf{C}_1 and \mathbf{C}_2 to check if (7) has a unique solution. As matrix $\mathbf{C}_2 = \mathbf{B}\underline{\mathbf{S}}\mathbf{B}^H$ is positive semi-definite, its eigenvalues include positive values and zeros [30]. In order to study the eigenvalues of \mathbf{C}_1 , Lemma 1 is introduced below.

²The invertibility of the matrix $\mathbf{H}^H\mathbf{\Lambda}_R^{-1}\mathbf{H}$ is guaranteed since \mathbf{H} has full column rank and $\mathbf{\Lambda}_R$ is positive definite.

Lemma 1. *If matrix $\mathbf{A}_1 \in \mathbb{R}^{n \times n}$ is symmetric (resp. Hermitian) positive definite and matrix $\mathbf{A}_2 \in \mathbb{R}^{n \times n}$ is symmetric (resp. Hermitian) positive semi-definite, the product $\mathbf{A}_1 \mathbf{A}_2$ is diagonalizable and all the eigenvalues of $\mathbf{A}_1 \mathbf{A}_2$ are non-negative.*

Proof. See proof in Appendix A. □

According to Lemma 1, since the matrix \mathbf{C}_1 is the product of a symmetric positive definite matrix $(\mathbf{H}^H \mathbf{\Lambda}_R^{-1} \mathbf{H})^{-1}$ and a symmetric semi-definite matrix $(\mathbf{LH})^H \mathbf{\Lambda}_L^{-1} \mathbf{LH}$, it is diagonalizable and all its eigenvalues are non-negative. As a consequence, the eigen-decomposition of \mathbf{C}_1 can be expressed as $\mathbf{C}_1 = \mathbf{Q} \mathbf{\Lambda}_1 \mathbf{Q}^{-1}$, where $\mathbf{\Lambda}_1 = \text{diag}(\lambda_1, \dots, \lambda_{\tilde{m}_\lambda})$ ($\text{diag}(\lambda_1, \dots, \lambda_{\tilde{m}_\lambda})$) is a diagonal matrix whose elements are $\lambda_1, \dots, \lambda_{\tilde{m}_\lambda}$ and $\lambda_i \geq 0, \forall i$. Therefore, as long as zero is not an eigenvalue of \mathbf{C}_1 (or equivalently \mathbf{C}_1 is invertible), any sum of eigenvalues of \mathbf{C}_1 and \mathbf{C}_2 is different from zero (more accurately, this sum is > 0), leading to the existence of a unique solution of (7).

However, the invertibility of \mathbf{C}_1 is not always guaranteed depending on the forms and dimensions of \mathbf{H} and \mathbf{L} . For example, if $n_\lambda < \tilde{m}_\lambda$, meaning the number of MS bands is smaller than the subspace dimension, the matrix $(\mathbf{LH})^H \mathbf{\Lambda}_L^{-1} \mathbf{LH}$ is rank deficient and thus there is not a unique solution of (7). In cases when \mathbf{C}_1 is singular, a regularization or prior information is necessary to be introduced to ensure (7) has a unique solution. In this section, we focus on the case when \mathbf{C}_1 is non-singular. The generalization to Bayesian estimators based on specific priors already considered in the literature will be elaborated in Section IV.

C. A classical algorithm for the Sylvester matrix equation

A classical algorithm for obtaining a solution of the SE is the Bartels-Stewart algorithm [29]. This algorithm decomposes \mathbf{C}_1 and \mathbf{C}_2 into Schur forms using a QR algorithm and solves the resulting triangular system via back-substitution. However, as the matrix $\mathbf{C}_2 = \mathbf{B} \mathbf{S} \mathbf{B}^H$ is very large for our application ($n \times n$, where n is the number of image pixels), it is unfeasible to construct the matrix \mathbf{C}_2 , let alone use the QR algorithm to compute its Schur form (which has the computational cost $\mathcal{O}(n^3)$ arithmetical operations). The next section proposes an innovative strategy to obtain an analytical expression of the SE (7) by exploiting the specific properties of the matrices \mathbf{C}_1 and \mathbf{C}_2 associated with the fusion

problem.

D. Proposed closed-form solution

Using the decomposition $\mathbf{C}_1 = \mathbf{Q}\mathbf{\Lambda}_C\mathbf{Q}^{-1}$ and multiplying both sides of (7) by \mathbf{Q}^{-1} leads to

$$\mathbf{\Lambda}_C\mathbf{Q}^{-1}\mathbf{U} + \mathbf{Q}^{-1}\mathbf{U}\mathbf{C}_2 = \mathbf{Q}^{-1}\mathbf{C}_3. \quad (8)$$

Right multiplying (8) by $\mathbf{F}\mathbf{D}$ on both sides and using the definitions of matrices \mathbf{C}_2 and \mathbf{B} yields

$$\mathbf{\Lambda}_1\mathbf{Q}^{-1}\mathbf{U}\mathbf{F}\mathbf{D} + \mathbf{Q}^{-1}\mathbf{U}\mathbf{F}\mathbf{D}(\mathbf{F}^H\mathbf{S}\mathbf{F}\mathbf{D}) = \mathbf{Q}^{-1}\mathbf{C}_3\mathbf{F}\mathbf{D} \quad (9)$$

where $\mathbf{D} = (\mathbf{D}^*)\mathbf{D}$ is a real diagonal matrix. Note that $\mathbf{U}\mathbf{F}\mathbf{D} = \mathbf{U}\mathbf{B}\mathbf{F} \in \mathbb{R}^{\tilde{m}\lambda \times n}$ can be interpreted as the Fourier transform of the blurred target image, which is a complex matrix. Eq. (9) can be regarded as a SE w.r.t. $\mathbf{Q}^{-1}\mathbf{U}\mathbf{F}\mathbf{D}$, which has a simpler form compared to (7) as $\mathbf{\Lambda}_C$ is a diagonal matrix.

The next step in our analysis is to simplify the matrix $\mathbf{F}^H\mathbf{S}\mathbf{F}\mathbf{D}$ appearing on the left hand side of (9). First, it is important to note that the matrix $\mathbf{F}^H\mathbf{S}\mathbf{F}\mathbf{D}$ has a specific structure since all its columns contain the same blocks (see Eq. (29) in Appendix B). Using this property, by multiplying left and right by specific matrices, one obtains a block matrix whose nonzero blocks are located in its first (block) row (see (12)). More precisely, introduce the following matrix

$$\mathbf{P} = \underbrace{\begin{bmatrix} \mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{I}_m & \mathbf{I}_m & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{I}_m \end{bmatrix}}_d \quad (10)$$

whose inverse³ can be easily computed

$$\mathbf{P}^{-1} = \underbrace{\begin{bmatrix} \mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{I}_m & \mathbf{I}_m & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{I}_m \end{bmatrix}}_d.$$

Right multiplying both sides of (9) by \mathbf{P}^{-1} leads to

$$\Lambda_C \bar{\mathbf{U}} + \bar{\mathbf{U}} \mathbf{M} = \bar{\mathbf{C}}_3 \quad (11)$$

where $\bar{\mathbf{U}} = \mathbf{Q}^{-1} \mathbf{U} \mathbf{F} \mathbf{D} \mathbf{P}^{-1}$, $\mathbf{M} = \mathbf{P} (\mathbf{F}^H \underline{\mathbf{S}} \mathbf{F} \underline{\mathbf{D}}) \mathbf{P}^{-1}$ and $\bar{\mathbf{C}}_3 = \mathbf{Q}^{-1} \mathbf{C}_3 \mathbf{F} \mathbf{D} \mathbf{P}^{-1}$. Eq. (11) is a Sylvester matrix equation w.r.t. $\bar{\mathbf{U}}$ whose solution is significantly easier than for (8) because the matrix \mathbf{M} has a simple form outlined in the following lemma.

Lemma 2. *The following equality holds*

$$\mathbf{M} = \mathbf{P} (\mathbf{F}^H \underline{\mathbf{S}} \mathbf{F} \underline{\mathbf{D}}) \mathbf{P}^{-1} = \frac{1}{d} \begin{bmatrix} \sum_{i=1}^d \underline{\mathbf{D}}_i & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \quad (12)$$

where the matrix $\underline{\mathbf{D}}$ has been partitioned as follows

$$\underline{\mathbf{D}} = \begin{bmatrix} \underline{\mathbf{D}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \underline{\mathbf{D}}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \underline{\mathbf{D}}_d \end{bmatrix}$$

with $\underline{\mathbf{D}}_i$ $m \times m$ real diagonal matrices.

