Bayesian Fusion of Multispectral and Hyperspectral Images Using a Block Coordinate Descent Method - Complementary results and supporting materials

Qi Wei, Nicolas Dobigeon, and Jean-Yves Tourneret

TECHNICAL REPORT – 2015, March

University of Toulouse, IRIT/INP-ENSEEIHT

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

Abstract

This paper studies a new Bayesian optimization algorithm for fusing hyperspectral and multispectral images. The hyperspectral image is supposed to be obtained by blurring and subsampling a high spatial and high spectral target image. The multispectral image is modeled as a spectral mixing version of the target image. By introducing appropriate priors for parameters and hyperparameters, the fusion problem is formulated within a Bayesian estimation framework, which is very convenient to model the noise and the target image. The high spatial resolution hyperspectral image is then inferred from its posterior distribution. To compute the Bayesian maximum a posteriori estimator associated with this posterior, an alternating direction method of multipliers within block coordinate descent algorithm is proposed. Simulation results demonstrate the efficiency of the proposed fusion method when compared with several state-of-the-art fusion techniques.

Index Terms

Fusion, multispectral and hyperspectral images, Bayesian estimation, block coordinate descent, alternating direction method of multipliers

The authors are with the University of Toulouse, IRIT/INP-ENSEEIHT, 2 rue Charles Camichel, BP 7122, 31071 Toulouse cedex 7, France (e-mail: {Qi Wei, Nicolas.Dobigeon, Jean-Yves.Tourneret}@enseeiht.fr).

I. INTRODUCTION

Image fusion has been a very active research topic during recent years in remote sensing [1]. A conventional fusion problem for remote sensing images is the pansharpening, which consists of fusing a high spatial resolution panchromatic (PAN) image and a low spatial resolution multispectral (MS) image. Recently, hyperspectral (HS) imaging, which consists of acquiring a same scene in several hundreds of contiguous spectral bands, has opened a new range of relevant applications, such as target detection [2] and spectral unmixing [3]. To take advantage of the good spectral properties of HS images, the problem of fusing HS and PAN images has been explored [4]. Many existing MS pansharpening algorithms have been adapted to HS pansharpening [5], [6]. Some methods have also been specifically designed to the HS pansharpening problem such as [7]–[9]. Conversely, the fusion of MS and HS images has been considered in fewer works. It is a challenging problem since the data to be processed are of high dimensionality and both spatial and spectral information is contained in multi-band images. Note that a lot of pansharpening methods, such as component substitution [10] and relative spectral contribution [11] are inapplicable or inefficient for the HS/MS fusion problem. Since the fusion problem is generally ill-posed, Bayesian inference can offer a convenient way to regularize the problem by defining an appropriate prior distribution for the scene of interest. Following this strategy, Hardie *et al.* proposed a Bayesian estimator for fusing the co-registered high spatial-resolution MS and high spectral-resolution HS images [12]. The estimator of [12] was implemented by Zhang *et al.* in the wavelet domain to improve denoising performance [13]. However, in both of these two works, the spectral response of MS sensors is not fully exploited. More recently, a hierarchical Bayesian model, explicitly taking advantage of the MS sensor spectral response, was proposed in [14], [15]. The Bayesian estimators associated with this model were computed from samples generated from the target posterior distribution using Markov chain Monte Carlo (MCMC) methods. However, Monte Carlo based methods are quite computationally intensive, which makes the implementation time-consuming.

In this work, we propose to address the problem of fusing HS and MS images following the Bayesian framework initially proposed in [14], with an optimization method. Based on the

posterior distribution of the unknown parameters, we propose to compute the MAP estimators of the unknown scene and the noise variances by using a block coordinate descent (BCD) method [16]. This descent method includes an alternating direction method of multipliers (ADMM) step. The ADMM step differs from the gradient method by introducing variable splitting and an augmented Lagrangian, which can solve the optimization problem analytically and alternately instead of descending along gradient direction.

The paper is organized as follows. Section II formulates the fusion problem. Section III introduces the hierarchical Bayesian model of [14] defined by the joint posterior distribution of the unknown image, its hyperparameters and the noise variances. Section IV studies a BCD algorithm based on an ADMM step to maximize the joint posterior distribution of the proposed fusion model. Simulation results are presented in Section V. Conclusions are reported in Section VI.

II. PROBLEM FORMULATION

Fusing HS and MS images consists of estimating an unknown high-spatial and high-spectral resolution image from a high-spatial low-spectral MS image and a low-spatial high-spectral HS image. The HS image $\mathbf{Y}_{\mathrm{H}}$ is supposed to be a blurred down-sampled and noisy version of the target image $\mathbf{X}$ whereas the MS image $\mathbf{Y}_{\mathrm{M}}$ is a spectrally degraded and noisy version of $\mathbf{X}$. As a consequence, the observation models associated with the HS and MS images can be written as [12]

$$
\begin{aligned}
\mathbf{Y}_{\mathrm{H}} &= \mathbf{X}\mathbf{B}\mathbf{S} + \mathbf{N}_{\mathrm{H}} \\
\mathbf{Y}_{\mathrm{M}} &= \mathbf{R}\mathbf{X} + \mathbf{N}_{\mathrm{M}}
\end{aligned}
\tag{1}
$$

