Bayesian Fusion of Multi-Band Images - 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-ENSEEIHT

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

Abstract

In this paper, a Bayesian fusion technique for remotely sensed multi-band images is presented. The observed images are related to the high spectral and high spatial resolution image to be recovered through physical degradations, e.g., spatial and spectral blurring and/or subsampling defined by the sensor characteristics. The fusion problem is formulated within a Bayesian estimation framework. An appropriate prior distribution exploiting geometrical consideration is introduced. To compute the Bayesian estimator of the scene of interest from its posterior distribution, a Markov chain Monte Carlo algorithm is designed to generate samples asymptotically distributed according to the target distribution. To efficiently sample from this high-dimension distribution, a Hamiltonian Monte Carlo step is introduced in the Gibbs sampling strategy. The efficiency of the proposed fusion method is evaluated with respect to several state-of-the-art fusion techniques. In particular, low spatial resolution hyperspectral and multispectral images are fused to produce a high spatial resolution hyperspectral image.

Index Terms

Part of this work has been supported by 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. Part of this work was presented during the IEEE ICASSP 2014 [1].

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

Fusion, super-resolution, multispectral and hyperspectral images, deconvolution, Bayesian estimation, Hamiltonian Monte Carlo algorithm.

I. Introduction

The problem of fusing a high spatial and low spectral resolution image with an auxiliary image of higher spectral but lower spatial resolution, also known as multi-resolution image fusion, has been explored for many years [2]. When considering remotely sensed images, an archetypal fusion task is the pansharpening, which generally consists of fusing a high spatial resolution panchromatic (PAN) image and low spatial resolution multispectral (MS) image. Pansharpening has been addressed in the image processing and remote sensing literatures for several decades and still remains an active topic [2]–[7]. More 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[8], such as target detection [9], classification [8] and spectral unmixing [10]. The visualization of HS images is also interesting to be explored [11]. Naturally, to take advantage of the newest benefits offered by HS images, the problem of fusing HS and PAN images has been explored [12]–[14]. Capitalizing on decades of experience in MS pansharpening, most of the HS pansharpening approaches merely adapt existing algorithms for PAN and MS fusion [15], [16]. Other methods are specifically designed to the HS pansharpening problem (see, e.g., [13], [17], [18]). Conversely, the fusion of MS and HS images has been considered in fewer research works and is still a challenging problem because of the high dimension of the data to be processed. Indeed, the fusion of MS and HS differs from traditional MS or HS pansharpening by the fact that more spatial and spectral information is contained in multi-band images. This additional information can be exploited to obtain a high spatial and spectral resolution image. In practice, the spectral bands of panchromatic images always cover the visible and infrared spectra. However, in several practical applications, the spectrum of MS data includes additional high-frequency spectral bands. For instance the MS data of WorldView-3 have spectral bands in the intervals $[400 \sim 1750]$nm and $[2145 \sim 2365]$nm whereas the PAN data are in the range $[450 \sim 800]$nm [19]. Another interesting example is the HS+MS suite (called Hyperspectral imager suite (HISUI)) that has been developed by the Japanese Ministry of

Economy, Trade, and Industry (METI) [20]. HISUI is the Japanese next-generation Earth-observing sensor composed of HS and MS imagers and will be launched by H-IIA rocket in 2015 or later as one of mission instruments onboard JAXA's ALOS-3 satellite. Some research activities have already been conducted for this practical multi-band fusion problem [21]. Noticeably, a lot of pansharpening methods, such as component substitution [22], relative spectral contribution [23] and high-frequency injection [24] are inapplicable or inefficient for the HS+MS fusion problem. To address the challenge raised by the high dimensionality of the data to be fused, innovative methods need to be developed. This is the main objective of this paper.

As demonstrated in [25], [26], the fusion of HS and MS images can be conveniently formulated within a Bayesian inference framework. Bayesian fusion allows an intuitive interpretation of the fusion process via the posterior distribution. Since the fusion problem is usually ill-posed, the Bayesian methodology offers a convenient way to regularize the problem by defining appropriate prior distribution for the scene of interest. Following this strategy, Hardie *et al.* proposed a Bayesian estimator for fusing co-registered high spatial-resolution MS and high spectral-resolution HS images [25]. To improve the denoising performance, Zhang *et. al* implemented the estimator of [25] in the wavelet domain [26]. In [27], Zhang *et al.* derived an expectation-maximization (EM) algorithm to maximize the posterior distribution of the unknown image via deblurring and denoising steps. An HS and MS image fusion strategy based on spectral unmixing has also been explored recently [28]