Proof. See Appendix B. □

³Note that left multiplying a matrix by \mathbf{P} corresponds to subtracting the first row blocks from all the other row blocks. Conversely, right multiplying by the matrix \mathbf{P}^{-1} means replacing the first (block) column by the sum of all the other (block) columns.

Finally, using the specific form of \mathbf{M} provided by Lemma 2, the solution $\bar{\mathbf{U}}$ of the SE (11) can be computed block-by-block as stated in the following theorem.

Theorem 1. Let $(\bar{\mathbf{C}}_3)_{l,j}$ denotes the j th block of the l th band of $\bar{\mathbf{C}}_3$ for any $l = 1, \dots, \tilde{m}_\lambda$. Then, the solution $\bar{\mathbf{U}}$ of the SE (11) can be decomposed as

$$\bar{\mathbf{U}} = \begin{bmatrix} \bar{\mathbf{u}}_{1,1} & \bar{\mathbf{u}}_{1,2} & \cdots & \bar{\mathbf{u}}_{1,d} \\ \bar{\mathbf{u}}_{2,1} & \bar{\mathbf{u}}_{2,2} & \cdots & \bar{\mathbf{u}}_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mathbf{u}}_{\tilde{m}_\lambda,1} & \bar{\mathbf{u}}_{\tilde{m}_\lambda,2} & \cdots & \bar{\mathbf{u}}_{\tilde{m}_\lambda,d} \end{bmatrix} \quad (13)$$

with

$$\bar{\mathbf{u}}_{l,j} = \begin{cases} (\bar{\mathbf{C}}_3)_{l,j} \left(\frac{1}{d} \sum_{i=1}^d \mathbf{D}_i + \lambda_C^l \mathbf{I}_n \right)^{-1}, & j = 1, \\ \frac{1}{\lambda_C^l} [(\bar{\mathbf{C}}_3)_{l,j} - \frac{1}{d} \bar{\mathbf{u}}_{l,1} \mathbf{D}_j], & j = 2, \dots, d. \end{cases} \quad (14)$$

Proof. See Appendix C. □

Note that $\mathbf{u}_{l,j} \in \mathbb{R}^{1 \times m}$ denotes the j th block of the l th band. Note also that the matrix $\frac{1}{d} \sum_{i=1}^d \mathbf{D}_i + \lambda_C^l \mathbf{I}_n$ appearing in the expression of $\bar{\mathbf{u}}_{l,1}$ is an $n \times n$ real diagonal matrix whose inversion is trivial. The final estimator of \mathbf{X} is obtained as follows⁴

$$\hat{\mathbf{X}} = \mathbf{H} \mathbf{Q} \bar{\mathbf{U}} \mathbf{P} \mathbf{D}^{-1} \mathbf{F}^H. \quad (15)$$

Algo. 1 summarizes the steps required to calculate the estimated image $\hat{\mathbf{X}}$.

E. Complexity Analysis

The most computationally expensive part of the proposed algorithm is the computation of matrices \mathbf{D} and $\bar{\mathbf{C}}_3$ because of the FFT and iFFT operations. Using the notation $\mathbf{C}_4 = \mathbf{Q}^{-1} (\mathbf{H}^H \mathbf{\Lambda}_R^{-1} \mathbf{H})^{-1}$, the matrix $\bar{\mathbf{C}}_3$ can be rewritten

$$\begin{aligned} \bar{\mathbf{C}}_3 &= \mathbf{C}_4 \left(\mathbf{H}^H \mathbf{\Lambda}_R^{-1} \mathbf{Y}_R (\mathbf{B}\mathbf{S})^H + (\mathbf{L}\mathbf{H})^H \mathbf{\Lambda}_L^{-1} \mathbf{Y}_L \right) \mathbf{B}\mathbf{F}\mathbf{P}^{-1} \\ &= \mathbf{C}_4 \left(\mathbf{H}^H \mathbf{\Lambda}_R^{-1} \mathbf{Y}_R \mathbf{S}^H \mathbf{F}\mathbf{D}^* + (\mathbf{L}\mathbf{H})^H \mathbf{\Lambda}_L^{-1} \mathbf{Y}_L \mathbf{F} \right) \mathbf{D}\mathbf{P}^{-1}. \end{aligned} \quad (16)$$

⁴It may happen that the diagonal matrix \mathbf{D} does not have full rank (containing zeros in diagonal) or is ill-conditioned (having very small numbers in diagonal), resulting from the property of blurring kernel. In this case, \mathbf{D}^{-1} can be replaced by $(\mathbf{D} + \tau \mathbf{I}_m)^{-1}$ for regularization purpose, where τ is a small penalty parameter [31].

Algorithm 1: Fast Fusion of Multi-band Images

Input: $\mathbf{Y}_L, \mathbf{Y}_R, \Lambda_L, \Lambda_R, \mathbf{L}, \mathbf{B}, \mathbf{S}, \mathbf{H}$
// Circulant matrix decomposition: $\mathbf{B} = \mathbf{FDF}^H$
1 $\mathbf{D} \leftarrow \text{Dec}(\mathbf{B});$
2 $\underline{\mathbf{D}} = \mathbf{D}^* \mathbf{D};$
// Calculate \mathbf{C}_1
3 $\mathbf{C}_1 \leftarrow (\mathbf{H}^H \Lambda_R^{-1} \mathbf{H})^{-1} \left((\mathbf{LH})^H \Lambda_L^{-1} \mathbf{LH} \right);$
// Eigen-decomposition of \mathbf{C}_1 : $\mathbf{C}_1 = \mathbf{Q} \Lambda_C \mathbf{Q}^{-1}$
4 $(\mathbf{Q}, \Lambda_C) \leftarrow \text{EigDec}(\mathbf{C}_1);$
// Calculate $\bar{\mathbf{C}}_3$
5 $\bar{\mathbf{C}}_3 \leftarrow \mathbf{Q}^{-1} (\mathbf{H}^H \Lambda_R^{-1} \mathbf{H})^{-1} (\mathbf{H}^H \Lambda_R^{-1} \mathbf{Y}_R (\mathbf{BS})^H + (\mathbf{LH})^H \Lambda_L^{-1} \mathbf{Y}_L) \mathbf{BFP}^{-1};$
// Calculate $\bar{\mathbf{U}}$ block by block (d blocks) and band by band (\tilde{m}_λ bands)
6 **for** $l = 1$ **to** \tilde{m}_λ **do**
 // Calculate the 1st block in l th band
7 $\bar{\mathbf{u}}_{l,1} = (\bar{\mathbf{C}}_3)_{l,1} \left(\frac{1}{d} \sum_{i=1}^d \underline{\mathbf{D}}_i + \lambda_C^l \mathbf{I}_n \right)^{-1};$
 // Calculate other blocks in l th band
8 **for** $j = 2$ **to** d **do**
9 $\bar{\mathbf{u}}_{l,j} = \frac{1}{\lambda_C^l} \left((\bar{\mathbf{C}}_3)_{l,j} - \frac{1}{d} \bar{\mathbf{u}}_{l,1} \underline{\mathbf{D}}_j \right);$
10 **end**
11 **end**
12 Set $\mathbf{X} = \mathbf{H} \mathbf{Q} \bar{\mathbf{U}} \mathbf{P} \mathbf{D}^{-1} \mathbf{F}^H;$
Output: \mathbf{X}

The most heavy step in computing (16) is the decomposition $\mathbf{B} = \mathbf{FDF}^H$ (or equivalently the FFT of the blurring kernel), which has a complexity of order $\mathcal{O}(n \log n)$. The calculations of $\mathbf{H}^H \Lambda_R^{-1} \mathbf{Y}_R \mathbf{S}^H \mathbf{F} \mathbf{D}^*$ and $(\mathbf{LH})^H \Lambda_L^{-1} \mathbf{Y}_L \mathbf{F}$ require one FFT operation each. All the other computations are made in the frequency domain. Note that the multiplication by $\mathbf{D} \mathbf{P}^{-1}$ has a cost of $\mathcal{O}(n)$ operations as \mathbf{D} is diagonal, and \mathbf{P}^{-1} reduces to block shifting and addition. The left multiplication with $\mathbf{Q}^{-1} (\mathbf{H}^H \Lambda_R^{-1} \mathbf{H})^{-1}$ is of order $\mathcal{O}(\tilde{m}_\lambda^2 n)$. Thus, the calculation of $\mathbf{C}_3 \mathbf{BFP}^{-1}$ has a total complexity of order $\mathcal{O}(n \cdot \max \{ \log n, \tilde{m}_\lambda^2 \})$.