where $\mathbf{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n] \in \mathbb{R}^{m_\lambda \times n}$ is the unknown full resolution image composed of m_λ bands and n pixels, $\mathbf{Y}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda \times m}$ is the HS image composed of m_λ bands and m pixels and $\mathbf{Y}_{\mathrm{M}} \in \mathbb{R}^{n_\lambda \times n}$ is the MS image composed of n_λ bands and n pixels. In (1), $\mathbf{B} \in \mathbb{R}^{n \times n}$ is a cyclic convolution operator acting on the bands that models the point spread function of the HS sensor and $\mathbf{S} \in \mathbb{R}^{n \times m}$ is a downsampling matrix. The matrix $\mathbf{R} \in \mathbb{R}^{n_\lambda \times m_\lambda}$ models the spectral response of the MS sensor. In this work, the noise matrices $\mathbf{N}_{\mathrm{H}} = [\mathbf{n}_{\mathrm{H},1}, \cdots \mathbf{n}_{\mathrm{H},m}]$ and

$\mathbf{N}_{\mathrm{M}} = [\mathbf{n}_{\mathrm{M},1}, \cdots \mathbf{n}_{\mathrm{M},n}]$ are assumed to be distributed according to matrix Gaussian distributions [17]

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

where $\mathbf{I}_c$ is the $c \times c$ identity matrix, $\mathbf{0}_a$ is the $a \times 1$ vector of zeros, and the diagonal matrices $\mathbf{\Lambda}_{\mathrm{H}} = \mathrm{diag}(s_{\mathrm{H},1}^2, \cdots, s_{\mathrm{H},m_\lambda}^2) \in \mathbb{R}^{m_\lambda \times m_\lambda}$ and $\mathbf{\Lambda}_{\mathrm{M}} = \mathrm{diag}(s_{\mathrm{M},1}^2, \cdots, s_{\mathrm{M},n_\lambda}^2) \in \mathbb{R}^{n_\lambda \times n_\lambda}$ correspond to band-dependent noise variances. The fusion problem consists of estimating the high-spatial resolution HS image $\mathbf{X}$ from the two available images $\mathbf{Y}_{\mathrm{H}}$ and $\mathbf{Y}_{\mathrm{M}}$ using the observation model (1). The proposed estimation scheme relies on a hierarchical Bayesian model introduced in Section III.

III. Hierarchical Bayesian model

A. Dimension Reduction

Because the HS bands are spectrally correlated, the HS vector $\boldsymbol{x}_i$ usually lives in a space whose dimension is much smaller than m_λ [3]. More precisely, the HS image can be rewritten as $\mathbf{X} = \mathbf{VU}$ 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

$$\begin{aligned}
\mathbf{Y}_{\mathrm{H}} &= \mathbf{VUBS} + \mathbf{N}_{\mathrm{H}} \\
\mathbf{Y}_{\mathrm{M}} &= \mathbf{RVU} + \mathbf{N}_{\mathrm{M}}.
\end{aligned} \tag{2}$$

Note that $\mathbf{V}$ is a full-column rank matrix whose rows span the space $\mathbb{R}^{\widetilde{m}_\lambda \times 1}$. In this work, we assume that the signal subspace has been previously identified, e.g. obtained after conducting a principal component analysis of the HS data. Then, the considered fusion problem is solved in this lower-dimensional subspace, by estimating the projected image $\mathbf{U}$.

B. Likelihood and prior distributions

Using the statistical properties of the noise matrices $\mathbf{N}_H$ and $\mathbf{N}_M$, the distributions of $\mathbf{Y}_H$ and $\mathbf{Y}_M$ are matrix Gaussian distributions, i.e.,

$$
\begin{aligned}
\mathbf{Y}_H | \mathbf{U}, \mathbf{\Lambda}_H &\sim \mathcal{MN}_{m_\lambda, m}(\mathbf{VUBS}, \mathbf{\Lambda}_H, \mathbf{I}_m) \\
\mathbf{Y}_M | \mathbf{U}, \mathbf{\Lambda}_M &\sim \mathcal{MN}_{n_\lambda, n}(\mathbf{RVU}, \mathbf{\Lambda}_M, \mathbf{I}_n).
\end{aligned}
\tag{3}
$$

The unknown parameter vector $\boldsymbol{\theta}$ associated with (3) is composed of the projected scene $\mathbf{U}$ and the noise variances $\boldsymbol{s}^2 = \left\{ s_{H,1}^2, \cdots, s_{H,m_\lambda}^2, s_{M,1}^2, \cdots, s_{M,n_\lambda}^2 \right\}$, i.e., $\boldsymbol{\theta} = \{\mathbf{U}, \boldsymbol{s}^2\}$. Appropriate prior distributions assigned to the unknown parameters are presented below.

Scene prior: Independent Gaussian prior distributions are assigned to the projected vectors $\boldsymbol{u}_i$ ($i = 1, \cdots, n$), i.e.,

$$
\boldsymbol{u}_i | \boldsymbol{\mu}_{\boldsymbol{u}_i}, \mathbf{\Sigma} \sim \mathcal{N} \left(\boldsymbol{\mu}_{\boldsymbol{u}_i}, \mathbf{\Sigma} \right).
\tag{4}
$$

The Gaussian prior has the advantage of being a conjugate distribution relative to the likelihood function, leading to simple computations of the Bayesian estimators derived from the posterior distribution of interest and has been used successfully in many image processing applications including image denoising [18] and image restoration [19].