In this paper, a prior knowledge accounting for artificial constraints related to the fusion problem is incorporated within the model via the prior distribution assigned to the scene to be estimated. Many strategies related to HS resolution enhancement have been proposed to define this prior distribution. For instance, in [5], the highly resolved image to be estimated is *a priori* modeled by an in-homogeneous Gaussian Markov random field (IGMRF). The parameters of this IGMRF are empirically estimated from a panchromatic image in the first step of the analysis. In [25] and related works [29], [30], a multivariate Gaussian distribution is proposed as prior distribution for the unobserved scene. The resulting conditional mean and covariance matrix can then be inferred using a standard clustering technique [25] or using a stochastic mixing model [29], [30], incorporating spectral mixing constraints to

improve spectral accuracy in the estimated high resolution image. In this paper, we propose to explicitly exploit the acquisition process of the different images. More precisely, the sensor specifications (i.e., spectral or spatial responses) are exploited to properly design the spatial or spectral degradations suffered by the image to be recovered [31]. Moreover, to define the prior distribution assigned to this image, we resort to geometrical considerations well admitted in the HS imaging literature devoted to the linear unmixing problem [32]. In particular, the high spatial resolution HS image to be estimated is assumed to live in a lower dimensional subspace, which is a suitable hypothesis when the observed scene is composed of a finite number of macroscopic materials.

Within a Bayesian estimation framework, two statistical estimators are generally considered. The minimum mean square error (MMSE) estimator is defined as the mean of the posterior distribution. Its computation generally requires intractable multidimensional integrations. Conversely, the maximum a posteriori (MAP) estimator is defined as the mode of the posterior distribution and is usually associated with a penalized maximum likelihood approach. Mainly due to the complexity of the integration required by the computation of the MMSE estimator (especially in high-dimension data space), most of the Bayesian estimators have proposed to solve the HS and MS fusion problem using a MAP formulation [25], [26], [33]. However, optimization algorithms designed to maximize the posterior distribution may suffer from the presence of local extrema, that prevents any guarantee to converge towards the actual maximum of the posterior.

In this paper, we propose to compute the MMSE estimator of the unknown scene by using samples generated by a Markov chain Monte Carlo (MCMC) algorithm. The posterior distribution resulting from the proposed forward model and the *a priori* modeling is defined in a high dimensional space, which makes difficult the use of any conventional MCMC algorithm, e.g., the Gibbs sampler [34] or the Metropolis-Hastings sampler [35]. To overcome this difficulty, a particular MCMC scheme, called Hamiltonian Monte Carlo (HMC) algorithm, is derived [36], [37]. It differs from the standard Metropolis-Hastings algorithm by exploiting Hamiltonian evolution dynamics to propose states with higher acceptance ratio, reducing the correlation between successive samples. Thus, the main contributions of this paper are two-fold. First, the paper presents a new hierarchical Bayesian fusion model whose parameters

and hyperparameters have to be estimated from the observed images. This model is defined by the likelihood, the priors and the hyper-priors detailed in the following sections. Second, a hybrid Gibbs sampler based on a Hamiltonian MCMC method is introduced to sample the desired posterior distribution. These samples are subsequently used to approximate a Bayesian estimator of the fused image.

The paper is organized as follows. Section II formulates the fusion problem in a Bayesian framework, with a particular attention to the forward model that exploits physical considerations. Section III derives the hierarchical Bayesian model to obtain the joint posterior distribution of the unknown image, its parameters and hyperparameters. In Section IV, the hybrid Gibbs sampler based on Hamiltonian MCMC is introduced to sample the desired posterior distribution. Simulations are conducted in Section V and conclusions are finally reported in Section VI.

II. Problem formulation

A. Notations and observation model

Let $\mathbf{Z}_1, \ldots, \mathbf{Z}_P$ denote a set of P multi-band images acquired by different optical sensors for a same scene $\mathbf{X}$. These measurements can be of different natures, e.g., PAN, MS and HS, with different spatial and/or spectral resolutions. The observed data $\mathbf{Z}_p$, $p = 1, \ldots, P$, are supposed to be degraded versions of the high-spectral and high-spatial resolution scene $\mathbf{X}$, according to the following observation model

$$\mathbf{Z}_p = \mathcal{F}_p(\mathbf{X}) + \mathbf{E}_p. \tag{1}$$

In (1), $\mathcal{F}_p(\cdot)$ is a linear or nonlinear transformation that models the degradation operated on $\mathbf{X}$. As previously assumed in numerous works (see for instance [5], [26], [33], [38], [39] among some recent contributions), these degradations may include spatial blurring, spatial decimation and spectral mixing which can all be modeled by linear transformations. In what follows, the remotely sensed images $\mathbf{Z}_p$ and the unobserved scene $\mathbf{X}$ are assumed to be pixelated images of sizes $n_{\mathrm{x},p} \times n_{\mathrm{y},p} \times n_{\lambda,p}$ and $m_{\mathrm{x}} \times m_{\mathrm{y}} \times m_{\lambda}$, respectively, where $\cdot_{\mathrm{x}}$ and $\cdot_{\mathrm{y}}$ refer to both spatial dimensions of the images, and $\cdot_{\lambda}$ is for the spectral dimension.