IV. GENERALIZATION TO BAYESIAN ESTIMATORS

As mentioned in Section III-A, if the matrix $\mathbf{B} \mathbf{S} \mathbf{B}^H$ is singular or ill-conditioned (e.g., when the number of MS bands is smaller than the dimension of the subspace spanned by the pixel vectors, i.e., $n_\lambda < \tilde{m}_\lambda$), a regularization or prior information $p(\mathbf{U})$ has to be introduced to ensure the Sylvester matrix equation (11) has a unique solution. The resulting estimator \mathbf{U}

can then be interpreted as a Bayesian estimator. Combining the likelihood (4) and the prior $p(\mathbf{U})$, Bayes theorem provides the posterior distribution of \mathbf{U}

$$\begin{aligned} p(\mathbf{U}|\Psi) &\propto p(\Psi|\mathbf{U})p(\mathbf{U}) \\ &\propto p(\mathbf{Y}_L|\mathbf{U})p(\mathbf{Y}_R|\mathbf{U})p(\mathbf{U}) \end{aligned}$$

where \propto means “proportional to” and where we have used the independence between the observation vectors \mathbf{Y}_L and \mathbf{Y}_R .

The mode of the posterior distribution $p(\mathbf{U}|\Psi)$ is the so-called MAP estimator which can be obtained by solving the following optimization problem

$$\arg \min_{\mathbf{U}} L(\mathbf{U}) \quad (17)$$

where

$$\begin{aligned} L(\mathbf{U}) = &\frac{1}{2} \text{tr} \left((\mathbf{Y}_R - \mathbf{HUBS})^T \Lambda_R^{-1} (\mathbf{Y}_R - \mathbf{HUBS}) \right) + \\ &\frac{1}{2} \text{tr} \left((\mathbf{Y}_L - \mathbf{LHU})^T \Lambda_L^{-1} (\mathbf{Y}_L - \mathbf{LHU}) \right) - \log p(\mathbf{U}). \end{aligned} \quad (18)$$

Different Bayesian estimators corresponding to different choices of $p(\mathbf{U})$ have been considered in the literature. These estimators are first recalled in the next sections. We will then show that the explicit solution of the SE derived in Section III can be used to compute the MAP estimator of \mathbf{U} for these prior distributions.

A. Gaussian prior

Gaussian priors have been used widely in image processing [32]–[34], and can be interpreted as a Tikhonov regularization [35]. Assume that a matrix normal distribution is assigned *a priori* to the projected target image \mathbf{U}

$$p(\mathbf{U}) = \mathcal{MN}_{\tilde{m}_{\lambda,n}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{I}_n) \quad (19)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and covariance matrix of the matrix normal distribution. Note that the covariance matrix $\boldsymbol{\Sigma}$ explores the correlations between HS band and controls the distance between \mathbf{U} and its mean $\boldsymbol{\mu}$. Forcing the derivative of $L(\mathbf{U})$ in (17) to be zero leads to the following SE

$$\mathbf{C}_1 \mathbf{U} + \mathbf{U} \mathbf{C}_2 = \mathbf{C}_3 \quad (20)$$

where

$$\begin{aligned}
\mathbf{C}_1 &= (\mathbf{H}^H \boldsymbol{\Lambda}_R^{-1} \mathbf{H})^{-1} \left((\mathbf{LH})^H \boldsymbol{\Lambda}_L^{-1} \mathbf{LH} + \boldsymbol{\Sigma}^{-1} \right) \\
\mathbf{C}_2 &= \mathbf{BSB}^H \\
\mathbf{C}_3 &= (\mathbf{H}^H \boldsymbol{\Lambda}_R^{-1} \mathbf{H})^{-1} (\mathbf{H}^H \boldsymbol{\Lambda}_R^{-1} \mathbf{Y}_R (\mathbf{BS})^H + \\
&\quad (\mathbf{LH})^H \boldsymbol{\Lambda}_L^{-1} \mathbf{Y}_L + \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}).
\end{aligned} \tag{21}$$

The matrix \mathbf{C}_1 is positive definite as long as the covariance matrix $\boldsymbol{\Sigma}^{-1}$ is positive definite. Algo. 1 can thus be adapted to a matrix normal prior case by simply replacing \mathbf{C}_1 and \mathbf{C}_3 by their new expressions defined in (21).

B. Non-Gaussian prior

The objective function $L(\mathbf{U})$ in (17) can be split into a data term $f(\mathbf{U})$ corresponding to the likelihood and a regularization term $\phi(\mathbf{U})$ corresponding to the non-Gaussian prior in a Bayesian framework as

$$L(\mathbf{U}) = f(\mathbf{U}) + \phi(\mathbf{U}) \tag{22}$$

where

$$\begin{aligned}
f(\mathbf{U}) &= \frac{1}{2} \text{tr} \left((\mathbf{Y}_R - \mathbf{HUBS})^T \boldsymbol{\Lambda}_R^{-1} (\mathbf{Y}_R - \mathbf{HUBS}) \right) \\
&\quad + \frac{1}{2} \text{tr} \left((\mathbf{Y}_L - \mathbf{LHU})^T \boldsymbol{\Lambda}_L^{-1} (\mathbf{Y}_L - \mathbf{LHU}) \right)
\end{aligned}$$

and

$$\phi(\mathbf{U}) = -\log p(\mathbf{U}).$$

The optimization of (22) w.r.t. \mathbf{U} can be solved efficiently by using an ADMM that consists of two steps: 1) solving a surrogate optimization problem associated with a Gaussian prior and 2) applying a proximity operator [36]. This strategy can be implemented in the image domain or in the frequency domain. The resulting algorithms, named as SE-within-ADMM (SE-ADMM) methods, are described below.

1) *Solution in image domain:* Eq. (22) can be rewritten as

$$L(\mathbf{U}, \mathbf{V}) = f(\mathbf{U}) + \phi(\mathbf{V}) \text{ s.t. } \mathbf{U} = \mathbf{V}.$$

The augmented Lagrangian associated with this problem is

$$L_\mu(\mathbf{U}, \mathbf{V}, \boldsymbol{\lambda}) = f(\mathbf{U}) + \phi(\mathbf{V}) + \boldsymbol{\lambda}^T(\mathbf{U} - \mathbf{V}) + \frac{\mu}{2}\|\mathbf{U} - \mathbf{V}\|_F^2 \quad (23)$$

or equivalently

$$L_\mu(\mathbf{U}, \mathbf{V}, \mathbf{W}) = f(\mathbf{U}) + \phi(\mathbf{V}) + \frac{\mu}{2}\|\mathbf{U} - \mathbf{V} - \mathbf{W}\|_F^2 \quad (24)$$

where \mathbf{W} is the scaled dual variable. This optimization problem can be solved by an ADMM as follows

$$\begin{aligned} (\mathbf{U}^{k+1}, \mathbf{V}^{k+1}) &= \arg \min_{\mathbf{U}, \mathbf{V}} f(\mathbf{U}) + \phi(\mathbf{V}) + \\ &\quad \frac{\mu}{2}\|\mathbf{U} - \mathbf{V} - \mathbf{W}^k\|_F^2 \\ \mathbf{W}^{k+1} &= \mathbf{W}^k - (\mathbf{U}^{k+1} - \mathbf{V}^{k+1}). \end{aligned}$$