The means $\boldsymbol{\mu}_{\boldsymbol{u}_i}$ are fixed using the interpolated HS image in the subspace of interest following the strategy of [12] and $\mathbf{\Sigma}$ is an unknown covariance matrix. The hyperparameter $\mathbf{\Sigma}$ is related to the regularization parameter of a penalized optimization problem, which adjusts the trade-off between the data-fitting term (likelihood) and the penalty term (prior). Instead of fixing $\mathbf{\Sigma}$ *a priori*, we propose to estimate it jointly with $\mathbf{U}$ from the data by defining a hierarchical Bayesian model, which requires to define prior for this hyperparameter.

Hyperparameter prior: Assigning a conjugate inverse-Wishart (IW) distribution to the covariance matrix $\mathbf{\Sigma}$ has provided interesting results in the signal/image processing literature. Following these works, an IW distribution has been chosen, i.e.,

$$
\mathbf{\Sigma} \sim \mathcal{IW}(\mathbf{\Psi}, \eta)
\tag{5}
$$

where $(\mathbf{\Psi}, \eta)^T$ are fixed to provide a reasonable prior for $\mathbf{\Sigma}$.

Noise variance priors: Conjugate inverse-gamma distributions are chosen as prior distributions for the noise variances $s_{\text{H},i}^2$ and $s_{\text{M},j}^2$

$$
\begin{aligned}
s_{\text{H},i}^2 | \nu_{\text{H}}, \gamma_{\text{H}} &\sim \mathcal{IG} \left(\tfrac{\nu_{\text{H}}}{2}, \tfrac{\gamma_{\text{H}}}{2} \right), i = 1, \cdots, m_\lambda \\
s_{\text{M},i}^2 | \nu_{\text{M}}, \gamma_{\text{M}} &\sim \mathcal{IG} \left(\tfrac{\nu_{\text{M}}}{2}, \tfrac{\gamma_{\text{M}}}{2} \right), i = 1, \cdots, n_\lambda.
\end{aligned}
\tag{6}
$$

These conjugate distributions allow one to obtain closed-form expressions for the conditional distributions $p\left(\boldsymbol{s}^2 | \mathbf{Y}_{\text{H}}, \mathbf{Y}_{\text{M}}\right)$ of the noise variances. Other motivations for using this kind of prior distribution can be found in [20]. In this work, we assume the variances $s_{\text{H},i}^2$ and $s_{\text{M},j}^2$ are *a priori* independent since the noise properties highly depend on the sensor characteristics.

C. Posterior distribution

Defining $\mathcal{Y} = \{\mathbf{Y}_{\text{H}}, \mathbf{Y}_{\text{M}}\}$ as the set of observed images, the joint posterior distribution of the unknown parameters and hyperparameters can be computed as

$$
\begin{aligned}
p\left(\boldsymbol{\theta}, \mathbf{\Sigma} | \mathcal{Y}\right) &\propto p\left(\mathcal{Y} | \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta} | \mathbf{\Sigma}\right) p\left(\mathbf{\Sigma}\right) \\
&\propto p\left(\mathbf{Y}_{\text{H}} | \boldsymbol{\theta}\right) p\left(\mathbf{Y}_{\text{M}} | \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta} | \mathbf{\Sigma}\right) p\left(\mathbf{\Sigma}\right)
\end{aligned}
$$

where the parameter prior is

$$
p\left(\boldsymbol{\theta} | \mathbf{\Sigma}\right) = \prod_{l=1}^{n} p\left(\boldsymbol{u}_l | \mathbf{\Sigma}\right) \prod_{i=1}^{m_\lambda} p\left(s_{\text{H},i}^2\right) \prod_{j=1}^{n_\lambda} p\left(s_{\text{M},j}^2\right).
$$

In practice, it is difficult to judge the ill-posed problem is well regularized or not. If the prior or hyperprior is too flat or non-informative, the ill-posedness may remain in the regularized optimization. According to Bayesian methodology [21, Section 5.2], 'it is often practical to start with a simple, relatively non-informative prior distribution on parameters and seek to add more prior information if there remains too much variation in the posterior distribution. This adjustment can be done by tuning the parameters of prior distribution, i.e., $(\nu_{\text{H}}, \gamma_{\text{H}})$ for $s_{\text{H},i}^2$, $(\nu_{\text{M}}, \gamma_{\text{M}})$ for $s_{\text{M},i}^2$ and $(\mathbf{\Psi}, \eta)$ for $\mathbf{\Sigma}$. Generally, more informative prior can regularize the ill-posed problem better and lead to more robust while more subjective results.