Moreover, in the right-hand side of (1), $\mathbf{E}_p$ stands for an additive error term that both reflects the mismodeling and the observation noise.

Classically, the observed image $\mathbf{Z}_p$ can be lexicographically ordered to build the $N_p \times 1$ vector $\mathbf{z}_p$, where $N_p = n_{\mathrm{x},p} n_{\mathrm{y},p} n_{\lambda,p}$ is the total number of measurements in the observed image $\mathbf{Z}_p$. For writing convenience, but without any loss of generality, the band interleaved by pixel (BIP) vectorization scheme (see [40, pp. 103–104] for a more detailed description of these data format conventions) is adopted in what follows (see paragraph III-B1). Considering a linear degradation, the observation equation (1) can be easily rewritten as follows

$$\mathbf{z}_p = \mathbf{F}_p \mathbf{x} + \mathbf{e}_p \tag{2}$$

where $\mathbf{x} \in \mathbb{R}^M$ and $\mathbf{e}_p \in \mathbb{R}^{N_p}$ are ordered versions of the scene $\mathbf{X}$ (with $M = m_{\mathrm{x}} m_{\mathrm{y}} m_{\lambda}$) and the noise term $\mathbf{E}_p$, respectively. In this work, the noise vector $\mathbf{e}_p$ will be assumed to be a band-dependent Gaussian sequence, i.e., $\mathbf{e}_p \sim \mathcal{N}\left(\mathbf{0}_{N_p}, \mathbf{\Lambda}_p\right)$ where $\mathbf{0}_{N_p}$ is an $N_p \times 1$ vector made of zeros and $\mathbf{\Lambda}_p = \mathbf{I}_{n_{\mathrm{x},p} n_{\mathrm{y},p}} \otimes \mathbf{S}_p$ is an $N_p \times N_p$ matrix where $\mathbf{I}_{n_{\mathrm{x},p} n_{\mathrm{y},p}} \in \mathbb{R}^{n_{\mathrm{x},p} n_{\mathrm{y},p} \times n_{\mathrm{x},p} n_{\mathrm{y},p}}$ is the identity matrix, $\otimes$ is the Kronecker product and $\mathbf{S}_p \in \mathbb{R}^{n_{\lambda,p} \times n_{\lambda,p}}$ is a diagonal matrix containing the noise variances, i.e., $\mathbf{S}_p = \mathrm{diag}\left[s_{p,1}^2 \cdots s_{p,n_{\lambda,p}}^2\right]$.

In (2), $\mathbf{F}_p$ is an $N_p \times M$ matrix that reflects the spatial and/or spectral degradation $\mathcal{F}_p(\cdot)$ operated on $\mathbf{x}$. As in [25], $\mathcal{F}_p(\cdot)$ can represent a spatial decimating operation. For instance, when applied to a single-band image (i.e., $n_{\lambda,p} = m_{\lambda} = 1$) with a decimation factor d in both spatial dimensions, it is easy to show that $\mathbf{F}_p$ is an $n_{\mathrm{x},p} n_{\mathrm{y},p} \times m_{\mathrm{x}} m_{\mathrm{y}}$ block diagonal matrix given in (3) with $m_{\mathrm{x}} = d n_{\mathrm{x},p}$ and $m_{\mathrm{y}} = d n_{\mathrm{y},p}$ [41]. Another example of degradation frequently encountered in the signal and image processing literature is spatial blurring [26], where $\mathcal{F}_p(\cdot)$ usually represents a 2-dimensional convolution by a kernel $\boldsymbol{\kappa}_p$. Similarly, when applied to a single-band image, $\mathbf{F}_p$ is an $n_x n_y \times n_x n_y$ (generally sparse) Toeplitz matrix.

$$\mathbf{F}_p = \frac{1}{d^2} \begin{pmatrix} \underbrace{11\ldots1}_{d^2} & & \\ & \ddots & \\ & & \underbrace{11\ldots1}_{d^2} \end{pmatrix} \tag{3}$$

The problem addressed in this paper consists of recovering the high-spectral and high-spatial resolution scene $\mathbf{x}$ by fusing the various spatial and/or spectral information provided by all the observed images $\mathbf{z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_P\}$.