More specifically, the updates of the derived ADMM algorithm are

$$\begin{aligned} \mathbf{U}^{k+1} &= \arg \min_{\mathbf{U}} f(\mathbf{U}) + \frac{\mu}{2}\|\mathbf{U} - \mathbf{V}^k - \mathbf{W}^k\|_F^2 \\ \mathbf{V}^{k+1} &= \text{prox}_{\phi, \mu}(\mathbf{U}^{k+1} - \mathbf{W}^k) \\ \mathbf{W}^{k+1} &= \mathbf{W}^k - (\mathbf{U}^{k+1} - \mathbf{V}^{k+1}). \end{aligned} \quad (25)$$

- **Update U:** Instead of using any iterative update method, the optimization w.r.t. \mathbf{U} can be solved analytically by using Algo. 1 as for the Gaussian prior investigated in Section IV-A. For this, we can set $\boldsymbol{\mu} = \mathbf{V}^k + \mathbf{W}^k$ and $\boldsymbol{\Sigma}^{-1} = \mu \mathbf{I}_{\tilde{m}_\lambda}$ in (21). However, the computational complexity of updating \mathbf{U} in each iteration is $\mathcal{O}(n \log n)$ because of the FFT and iFFT steps required for computing $\bar{\mathbf{C}}_3$ and \mathbf{U} from $\bar{\mathbf{U}}$.
- **Update V:** The update of \mathbf{V} requires computing a proximity operator, which depends on the form of $\phi(\mathbf{V})$. When the regularizer $\phi(\mathbf{V})$ is simple enough, the proximity operator can be evaluated analytically. For example, if $\phi(\mathbf{V}) \equiv \|\mathbf{V}\|_1$, then

$$\text{prox}_{\phi, \mu}(\mathbf{U}^{k+1} - \mathbf{W}^k) = \text{soft}\left(\mathbf{U}^{k+1} - \mathbf{W}^k, \frac{1}{\mu}\right)$$

where soft is the soft-thresholding function defined as

$$\text{soft}(g, \tau) = \text{sign}(g) \max(|g| - \tau, 0).$$

More examples of proximity computations can be found in [36].

- **Update W:** The update of \mathbf{W} is simply a matrix addition whose implementation has a small computational cost.

2) *Solution in frequency domain:* Recalling that $\mathbf{B} = \mathbf{FDF}^H$, a less computationally expensive solution is obtained by rewriting $L(\mathbf{U})$ in (22) as

$$L(\mathcal{U}, \mathcal{V}) = f(\mathcal{U}) + \phi(\mathcal{V}) \text{ s.t. } \mathcal{U} = \mathcal{V}$$

where $\mathcal{U} = \mathbf{UF}$ is the Fourier transform of \mathbf{U} and

$$\begin{aligned} f(\mathcal{U}) &= \frac{1}{2} \text{tr} \left((\mathbf{Y}_R - \mathbf{H}\mathcal{U}\mathbf{D}\mathbf{F}^H\mathbf{S})^T \boldsymbol{\Lambda}_R^{-1} (\mathbf{Y}_R - \mathbf{H}\mathcal{U}\mathbf{D}\mathbf{F}^H\mathbf{S}) \right) \\ &\quad + \frac{1}{2} \text{tr} \left((\mathbf{Y}_L - \mathbf{LH}\mathcal{U}\mathbf{F}^H)^T \boldsymbol{\Lambda}_L^{-1} (\mathbf{Y}_L - \mathbf{LH}\mathcal{U}\mathbf{F}^H) \right) \end{aligned}$$

and

$$\phi(\mathcal{V}) = -\log p(\mathcal{V}).$$

Thus, the ADMM updates, defined in the image domain by (25), can be rewritten in the frequency domain as

$$\begin{aligned} \mathcal{U}^{k+1} &= \arg \min_{\mathcal{U}} f(\mathcal{U}) + \frac{\mu}{2} \|\mathcal{U} - \mathcal{V}^k - \mathcal{W}^k\|_F^2 \\ \mathcal{V}^{k+1} &= \text{prox}_{\phi, \mu}(\mathcal{U}^{k+1} - \mathcal{W}^k) \\ \mathcal{W}^{k+1} &= \mathcal{W}^k - (\mathcal{U}^{k+1} - \mathcal{V}^{k+1}). \end{aligned} \tag{26}$$

At the $(k+1)$ th ADMM iteration, updating \mathcal{U} can be efficiently conducted thanks to a SE solver similar to Algo. 1, where the matrix $\bar{\mathbf{C}}_3$ is defined by

$$\bar{\mathbf{C}}_3 = \mathbf{C}_s + \mathbf{C}_c (\mathcal{V}^k + \mathcal{W}^k) \mathbf{DP}^{-1} \tag{27}$$

with

$$\begin{aligned} \mathbf{C}_s &= \mathbf{Q}^{-1} (\mathbf{H}^H \boldsymbol{\Lambda}_R^{-1} \mathbf{H})^{-1} (\mathbf{H}^H \boldsymbol{\Lambda}_R^{-1} \mathbf{Y}_R \mathbf{S}^H \mathbf{F} \mathbf{D}^H \\ &\quad + (\mathbf{LH})^H \boldsymbol{\Lambda}_L^{-1} \mathbf{Y}_L \mathbf{F}) \mathbf{DP}^{-1} \\ \mathbf{C}_c &= \mathbf{Q}^{-1} (\mathbf{H}^H \boldsymbol{\Lambda}_R^{-1} \mathbf{H})^{-1} \boldsymbol{\Sigma}^{-1}. \end{aligned}$$

Note that the update of $\bar{\mathbf{C}}_3$ does not require any FFT computation since \mathbf{C}_s and \mathbf{C}_c can be calculated once and are not updated in the ADMM iterations.

C. Hierarchical Bayesian framework

When using a Gaussian prior, a hierarchical Bayesian framework can be constructed by introducing a hyperprior to the hyperparameter vector $\Phi = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$. Several priors have been investigated in the literature based on generalized Gaussian distributions, sparsity-promoted ℓ_1 or ℓ_0 regularizations, ℓ_2 smooth regularization, or TV regularization. Denoting as $p(\Phi)$ the prior of Φ , the optimization w.r.t. \mathbf{U} can be replaced by an optimization w.r.t. (\mathbf{U}, Φ) as follows

$$\begin{aligned} (\mathbf{U}, \Phi) &= \arg \max_{\mathbf{U}, \Phi} p(\mathbf{U}, \Phi | \Psi) \\ &= \arg \max_{\mathbf{U}, \Phi} p(\mathbf{Y}_L | \mathbf{U}) p(\mathbf{Y}_R | \mathbf{U}) p(\mathbf{U} | \Phi) p(\Phi). \end{aligned}$$

A standard way of solving this problem is to optimize alternatively between \mathbf{U} and Φ using the following updates

$$\begin{aligned} \mathbf{U}^{k+1} &= \arg \max_{\mathbf{U}} p(\mathbf{Y}_L | \mathbf{U}) p(\mathbf{Y}_R | \mathbf{U}) p(\mathbf{U} | \Phi^k) \\ \Phi^{k+1} &= \arg \max_{\Phi} p(\mathbf{U}^{k+1} | \Phi) p(\Phi). \end{aligned}$$

The update of \mathbf{U}^{k+1} can be solved using Algo. 1 whereas the update of Φ depends on the form of the hyperprior $p(\Phi)$. The derived optimization method is named as a SE-within-block coordinate descent (SE-BCD) method.

It is interesting to note that the strategy of Section IV-B proposed to handle the case of a non-Gaussian prior can be interpreted as a special case of a hierarchical updating. Indeed, if we interpret $\mathbf{V} + \mathbf{d}$ and $\frac{1}{\mu} \mathbf{I}_{\tilde{m}_\lambda}$ in (24) as the mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, the ADMM update (25) can be considered as the iterative updates of \mathbf{U} and $\boldsymbol{\mu} = \mathbf{V} + \mathbf{d}$ with fixed $\boldsymbol{\Sigma} = \frac{1}{\mu} \mathbf{I}_{\tilde{m}_\lambda}$.

V. EXPERIMENTAL RESULTS

This section applies the proposed fusion method to three kinds of priors that have been investigated in [19], [16] and [20] for the fusion of multi-band images. Note that these three methods require to solve a minimization problem similar to (17). All the algorithms have been implemented using MATLAB R2013A on a computer with Intel(R) Core(TM) i7-2600 CPU@3.40GHz and 8GB RAM. The MATLAB codes and all the simulation results are

available in the first author's homepage⁵.