The two classical estimators considered within a Bayesian estimation framework are the minimum mean square error (MMSE) and maximum a posteriori (MAP) estimators. However, for the considered fusion problem, deriving closed-form expressions for these estimators is difficult. An alternative for approximating the MMSE estimator consists of resorting to Monte Carlo integration. However, this strategy is computationally intensive due to the high dimensionality of the problem. Instead, in this work, an optimization algorithm is designed to maximize $p(\boldsymbol{\theta}, \boldsymbol{\Sigma}|\boldsymbol{\mathcal{Y}})$ providing the MAP estimator of $(\boldsymbol{\theta}, \boldsymbol{\Sigma})$. The negative logarithm of the joint posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\Sigma}|\boldsymbol{\mathcal{Y}})$ is given as

$$
\begin{aligned}
L(\mathbf{U}, \boldsymbol{s}^2, \boldsymbol{\Sigma}) = & -\log p(\boldsymbol{\theta}, \boldsymbol{\Sigma}|\boldsymbol{\mathcal{Y}}) = \\
& -\log p(\mathbf{Y}_{\mathrm{H}}|\boldsymbol{\theta}) - \log p(\mathbf{Y}_{\mathrm{M}}|\boldsymbol{\theta}) - \sum_{l=1}^{n} \log p(\boldsymbol{u}_l|\boldsymbol{\Sigma}) \\
& -\sum_{i=1}^{m_\lambda} \log p(s_{\mathrm{H},i}^2) - \sum_{j=1}^{n_\lambda} \log p(s_{\mathrm{M},j}^2) - \log p(\boldsymbol{\Sigma}) - C
\end{aligned}
\tag{7}
$$

where C is a constant. The MAP estimator of the unknown model parameters can then be obtained by minimizing the function $L(\mathbf{U}, \boldsymbol{s}^2, \boldsymbol{\Sigma})$ with respect to $\mathbf{U}$, $\boldsymbol{s}^2$ and $\boldsymbol{\Sigma}$. To solve this multivariate optimization problem, we propose to use a BCD algorithm whose details are given in the following section.

IV. BLOCK COORDINATE DESCENT METHOD

BCD consists of optimizing with respect to (w.r.t.) the unknown parameters iteratively, which can be easily implemented in the considered fusion problem (see Algorithm 1). Contrary to gradient based optimization methods, BCD does not require any stepsize tuning, which makes the algorithm more usable by practitioners. BCD is known to converge to a stationary point of the target cost function to be optimized provided that this target function has a unique minimum point with respect to each variable [16, Prop. 2.7.1], which is the case for the criterion in (7). The *three* steps of the BCD algorithm are detailed below.

Algorithm 1: Block coordinated descent algorithm

Input: $\mathbf{Y}_H$, $\mathbf{Y}_M$, $\widetilde{m}_\lambda$, $\mathbf{B}$, $\mathbf{S}$, $\mathbf{R}$, s_0^2, $\boldsymbol{\Sigma}_0$

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

2 $\quad$ $\mathbf{U}_t = \arg\min_{\mathbf{U}} L(\mathbf{U}, s_{t-1}^2, \boldsymbol{\Sigma}_{t-1})$; $\qquad\qquad$ `/* See Section IV-A */`

3 $\quad$ $s_t^2 = \arg\min_{s^2} L(\mathbf{U}_t, s^2, \boldsymbol{\Sigma}_{t-1})$; $\qquad\qquad$ `/* See Section IV-B */`

4 $\quad$ $\boldsymbol{\Sigma}_t = \arg\min_{\boldsymbol{\Sigma}} L(\mathbf{U}_t, s_t^2, \boldsymbol{\Sigma})$; $\qquad\qquad\quad$ `/* See Section IV-C */`

5 end

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

A. Optimization with respect to U

The optimization w.r.t. to $\mathbf{U}$ consists of minimizing

$$
\begin{aligned}
L_{\mathbf{U}}(\mathbf{U}) = \tfrac{1}{2}\|\boldsymbol{\Lambda}_H^{-\frac{1}{2}}\left(\mathbf{Y}_H - \mathbf{V}\mathbf{U}\mathbf{B}\mathbf{S}\right)\|_F^2 + \\
\tfrac{1}{2}\|\boldsymbol{\Lambda}_M^{-\frac{1}{2}}\left(\mathbf{Y}_M - \mathbf{R}\mathbf{V}\mathbf{U}\right)\|_F^2 + \tfrac{1}{2}\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\mathbf{U} - \boldsymbol{\mu}_{\mathbf{U}}\right)\|_F^2.
\end{aligned}
\tag{8}
$$

Determining $\mathbf{U}$ which makes the gradient of $L_{\mathbf{U}}(\mathbf{U})$ equal to zero is not straightforward, mainly due to left- and right-side linear operators applied to $\mathbf{U}$ and the size of the matrices involved in the computation. In fact, it is equivalent to solve the following Sylvester equation [22]

$$
\begin{aligned}
\mathbf{V}^T\boldsymbol{\Lambda}_H^{-1}\mathbf{V}\mathbf{U}\mathbf{B}\mathbf{S}\left(\mathbf{B}\mathbf{S}\right)^T + \left(\left(\mathbf{R}\mathbf{V}\right)^T\boldsymbol{\Lambda}_M^{-1}\mathbf{R}\mathbf{V} + \boldsymbol{\Sigma}^{-1}\right)\mathbf{U} \\
= \mathbf{V}^T\boldsymbol{\Lambda}_H^{-1}\mathbf{Y}_H\left(\mathbf{B}\mathbf{S}\right)^T + \left(\mathbf{R}\mathbf{V}\right)^T\boldsymbol{\Lambda}_M^{-1}\mathbf{Y}_M + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{\mathbf{U}}.
\end{aligned}
\tag{9}
$$

As the Schur decomposition of matrix $\mathbf{B}\mathbf{S}\left(\mathbf{B}\mathbf{S}\right)^T$ is not easy achieved, the analytic solution of the Sylvester equation is almost not solvable.