A. Fusion Quality Metrics

To evaluate the quality of the proposed fusion strategy, five image quality measures have been investigated. Referring to [16], [37], we propose to use the restored signal to noise ratio (RSNR), the averaged spectral angle mapper (SAM), the universal image quality index (UIQI), the relative dimensionless global error in synthesis (ERGAS) and the degree of distortion (DD) as quantitative measures. The RSNR is defined by the negative logarithm of the distance between the estimated and reference images. The larger RSNR, the better the fusion. The definition of SAM, UIQI, ERGAS and DD can be found in [16]. The smaller SAM, ERGAS and DD, the better the fusion. The larger UIQI, the better the fusion.

B. Fusion of HS and MS images

The reference image considered here as the high-spatial and high-spectral image is a $256 \times 128 \times 93$ HS image acquired over Pavia, Italy, by the reflective optics system imaging spectrometer (ROSIS). This image was initially composed of 115 bands that have been reduced to 93 bands after removing the water vapor absorption bands. A composite color image of the scene of interest is shown in the top right of Fig. 1.

Our objective is to reconstruct the high-spatial high-spectral image \mathbf{X} from a low-spatial high-spectral HS image \mathbf{Y}_R and a high-spatial low-spectral MS image \mathbf{Y}_L . First, \mathbf{Y}_R has been generated by applying a 5×5 averaging filter and by down-sampling every 4 pixels in both vertical and horizontal directions for each band of the reference image. Second, a 4-band MS image \mathbf{Y}_L has been obtained by filtering \mathbf{X} with the LANDSAT-like reflectance spectral responses [38]. The HS and MS images are both contaminated by zero-mean additive Gaussian noises. Our simulations have been conducted with $\text{SNR}_{H,i} = 35\text{dB}$ for the first 43 bands of the HS image and $\text{SNR}_{H,i} = 30\text{dB}$ for the remaining 50 bands, with

$$\text{SNR}_{H,i} = 10 \log \left(\frac{\|(\mathbf{XBS})_i\|_F^2}{s_{H,i}^2} \right).$$

⁵<http://wei.perso.enseeiht.fr/>

For the MS image

$$\text{SNR}_{M,j} = 10 \log \left(\frac{\|(\mathbf{L}\mathbf{X})_j\|_F^2}{s_{M,j}^2} \right) = 30\text{dB}$$

for all spectral bands.

The observed HS and MS images are shown in the top left and middle figures of Fig. 1. 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 subspace transformation matrix \mathbf{H} has been defined as the PCA following the strategy of [16].

1) *Example 1: HS+MS fusion with a naive Gaussian prior:* We first consider a Bayesian fusion model initially proposed in [18]. This method assumed a naive Gaussian prior for the target image, leading to an ℓ_2 -regularization of the fusion problem. The mean of this Gaussian prior was fixed to an interpolated HS image. The covariance matrix of the Gaussian prior can be fixed *a priori* (supervised fusion) or estimated jointly with the unknown image within a hierarchical Bayesian method (unsupervised fusion). The estimator studied in [18] was based on a hybrid Gibbs sampler generating samples distributed according to the posterior of interest. An ADMM step embedded in a BCD method (ADMM-BCD) was also proposed in [19] providing a significant computational cost reduction. This section compares the performance of this ADMM-BCD algorithm with the performances of the SE-based methods for these fusion problems.

For the supervised case, the explicit solution of the SE can be constructed directly following the Gaussian prior-based generalization in Section IV-A. Conversely, for the unsupervised case, the generalized version denoted SE-BCD and described in Section IV-C is exploited, which requires embedding the closed-form solution into a BCD algorithm (refer [19] for more details). The estimated images obtained with the different algorithms are depicted in Fig. 1 and are visually very similar. More quantitative results are reported in Table II and confirm the similar performance of these methods in terms of the various fusion quality measures (RSNR, UIQI, SAM, ERGAS and DD). However, the computational time of the proposed algorithm is reduced by a factor larger than 100 (supervised) and 50 (unsupervised) due to the existence of a closed-form solution for the Sylvester matrix equation.

2) *Example 2: HS+MS fusion with a sparse representation:* This section investigates a Bayesian fusion model based on a sparse representation introduced in [16]. The basic idea of this approach was to design a prior that results from the sparse decomposition of the target image on a set of dictionaries learned empirically. Some parameters needed to be adjusted by the operator (regularization parameter, dictionaries and supports) whereas the other parameters (sparse codes) were jointly estimated with the target image. In [16], the MAP estimator associated with this model was reached using an optimization algorithm that consists of an ADMM step embedded in a BCD method (ADMM-BCD). Using the strategy proposed in Section IV-C, this ADMM step can be avoided by exploiting the SE solution. Thus, the performance of the ADMM-BCD algorithm in [16] is compared with the performance of the SE-BCD scheme as described in Section IV-C. As shown in Fig. 1 and Table II, the performance of both algorithms is quite similar. However, the proposed solution exhibits a significant complexity reduction.

3) *Example 3: HS+MS Fusion with TV regularization:* The third experiment is based on a TV regularization (can be interpreted as a specific instance of a non-Gaussian prior) studied in [20]. The regularization parameter of this method needs to be fixed by the user. The ADMM-based method investigated in [20] requires to compute a TV-based proximity operator (which increases the computational cost when compared to the previous algorithms). To solve this optimization problem, the frequency domain SE solution derived in Section IV-B2 can be embedded in an ADMM algorithm. The fusion results obtained with the ADMM method of [20] and the proposed SE-ADMM method are shown in Fig. 1 and are quite similar. Table II confirms this similarity more quantitatively by using the quality measures introduced in Section V-A. Note that the computational time obtained with the proposed explicit fusion solution is reduced when compared to the ADMM method. In order to complement this analysis, the convergence speeds of the SE-ADMM algorithm and the ADMM method of [20] are studied by analyzing the evolution of the objective function for the two fusion solutions. Fig. 2 shows that the SE-ADMM algorithm converges faster at the starting phase and gives smoother convergence result.

C. Hyperspectral Pansharpening

When $n_\lambda = 1$, the fusion of HS and MS images reduces to the HS pansharpening (HS+PAN) problem, which is the extension of conventional pansharpening (MS+PAN) and has become an important and popular application in the area of remote sensing [6]. In order to show that the proposed method is also applicable to this problem, we conducted HS and PAN image fusion for another HS dataset. The reference image, considered here as the high-spatial and high-spectral image, is an hyperspectral image acquired over Moffett field, CA, in 1994 by the JPL/NASA airborne visible/infrared imaging spectrometer (AVIRIS) [39]. This image was initially composed of 224 bands that have been reduced to 176 bands after removing the water vapor absorption bands. The HS image has been generated by applying a 5×5 averaging filter on each band of the reference image. Besides, a PAN image is obtained by successively averaging the adjacent bands in visible bands (1~41 bands) according to realistic spectral responses. In addition, the HS and PAN images have been both contaminated by zero-mean additive Gaussian noises. The SNR of the HS image is 35dB for the first 126 bands and 30dB for the last remaining bands.

Like in Section V-B, three kinds of priors are chosen to regularize this ill-posed inverse problem. The ADMM based optimization and the SE based version are compared to solve the pansharpening problem. The results are displayed in Fig. 3 whereas more quantitative results are reported in Table III. Again, the proposed SE-based method provides similar qualitative and quantitative fusion results with a significant computational cost reduction.

VI. CONCLUSION

This paper developed a fast multi-band image fusion method based on an explicit solution of a Sylvester equation. This method was applied to both the fusion of multispectral and hyperspectral images and to the fusion of panchromatic and hyperspectral images. Coupled with the alternating direction method of multipliers and block coordinate descent, the proposed algorithm can be easily generalized to compute Bayesian estimators for the fusion problem. Besides, the analytical solution can be embedded in block coordinate descent algorithm to solve the hierarchical Bayesian estimation problem. Numerical experiments showed that the proposed fast fusion method compares competitively with the ADMM based methods,

with the advantage of reducing the computational complexity significantly. Future work will consist of incorporating learning of the subspace transform matrix \mathbf{H} into the fusion scheme. Implementing the proposed fusion scheme in real datasets will also be interesting.

APPENDIX A PROOF OF LEMMA 1

As \mathbf{A}_1 is symmetric (resp. Hermitian) positive definite, \mathbf{A}_1 can be decomposed as $\mathbf{A}_1 = \mathbf{A}_1^{\frac{1}{2}} \mathbf{A}_1^{\frac{1}{2}}$, where $\mathbf{A}_1^{\frac{1}{2}}$ is also symmetric (resp. Hermitian) positive definite thus invertible. Therefore, we have

$$\mathbf{A}_1 \mathbf{A}_2 = \mathbf{A}_1^{\frac{1}{2}} \left(\mathbf{A}_1^{\frac{1}{2}} \mathbf{A}_2 \mathbf{A}_1^{\frac{1}{2}} \right) \mathbf{A}_1^{-\frac{1}{2}}. \quad (28)$$

As $\mathbf{A}_1^{\frac{1}{2}}$ and \mathbf{A}_2 are both symmetric (resp. Hermitian) matrices, $\mathbf{A}_1^{\frac{1}{2}} \mathbf{A}_2 \mathbf{A}_1^{\frac{1}{2}}$ is also a symmetric (resp. Hermitian) matrix that can be diagonalized. As a consequence, $\mathbf{A}_1 \mathbf{A}_2$ is similar to a diagonalizable matrix, and thus it is diagonalizable.

Similarly, \mathbf{A}_2 can be written as $\mathbf{A}_2 = \mathbf{A}_2^{\frac{1}{2}} \mathbf{A}_2^{\frac{1}{2}}$, where $\mathbf{A}_2^{\frac{1}{2}}$ is positive semi-definite. Thus, $\mathbf{A}_1^{\frac{1}{2}} \mathbf{A}_2 \mathbf{A}_1^{\frac{1}{2}} = \mathbf{A}_1^{\frac{1}{2}} \mathbf{A}_2^{\frac{1}{2}} \mathbf{A}_2^{\frac{1}{2}} \mathbf{A}_1^{\frac{1}{2}}$ is positive semi-definite showing that all its eigenvalues are non-negative. As similar matrices share equal similar eigenvalues, the eigenvalues of $\mathbf{A}_1 \mathbf{A}_2$ are non-negative.

APPENDIX B PROOF OF LEMMA 2

First, we introduce the following lemma.

Lemma 3. *The following equality holds*

$$\mathbf{F}^H \underline{\mathbf{S}} \mathbf{F} = \frac{1}{d} \mathbf{J}_d \otimes \mathbf{I}_m$$

where \mathbf{F} and $\underline{\mathbf{S}}$ are defined as in Section III-A, \mathbf{J}_d is the $d \times d$ matrix of ones and \mathbf{I}_m is the $m \times m$ identity matrix.

Proof. See Appendix D. □

According to Lemma 3, we have

$$\mathbf{F}^H \underline{\mathbf{S}} \mathbf{F} \underline{\mathbf{D}} = \frac{1}{d} (\mathbf{J}_d \otimes \mathbf{I}_m) \underline{\mathbf{D}} = \frac{1}{d} \begin{bmatrix} \underline{\mathbf{D}}_1 & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \\ \vdots & \vdots & \ddots & \vdots \\ \underline{\mathbf{D}}_1 & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \end{bmatrix} \quad (29)$$

Thus, multiplying (29) by \mathbf{P} on the left side and by \mathbf{P}^{-1} on the right side leads to

$$\begin{aligned} \mathbf{M} &= \mathbf{P} (\mathbf{F}^H \underline{\mathbf{S}} \mathbf{F} \underline{\mathbf{D}}) \mathbf{P}^{-1} \\ &= \frac{1}{d} \begin{bmatrix} \underline{\mathbf{D}}_1 & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \mathbf{P}^{-1} \\ &= \frac{1}{d} \begin{bmatrix} \sum_{i=1}^d \underline{\mathbf{D}}_i & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \end{aligned}$$

APPENDIX C

PROOF OF THEOREM 1

Substituting (12) and (13) into (11) leads to (31), where

$$\bar{\mathbf{C}}_3 = \begin{bmatrix} (\bar{\mathbf{C}}_3)_{1,1} & (\bar{\mathbf{C}}_3)_{1,2} & \cdots & (\bar{\mathbf{C}}_3)_{1,d} \\ (\bar{\mathbf{C}}_3)_{2,1} & (\bar{\mathbf{C}}_3)_{2,2} & \cdots & (\bar{\mathbf{C}}_3)_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ (\bar{\mathbf{C}}_3)_{d,1} & (\bar{\mathbf{C}}_3)_{d,2} & \cdots & (\bar{\mathbf{C}}_3)_{d,d} \end{bmatrix}. \quad (30)$$

Identifying the first (block) columns of (31) allows us to compute the element $\bar{\mathbf{u}}_{l,1}$ for $l = 1, \dots, d$ as follows

$$\bar{\mathbf{u}}_{l,1} = (\bar{\mathbf{C}}_3)_{l,1} \left(\frac{1}{d} \sum_{i=1}^d \underline{\mathbf{D}}_i + \lambda_C^l \mathbf{I}_n \right)^{-1}$$

for $l = 1, \dots, \tilde{m}_\lambda$. Using the values of $\bar{\mathbf{u}}_{l,1}$ determined above, it is easy to obtain $\bar{\mathbf{u}}_{l,2}, \dots, \bar{\mathbf{u}}_{l,d}$ as

$$\bar{\mathbf{u}}_{l,j} = \frac{1}{\lambda_C^l} \left[(\bar{\mathbf{C}}_3)_{l,j} - \frac{1}{d} \bar{\mathbf{u}}_{l,1} \mathbf{D}_j \right]$$

for $l = 1, \dots, \tilde{m}_\lambda$ and $j = 2, \dots, d$.

APPENDIX D PROOF OF LEMMA 3

The n dimensional DFT matrix \mathbf{F} can be written explicitly as follows

$$\mathbf{F} = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \omega^3 & \dots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \omega^6 & \dots & \omega^{2(n-1)} \\ 1 & \omega^3 & \omega^6 & \omega^9 & \dots & \omega^{3(n-1)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \omega^{3(n-1)} & \dots & \omega^{(n-1)(n-1)} \end{bmatrix}$$

where $\omega = e^{-\frac{2\pi i}{n}}$ is a primitive n th root of unity in which $i = \sqrt{-1}$. The matrix $\underline{\mathbf{S}}$ can also be written as follows

$$\underline{\mathbf{S}} = \mathbf{E}_1 + \mathbf{E}_{1+d} + \dots + \mathbf{E}_{1+(m-1)d}$$

where $\mathbf{E}_i \in \mathbb{R}^{n \times n}$ is a matrix containing only one non-zero element equal to 1 located at the i th row and i th column as follows

$$\mathbf{E}_i = \begin{bmatrix} 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{bmatrix}.$$

It is obvious that \mathbf{E}_i is an idempotent matrix, i.e., $\mathbf{E}_i = \mathbf{E}_i^2$. Thus, we have

$$\mathbf{F}^H \mathbf{E}_i \mathbf{F} = (\mathbf{E}_i \mathbf{F})^H \mathbf{E}_i \mathbf{F} = [\mathbf{0}^T \cdots \mathbf{f}_i^H \cdots \mathbf{0}^T] \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{f}_i \\ \vdots \\ \mathbf{0} \end{bmatrix} = \mathbf{f}_i^H \mathbf{f}_i$$

where $\mathbf{f}_i = \frac{1}{\sqrt{n}} [1 \ \omega^{i-1} \ \omega^{2(i-1)} \ \omega^{3(i-1)} \cdots \omega^{(n-1)(i-1)}]$ is the i th row of the matrix \mathbf{F} and $\mathbf{0} \in \mathbb{R}^{1 \times n}$ is the zero vector of dimension $1 \times n$. Straightforward computations lead to

$$\mathbf{f}_i^H \mathbf{f}_i = \frac{1}{n} \begin{bmatrix} 1 & \omega^{i-1} & \cdots & \omega^{(i-1)(n-1)} \\ \omega^{-(i-1)} & 1 & \cdots & \omega^{(i-1)(n-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \omega^{-(i-1)(n-1)} & \omega^{-(i-1)(n-2)} & \cdots & 1 \end{bmatrix}.$$