Fortunately, this kind of optimization problem has been solved efficiently by the ADMM method [23]. After defining 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 (8) can be

written as

$$L_{\mathbf{U}}(\mathbf{U}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3) =$$

$$\frac{1}{2}\left\|\mathrm{H}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{V}\mathbf{V}_1\mathbf{S})\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U}\mathbf{B} - \mathbf{V}_1 - \mathbf{G}_1\right\|_F^2 +$$

$$\frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{R}\mathbf{V}\mathbf{V}_2)\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U} - \mathbf{V}_2 - \mathbf{G}_2\right\|_F^2 +$$

$$\frac{1}{2}\left\|\mathbf{\Sigma}^{-\frac{1}{2}}\left(\boldsymbol{\mu}_{\mathbf{U}} - \mathbf{V}_3\right)\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U} - \mathbf{V}_3 - \mathbf{G}_3\right\|_F^2.$$

The iterative update of $\mathbf{U}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$ can be achieved with the split augmented Lagrangian shrinkage algorithm (SALSA) [24], [25], which is an instance of the ADMM algorithm with convergence guaranty. The detailed implementation of the SALSA scheme for the fusion problem are summarized in Algorithm 2.

Algorithm 2: SALSA step

---

1 **Define:** $\mathbf{Ds} \in \{0, 1\}^n$ such that $\mathbf{Ds}(i) = \begin{cases} 1 & \text{if pixel } i \text{ is sampled,} \\ 0 & \text{otherwise;} \end{cases}$

2 **Initialization**: choose $\mu > 0$, $\mathbf{U}^{(0)}, \mathbf{V}_1^{(0)}, \mathbf{V}_1^{(0)}, \mathbf{V}_1^{(0)}, \mathbf{V}_1^{(0)}, \mathbf{D}_1^{(0)}, \mathbf{D}_2^{(0)}, \mathbf{D}_3^{(0)}$

3 **for** $k = 0, 1, \dots$ **to** *stopping rule* **do**

    /* optimize w.r.t $\mathbf{U}$ (light with FFT)   */

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

    /* optimize w.r.t $\mathbf{V}_1$ */

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

    /* Update $\mathbf{V}_1$ according to downsampling   */

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

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

    /* optimize w.r.t $\mathbf{V}_2$ */

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

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

    /* optimize w.r.t $\mathbf{V}_3$ */

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

11    $\mathbf{V}_3^{(k+1)} \leftarrow \left(\mathbf{\Sigma}^{-1} + \mu\mathbf{I}\right)^{-1} \left(\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_{\mathbf{U}} + \mu\boldsymbol{\nu}_3\right)$;

    /* update Lagrange multipliers   */

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

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

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

15 **end**

B. Optimization with respect to s^2

The optimization w.r.t. s^2 is decomposed into $(m_\lambda + n_\lambda)$ parallel optimizations w.r.t. $\left\{ s_{\mathrm{H},j}^2 \right\}_{j=1}^{m_\lambda}$ and $\left\{ s_{\mathrm{M},j}^2 \right\}_{j=1}^{n_\lambda}$ thanks to the criterion separability

$$
\begin{aligned}
L_{\boldsymbol{s}^2}(\boldsymbol{s}^2) = \\
\left(\tfrac{\nu_{\mathrm{H}}+n}{2} + 1 \right) \sum_{i=1}^{m_\lambda} \log s_{\mathrm{H},i}^2 + \sum_{i=1}^{m_\lambda} \tfrac{\gamma_{\mathrm{H}}+\|(\mathbf{Y}_{\mathrm{H}}-\mathbf{V}\mathbf{U}_{t-1}\mathbf{B}\mathbf{S})_i\|_F^2}{2s_{\mathrm{H},i}^2} \\
\left(\tfrac{\nu_{\mathrm{M}}+n}{2} + 1 \right) \sum_{j=1}^{n_\lambda} \log s_{\mathrm{M},j}^2 + \sum_{j=1}^{n_\lambda} \tfrac{\gamma_{\mathrm{M}}+\|(\mathbf{Y}_{\mathrm{M}}-\mathbf{R}\mathbf{V}\mathbf{U}_{t-1})_j\|_F^2}{2s_{\mathrm{M},j}^2}.
\end{aligned}
$$

Computing the derivatives of $L_{\boldsymbol{s}^2}(\boldsymbol{s}^2)$ w.r.t. $s_{\mathrm{H},i}^2$ and $s_{\mathrm{M},j}^2$ and forcing them to be zero leads to the update rules

$$
\begin{aligned}
s_{\mathrm{H},i}^2 &= \tfrac{1}{\nu_{\mathrm{H}}+n+2} \left(\gamma_{\mathrm{H}} + \| \left(\mathbf{Y}_{\mathrm{H}} - \mathbf{V}\mathbf{U}_{t-1}\mathbf{B}\mathbf{S} \right)_i \|_F^2 \right) \\
s_{\mathrm{M},j}^2 &= \tfrac{1}{\nu_{\mathrm{M}}+n+2} \left(\gamma_{\mathrm{M}} + \| \left(\mathbf{Y}_{\mathrm{M}} - \mathbf{R}\mathbf{V}\mathbf{U}_{t-1} \right)_j \|_F^2 \right).
\end{aligned}
$$