Using the ω 's property $\sum_{i=1}^n \omega^i = 0$ and $n = md$ leads to

$$\begin{aligned} & \mathbf{f}_1^H \mathbf{f}_1 + \mathbf{f}_{1+d}^H \mathbf{f}_{1+d} + \cdots + \mathbf{f}_{1+(m-1)d}^H \mathbf{f}_{1+(m-1)d} \\ &= \frac{1}{n} \begin{bmatrix} \begin{bmatrix} m & 0 & \cdots & 0 \\ 0 & m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m \end{bmatrix} & \cdots & \begin{bmatrix} m & 0 & \cdots & 0 \\ 0 & m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m \end{bmatrix} \\ \vdots & \ddots & \vdots \\ \begin{bmatrix} m & 0 & \cdots & 0 \\ 0 & m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m \end{bmatrix} & \cdots & \begin{bmatrix} m & 0 & \cdots & 0 \\ 0 & m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m \end{bmatrix} \end{bmatrix} \\ &= \frac{1}{d} \begin{bmatrix} \mathbf{I}_m & \cdots & \mathbf{I}_m \\ \vdots & \ddots & \vdots \\ \mathbf{I}_m & \cdots & \mathbf{I}_m \end{bmatrix} \\ &= \frac{1}{d} \mathbf{J}_d \otimes \mathbf{I}_m. \end{aligned}$$

REFERENCES

- [1] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan 2002.
- [2] K. Navulur, *Multispectral Image Analysis Using the Object-Oriented Paradigm*, ser. Remote Sensing Applications Series. Boca Raton, FL: CRC Press, 2006.
- [3] R. Bacon, Y. Copin, G. Monnet, B. W. Miller, J. Allington-Smith, M. Bureau, C. M. Carollo, R. L. Davies, E. Emsellem, H. Kuntschner *et al.*, "The sauron project–i. the panoramic integral-field spectrograph," *Monthly Notices of the Royal Astronomical Society*, vol. 326, no. 1, pp. 23–35, 2001.
- [4] C.-I. Chang, *Hyperspectral data exploitation: theory and applications*. New York: John Wiley & Sons, 2007.
- [5] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "25 years of pansharpening: a critical review and new developments," in *Signal and Image Processing for Remote Sensing*, 2nd ed., C. H. Chen, Ed. Boca Raton, FL: CRC Press, 2011, ch. 28, pp. 533–548.
- [6] L. Loncan, L. B. Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. Licciardi, M. Simoes, J.-Y. Tourneret, M. Veganzones, G. Vivone, Q. Wei, and N. Yokoya, "Introducing hyperspectral pansharpening," *IEEE Geosci. and Remote Sensing Mag.*, 2015, submitted.
- [7] N. Yokoya, N. Mayumi, and A. Iwasaki, "Cross-calibration for data fusion of eo-1/hyperion and terra/aster," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 419–426, 2013.
- [8] 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.
- [9] 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. and Remote Sens.*, vol. 42, no. 6, pp. 1291–1299, 2004.
- [10] 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.
- [11] 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, 1999.
- [12] 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.
- [13] Y. Zhang, A. Duijster, and P. Scheunders, "A Bayesian restoration approach for hyperspectral images," *IEEE Trans. Geosci. and Remote Sens.*, vol. 50, no. 9, pp. 3453 –3462, Sep. 2012.
- [14] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. and Remote Sens.*, vol. 50, no. 2, pp. 528–537, 2012.
- [15] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of hyperspectral and multispectral images," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Florence, Italy, May 2014.
- [16] Q. Wei, J. M. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. and Remote Sens.*, 2015, to appear.

- [17] N. Yokoya and A. Iwasaki, "Hyperspectral and multispectral data fusion mission on hyperspectral imager suite (HISUI)," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Melbourne, Australia, July 2013, pp. 4086–4089.
- [18] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multi-band images," *IEEE J. Sel. Topics Signal Process.*, 2015, to appear.
- [19] —, "Bayesian fusion of multispectral and hyperspectral images using a block coordinate descent method," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS)*, Tokyo, Japan, June 2015, submitted.
- [20] M. Simoes, J. Bioucas-Dias, L. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. and Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, June 2015.
- [21] C. P. Robert, *The Bayesian Choice: from Decision-Theoretic Motivations to Computational Implementation*, 2nd ed., ser. Springer Texts in Statistics. New York, NY, USA: Springer-Verlag, 2007.
- [22] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*, 3rd ed. Boca Raton, FL: CRC press, 2013.
- [23] M. Afonso, J. M. 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, 2011.
- [24] 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. and Remote Sens.*, vol. 36, no. 3, pp. 898–912, 1998.
- [25] J. M. Bioucas-Dias and J. M. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. and Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, 2008.
- [26] C. L. Lawson and R. J. Hanson, *Solving least squares problems*. Englewood Cliffs, NJ: Prentice-hall, 1974, vol. 161.
- [27] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1646–1658, 1997.
- [28] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," vol. 20, no. 3, pp. 21–36, 2003.
- [29] R. H. Bartels and G. Stewart, "Solution of the matrix equation $AX + XB = C$," *Communications of the ACM*, vol. 15, no. 9, pp. 820–826, 1972.
- [30] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge, UK: Cambridge university press, 2012.
- [31] R. L. Lagendijk and J. Biemond, *Iterative identification and restoration of images*. New York: Kluwer: Springer Science & Business Media, 1990, vol. 118.
- [32] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1621–1633, Dec. 1997.
- [33] M. T. Eismann and R. C. Hardie, "Application of the stochastic mixing model to hyperspectral resolution enhancement," *IEEE Trans. Geosci. and Remote Sens.*, vol. 42, no. 9, pp. 1924–1933, Sept. 2004.
- [34] N. A. Woods, N. P. Galatsanos, and A. K. Katsaggelos, "Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 201–213, Jan. 2006.
- [35] A. Tikhonov and V. Arsenin, *Solutions of ill-posed problems*, ser. Scripta series in mathematics. Winston, 1977.

- [36] P. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, ser. Springer Optimization and Its Applications, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. Springer New York, 2011, pp. 185–212.
- [37] 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.
- [38] 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.
- [39] 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.



Fig. 1. HS+MS fusion results. Top: HS image (1st column), MS image (2nd column) and reference image (3rd column). Middle and bottom: state-of-the-art-methods and corresponding proposed fast fusion methods, respectively, with various regularizations: supervised naive Gaussian prior (1st column), unsupervised naive Gaussian prior (2nd column), sparse representation (3rd column) and TV (4th column).

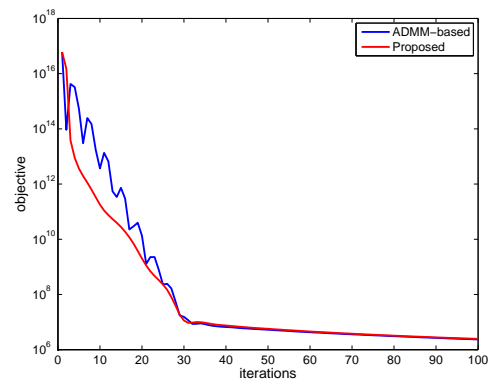


Fig. 2. Convergence speed of the ADMM [20] and the proposed SE-ADMM with the TV-regularization.

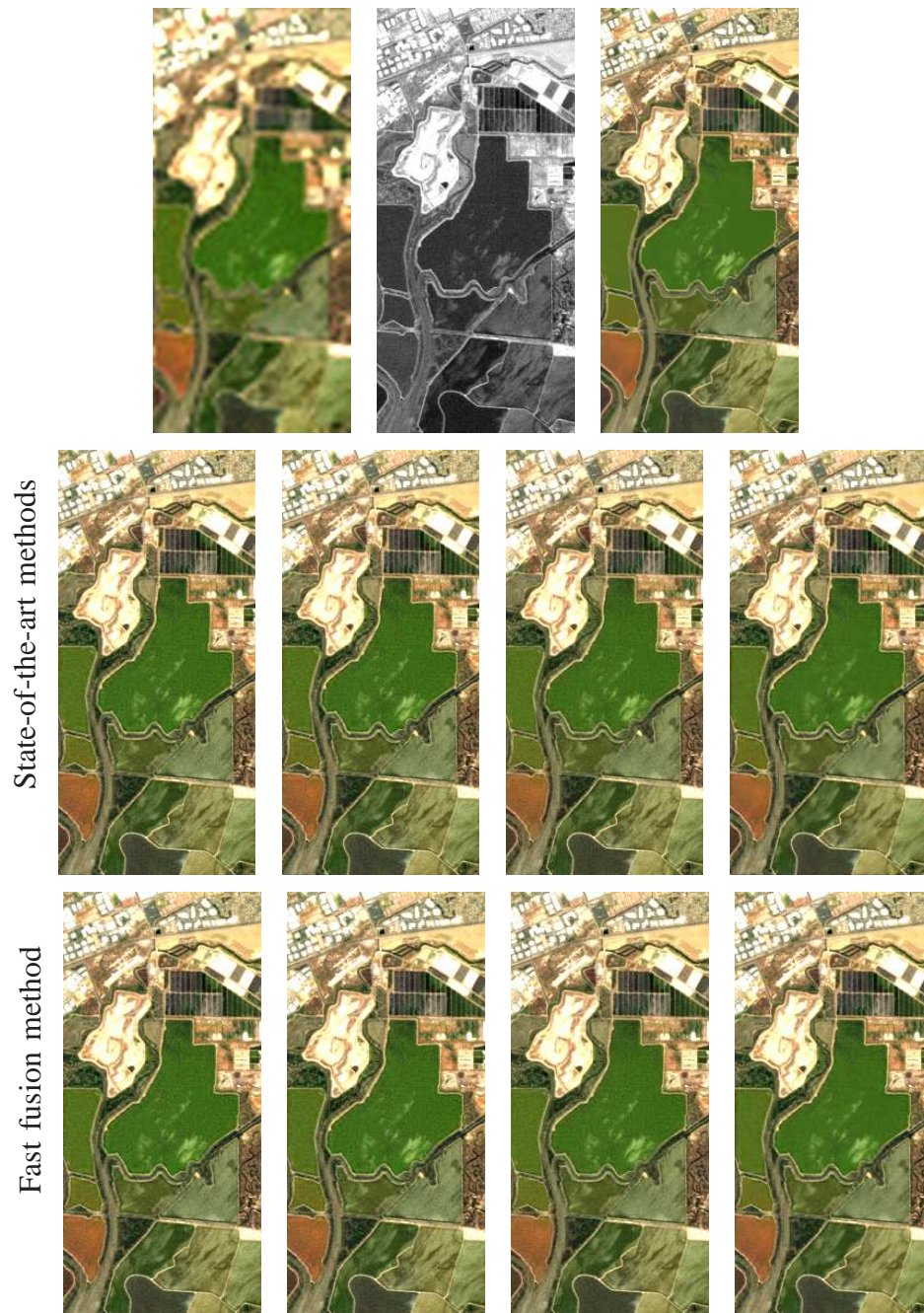


Fig. 3. Hyperspectral pansharpener results. Top: HS image (1st column), PAN image (2nd column) and reference image (3rd column). Middle and bottom: state-of-the-art-methods and corresponding proposed fast fusion methods, respectively, with various regularizations: supervised naive Gaussian prior (1st column), unsupervised naive Gaussian prior (2nd column), sparse representation (3rd column) and TV (4th column).

$$\left[\begin{array}{cccc}
\bar{\mathbf{u}}_{1,1} \left(\frac{1}{d} \sum_{i=1}^d \underline{\mathbf{D}}_i + \lambda_C^1 \mathbf{I}_n \right) & \lambda_C^1 \bar{\mathbf{u}}_{1,2} + \frac{1}{d} \bar{\mathbf{u}}_{1,1} \underline{\mathbf{D}}_2 & \cdots & \lambda_C^1 \bar{\mathbf{u}}_{1,d} + \frac{1}{d} \bar{\mathbf{u}}_{1,1} \underline{\mathbf{D}}_d \\
\bar{\mathbf{u}}_{2,1} \left(\frac{1}{d} \sum_{i=1}^d \underline{\mathbf{D}}_i + \lambda_C^2 \mathbf{I}_n \right) & \lambda_C^2 \bar{\mathbf{u}}_{2,2} + \frac{1}{d} \bar{\mathbf{u}}_{2,1} \underline{\mathbf{D}}_2 & \cdots & \lambda_C^2 \bar{\mathbf{u}}_{2,d} + \frac{1}{d} \bar{\mathbf{u}}_{2,1} \underline{\mathbf{D}}_d \\
\vdots & \vdots & \ddots & \vdots \\
\bar{\mathbf{u}}_{\tilde{m}_\lambda,1} \left(\frac{1}{d} \sum_{i=1}^d \underline{\mathbf{D}}_i + \lambda_C^{\tilde{m}_\lambda} \mathbf{I}_n \right) & \lambda_C^{\tilde{m}_\lambda} \bar{\mathbf{u}}_{\tilde{m}_\lambda,2} + \frac{1}{d} \bar{\mathbf{u}}_{\tilde{m}_\lambda,1} \underline{\mathbf{D}}_2 & \cdots & \lambda_C^{\tilde{m}_\lambda} \bar{\mathbf{u}}_{\tilde{m}_\lambda,d} + \frac{1}{d} \bar{\mathbf{u}}_{\tilde{m}_\lambda,1} \underline{\mathbf{D}}_d
\end{array} \right] = \bar{\mathbf{C}}_3 \quad (31)$$

TABLE I
NOTATIONS

Notation	Definition	Relation
m_r	row number of \mathbf{Y}_R	$m_r = n_r/d_r$
m_c	column number of \mathbf{Y}_R	$m_c = n_c/d_c$
m	number of pixels in each band of \mathbf{Y}_R	$m = m_r \times m_c$
n_r	row number of \mathbf{Y}_L	$n_r = m_r \times d_r$
n_c	column number of \mathbf{Y}_L	$n_c = m_c \times d_c$
n	number of pixels in each band of \mathbf{Y}_L	$n = n_r \times n_c$
d_r	decimation factor in row	$d_r = n_r/m_r$
d_c	decimation factor in column	$d_c = n_c/m_c$
d	decimation factor	$m = d_r \times d_c$

TABLE II
PERFORMANCE OF THE FUSION METHODS: RSNR (IN DB), UIQI, SAM (IN DEGREE), ERGAS, DD (IN 10^{-3}) AND TIME (IN SECOND).

Regularization	Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
supervised naive Gaussian	ADMM [19]	29.295	0.9906	1.556	0.892	7.121	88.58
	Proposed SE	29.366	0.9908	1.552	0.880	7.094	0.57
unsupervised naive Gaussian	ADMM-BCD [19]	29.078	0.9902	1.595	0.914	7.279	53.90
	Proposed SE-BCD	29.076	0.9902	1.624	0.914	7.369	1.07
sparse representation	ADMM-BCD [16]	29.575	0.9912	1.474	0.860	6.833	96.09
	Proposed SE-BCD	29.600	0.9913	1.472	0.856	6.820	23.72
TV	ADMM [20]	29.470	0.9911	1.503	0.861	6.923	138.98
	Proposed SE-ADMM	29.626	0.9914	1.478	0.846	6.795	93.94

TABLE III
PERFORMANCE OF THE PANSHARPENING METHODS: RSNR (IN DB), UIQI, SAM (IN DEGREE), ERGAS, DD (IN 10^{-2}) AND TIME (IN SECOND).

Regularization	Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
supervised naive Gaussian	ADMM [19]	18.637	0.9802	4.933	3.561	1.303	906.96
	Proposed SE	18.655	0.9802	4.935	3.553	1.302	0.59
unsupervised naive Gaussian	ADMM-BCD [19]	18.637	0.9802	4.933	3.561	1.304	849.14
	Proposed SE-BCD	18.652	0.9802	4.933	3.555	1.302	1.53
sparse representation	ADMM-BCD [16]	19.083	0.9818	4.691	3.379	1.227	1941.20
	Proposed SE-BCD	19.092	0.9819	4.676	3.375	1.226	116.2
TV	ADMM [20]	18.548	0.9798	4.609	3.594	1.259	295.99
	Proposed SE-ADMM	18.651	0.9803	4.571	3.553	1.240	214.75