C. Optimization with respect to $\boldsymbol{\Sigma}$

Fixing $\mathbf{U}$ and $\boldsymbol{s}^2$, the objective function is

$$
\begin{aligned}
L_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}) &= \tfrac{\eta+\widetilde{m}_\lambda+n+1}{2} \log|\boldsymbol{\Sigma}| \\
&\quad + \tfrac{1}{2}\mathrm{tr}\left(\left(\sum_{i=1}^n \left(\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i} \right) \left(\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i} \right)^T + \boldsymbol{\Psi} \right) \boldsymbol{\Sigma}^{-1} \right)
\end{aligned}
$$

where $\mathrm{tr}(\cdot)$ is the trace operator. The maximum of this function is obtained for

$$
\boldsymbol{\Sigma}_t = \frac{\left(\mathbf{U}_{t-1} - \boldsymbol{\mu}_{\mathbf{U}} \right) \left(\mathbf{U}_{t-1} - \boldsymbol{\mu}_{\mathbf{U}} \right)^T + \boldsymbol{\Psi}}{\eta + \widetilde{m}_\lambda + n + 1}.
$$

D. Relationship with the MCMC method of [14]

It is worthy to note that the proposed optimization procedure is structured similarly to the Gibbs sampler developed in [14] to solve the fusion problem. Indeed, the BCD method can be interpreted as a deterministic counterpart of the Gibbs sampler, consisting of replacing the stochastic sampling procedures according to the conditional posterior distributions of the target distribution by iterative evaluations of their modes. However, the BCD method requires much fewer computation resources when compared with Monte Carlo-based methods, which is crucial for practical implementations.

V. Simulation results

This section presents numerical results obtained with the proposed fusion algorithm. The reference image, considered here as the high spatial and high spectral resolution image to be recovered, is an HS image acquired over Moffett field, CA, in 1994 by the JPL/NASA airborne visible/infrared imaging spectrometer (AVIRIS). This image is of size 128×64 and was initially composed of 224 bands that have been reduced to 177 bands after removing the water vapor absorption bands. A composite color image of the scene of interest is shown in the top right panel of Fig. 2.

A. Simulation scenario

We propose to reconstruct the reference HS image $\mathbf{X}$ from two HS and MS images $\mathbf{Y}_\mathrm{H}$ and $\mathbf{Y}_\mathrm{M}$. First, a high-spectral and low-spatial resolution image $\mathbf{Y}_\mathrm{H}$ has been generated by applying a 5×5 averaging filter and by down-sampling every 4 pixels in both vertical and horizontal direction for each band of $\mathbf{X}$. Second, a 7-band MS image $\mathbf{Y}_\mathrm{M}$ has been obtained by filtering $\mathbf{X}$ with the LANDSAT-like reflectance spectral responses [26]. The HS and MS images are both contaminated by additive centered Gaussian noises. The simulations have been conducted with $\mathrm{SNR}_{\mathrm{H},j} = 35\mathrm{dB}$ for the first 127 bands and $\mathrm{SNR}_{\mathrm{H},j} = 30\mathrm{dB}$ for the remaining 50 bands of the HS image, where $\mathrm{SNR}_{\mathrm{H},j} = 10 \log \left(\frac{\|[\mathbf{XBS}]_j\|_F^2}{s_{\mathrm{H},j}^2} \right)$. For the MS image, the noise level has been adjusted to obtain $\mathrm{SNR}_{\mathrm{M},j} = 10 \log \left(\frac{\|[\mathbf{RX}]_j\|_F^2}{s_{\mathrm{M},j}^2} \right) = 30\mathrm{dB}$ in all the spectral bands. The observed HS and MS images are shown in the top left and right of Fig. 2 (note that the HS image has been interpolated for better visualization and that the MS image has been displayed using an arbitrary color composition). To learn the projection matrix $\mathbf{V}$, a PCA has been conducted, i.e., the $\widetilde{m}_\lambda = 10$ most discriminant vectors associated with the 10 largest eigenvalues of the sample covariance matrix of the HS image have been computed. These 10 vectors lead to 99.89% of the information contained in the HS image.

B. Hyperparameter Selection

Similarly to the Monte Carlo-based approach in [14], the proposed algorithm requires to tune the hyperparameters γ_H and ν_H (γ_M and ν_M, resp.) associated with the HS (MS, resp.) noise variance priors, and $\boldsymbol{\Psi}$ and η associated with the image covariance matrix prior. The strategy adopted to adjust these hyperparameters is detailed in what follows.

The noise variance hyperparameters can be tuned based on some prior knowledge. Indeed, since the $\text{SNR}_{H,i}$ are approximately known in practice, rough estimations of $s_{H,i}^2$ can be obtained as $\tilde{s}_{H,i}^2 = \frac{\|(\mathbf{Y}_H)_i\|_F^2}{10^{\text{SNR}_{H,i}/10}}$. Following a 1st-order moment technique, these rough estimations can be used to adjust the means of the prior in (6), leading to $\gamma_H = (\nu_H - 2)\,\tilde{s}_{H,i}^2$. A similar strategy is adopted to define γ_M.

The values of ν_H and ν_M are related to the variances of the inverse-gamma distributions (6), since $\text{var}\left[s_{\cdot,i}^2\right] = \frac{2\gamma_{\cdot}^2}{(\nu_{\cdot}-2)^2(\nu_{\cdot}-4)}$. It has been observed that assigning non-informative priors for both noise variances of the HS and MS images (i.e., fixing ν_H and ν_M to both small values) lead to poor fusion results, since these priors are not sufficiently informative to regularize the ill-posed problem. In this work, informative priors for $s_{M,j}^2$ have been chosen by fixing $\nu_M = 10n$ and flat priors for $s_{H,j}^2$ have been chosen by fixing $\nu_H = 3$.

The prior distribution assigned to $\boldsymbol{\Sigma}$ depends on the parameters $\boldsymbol{\Psi}$ and η. As there is no available prior knowledge about $\boldsymbol{\Sigma}$, the two parameters η and $\boldsymbol{\Psi}$ have been chosen in order to obtain a non-informative prior for $\boldsymbol{\Sigma}$. More precisely, $\eta = \widetilde{m}_\lambda + 3$ and $\boldsymbol{\Psi} = (\eta + \widetilde{m}_\lambda + 1)\,\mathbf{I}_{\widetilde{m}_\lambda}$.

C. Fusion performance

To evaluate the quality of the proposed fusion strategy, three image quality measures have been investigated. Referring to [13], [14], the root mean square error (RMSE), the averaged spectral angle mapper (SAM) and the universal image quality index (UIQI) are used as quantitative measures. The RMSE is defined by the distance between the estimated and reference images, while the definitions of SAM and UIQI can be found in [13]. The smaller RMSE and SAM, the better the fusion. The larger UIQI, the better the fusion.

The experiments compare the proposed algorithm with three state-of-the-art fusion algorithms

[12]–[14]. Note that the fusion method in [14] can be considered as the Monte Carlo-based counterpart of the proposed method, since both methods share the same hierarchical Bayesian model. Results obtained with these algorithms are depicted in Fig. 2 and quantitative results are reported in Table I. These results show that the proposed method provides better results than the methods of [12], [13] and competitive results when compared with the method in [14]. However, as observed by comparing the execution times reported in Table I, the proposed optimization algorithm is significantly faster than the method of [14].

The estimation of noise variances for both HS bands and MS bands are shown in Fig. 3. These results show that the noise variances for different bands can be tracked with tolerant discrepancy.

VI. CONCLUSION

This paper proposed to maximize the posterior distribution associated with a hierarchical Bayesian model for fusing multispectral and hyperspectral images using a block coordinate descent (BCD) method. The high spatial and high spectral resolution image to be recovered was defined in a lower-dimensional subspace, identified by a PCA applied to the hyperspectral image. The joint optimization was conducted iteratively with respect to the image to be recovered, the noise variances and the image prior covariance matrix. One particularity of the proposed BCD algorithm was to involve an ADMM step for estimating the unknown image. Numerical experiments showed that the proposed method compares competitively with other state-of-the-art methods, with the great advantage of reducing the computational complexity when compared with a Monte Carlo-based counterpart method. It is interesting to note that recently the proposed framework has been successfully used to incorporate a sparse prior [27]. A related acceleration to achieve fast fusion is also noteworthy [28].

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

[3] 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 Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, 2012.

[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] M. Moeller, T. Wittman, and A. L. Bertozzi, "A variational approach to hyperspectral image fusion," in *Proc. SPIE Defense, Security, and Sensing*.   International Society for Optics and Photonics, 2009, p. 73341E.

[6] 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.

[7] G. A. Licciardi, M. M. Khan, J. Chanussot, A. Montanvert, L. Condat, and C. Jutten, "Fusion of hyperspectral and panchromatic images using multiresolution analysis and nonlinear PCA band reduction," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, pp. 1–17, 2012.

[8] 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, pp. 1–22, May 2012.

[9] 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. Remote Sens. Mag.*, submitted.

[10] 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.

[11] 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.

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

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

[14] 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.

[15] ——, "Bayesian fusion of multi-band images," *IEEE J. Sel. Topics Signal Process.*, to appear.

[16] D. P. Bertsekas, *Nonlinear programming*.   Athena Scientific, 1999.

[17] A. P. Dawid, "Some matrix-variate distribution theory: notational considerations and a Bayesian application," *Biometrika*, vol. 68, no. 1, pp. 265–274, 1981.

[18] P. J. Liu, "Using Gaussian process regression to denoise images and remove artefacts from microarray data," Ph.D. dissertation, University of Toronto, 2007.

[19] 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.

[20] A. Gelman, "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)," *Bayesian analysis*, vol. 1, no. 3, pp. 515–534, 2006.

[21] 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.

[22] 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.

[23] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

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

[25] 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.

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

[28] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Trans. Image Process.*, submitted.
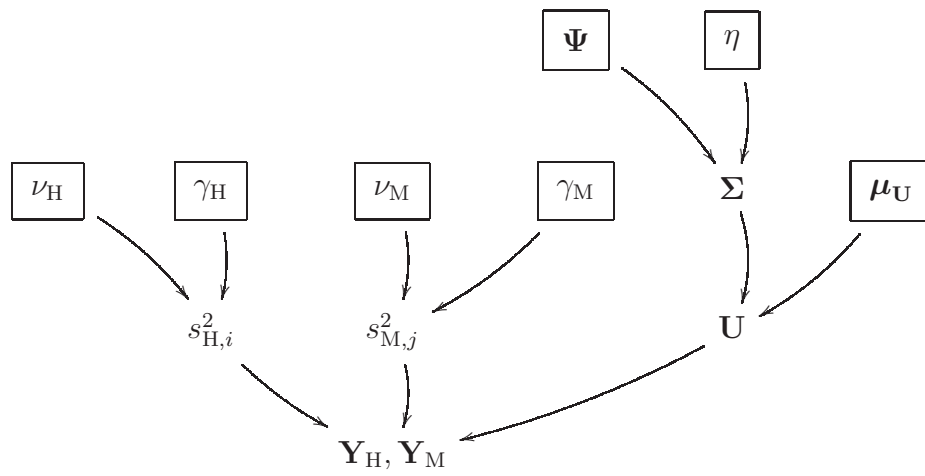
Fig. 1. DAG for the parameter and hyperparameter priors (the fixed parameters appear in boxes).
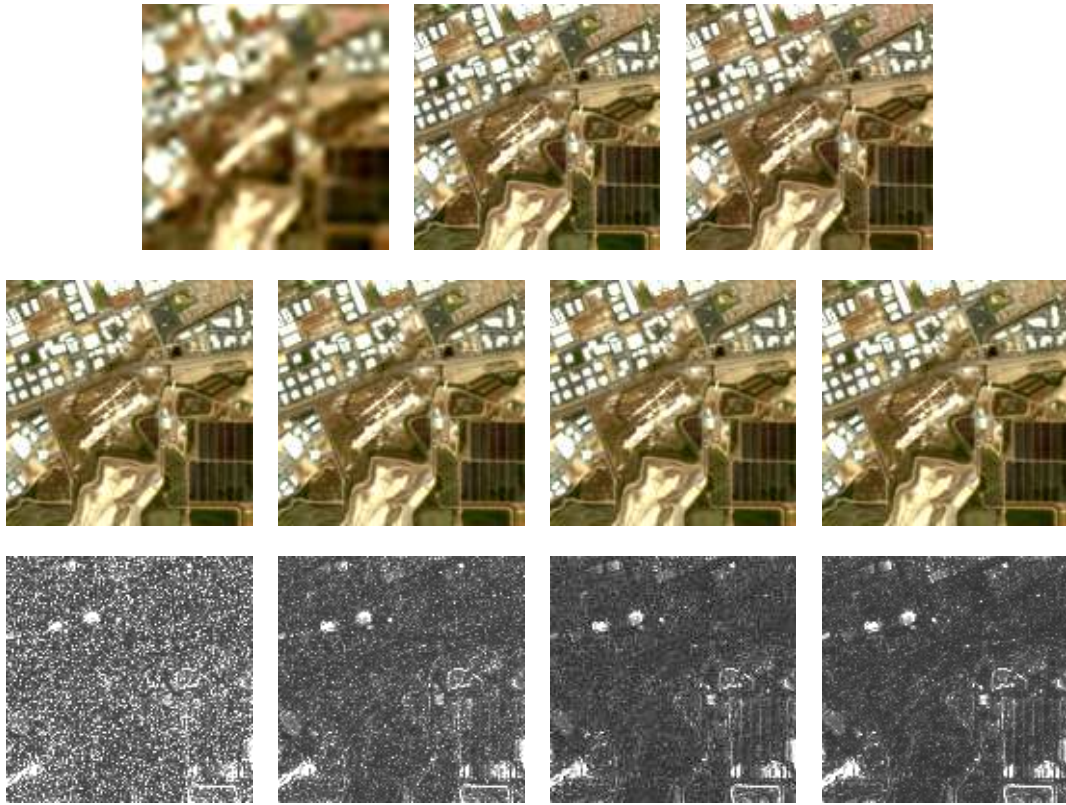
Fig. 2. Fusion results. Top, left: HS image. Top, middle: MS image. Top, right: Reference image. Middle, 1: MAP estimator [12]. Middle, 2: Wavelet MAP estimator [13]. Middle, 3: MMSE estimator. Middle, 4: Proposed method. Bottom: The corresponding RMSE maps (More black, smaller errors; more white, larger errors).
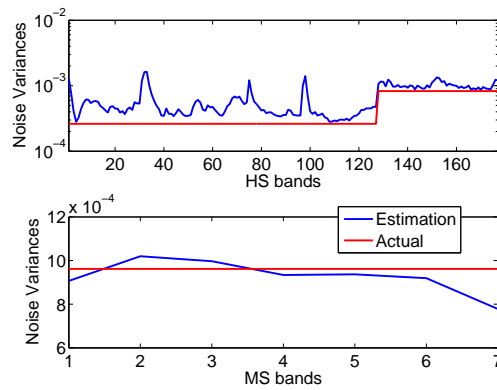

Fig. 3. Noise variances and their MMSE estimates. Top: HS Image ($SNR_2 = 30$dB). Bottom: MS Image ($SNR_1 = 30$dB).

TABLE I
PERFORMANCE OF THE FUSION METHODS: RSNR ($\times 10^{-2}$), UIQI, SAM ($°$) AND TIME (SECOND).

Methods	RMSE	UIQI	SAM	Time
Hardie [12]	6.96	0.9932	5.15	3
Zhang [13]	5.68	0.9956	4.22	72
MCMC [14]	5.06	0.9971	3.73	6228
Proposed	5.10	0.9971	3.74	96