B. Bayesian estimation of $\mathbf{x}$

In this work, we propose to estimate the unknown scene $\mathbf{x}$ within a Bayesian estimation framework. In this statistical estimation scheme, the fused highly-resolved image $\mathbf{x}$ is inferred through its posterior distribution $f(\mathbf{x}|\mathbf{z})$. Given the observed data, this target distribution can be derived from the likelihood function $f(\mathbf{z}|\mathbf{x})$ and the prior distribution $f(\mathbf{x})$ by using the Bayes' formula $f(\mathbf{x}|\mathbf{z}) \propto f(\mathbf{z}|\mathbf{x})\, f(\mathbf{x})$, where $\propto$ means " proportional to". Based on the posterior distribution, several estimators of the scene $\mathbf{x}$ can be investigated. For instance, maximizing $f(\mathbf{x}|\mathbf{z})$ leads to the MAP estimator $\hat{\mathbf{x}}_{\mathrm{MAP}} = \arg\max_{\mathbf{x}} f(\mathbf{x}|\mathbf{z})$. This estimator has been widely exploited for HS image enhancement (see for instance [25], [29], [30] or more recently [5], [26]). This work proposes to focus on the first moment of the posterior distribution $f(\mathbf{x}|\mathbf{z})$, which is known as the posterior mean estimator or the *minimum mean square error* estimator $\hat{\mathbf{x}}_{\mathrm{MMSE}}$. This estimator is defined as

$$\hat{\mathbf{x}}_{\mathrm{MMSE}} = \int \mathbf{x} f(\mathbf{x}|\mathbf{z})\, d\mathbf{x} = \frac{\int \mathbf{x} f(\mathbf{z}|\mathbf{x})\, f(\mathbf{x})\, d\mathbf{x}}{\int f(\mathbf{z}|\mathbf{x})\, f(\mathbf{x})\, d\mathbf{x}}. \tag{4}$$

In this work, we propose a flexible and relevant statistical model to solve the fusion problem. Deriving the corresponding Bayesian estimators $\hat{\mathbf{x}}_{\mathrm{MMSE}}$ defined in (4), requires the definition of the likelihood function $f(\mathbf{z}|\mathbf{x})$ and the prior distribution $f(\mathbf{x})$. These quantities are detailed in the next section. To facilitate reading, notations have been summarized in Table I.

III. Hierarchical Bayesian model

A. Likelihood function

The statistical properties of the noise vectors $\mathbf{e}_p$ $(p = 1, \ldots, P)$ allow one to state that the observed vector $\mathbf{z}_p$ is normally distributed with mean vector $\mathbf{F}_p\mathbf{x}$ and covariance matrix $\mathbf{\Lambda}_p$. Consequently, the likelihood function, that represents a data fitting term relative to the

observed vector $\mathbf{z}_p$, can be easily derived leading to

$$\begin{aligned} f\left(\mathbf{z}_p|\mathbf{x},\mathbf{\Lambda}_p\right) &= (2\pi)^{-\frac{N_p}{2}}\left|\mathbf{\Lambda}_p\right|^{-\frac{n_{\mathrm{x},p}n_{\mathrm{y},p}}{2}} \\ &\quad\times\exp\left(-\frac{1}{2}\left(\mathbf{z}_p-\mathbf{F}_p\mathbf{x}\right)^T\mathbf{\Lambda}_p^{-1}\left(\mathbf{z}_p-\mathbf{F}_p\mathbf{x}\right)\right) \end{aligned} \tag{5}$$

where $|\mathbf{\Lambda}_p|$ is the determinant of the matrix $\mathbf{\Lambda}_p$. As mentioned in the previous section, the collected measurements $\mathtt{z}$ may have been acquired by different (possibly heterogeneous) sensors. Therefore, the observed vectors $\mathbf{z}_1,\ldots,\mathbf{z}_P$ can be generally assumed to be independent, conditionally upon the unobserved scene $\mathbf{x}$ and the noise covariances $\mathbf{\Lambda}_1,\ldots,\mathbf{\Lambda}_p$. As a consequence, the joint likelihood function of the observed data is

$$f\left(\mathbf{z}|\mathbf{x},\mathbf{\Lambda}\right)=\prod_{p=1}^{P}f\left(\mathbf{z}_p|\mathbf{x},\mathbf{\Lambda}_p\right) \tag{6}$$

with $\mathbf{\Lambda}=\left(\mathbf{\Lambda}_1,\ldots,\mathbf{\Lambda}_P\right)^T$.

B. Prior distributions

The unknown parameters are the scene $\mathbf{x}$ to be recovered and the noise covariance matrix $\mathbf{\Lambda}$ relative to each observation. In this section, prior distributions are introduced for these parameters.

1) Scene prior: Following a BIP strategy, the vectorized image $\mathbf{x}$ can be decomposed as $\mathbf{x}=\left[\boldsymbol{x}_1^T,\boldsymbol{x}_2^T,\cdots,\boldsymbol{x}_{m_{\mathrm{x}}m_{\mathrm{y}}}^T\right]^T$, where $\boldsymbol{x}_i=\left[x_{i,1},x_{i,2},\cdots,x_{i,m_\lambda}\right]^T$ is the $m_\lambda\times 1$ vector corresponding to the ith spatial location (with $i=1,\cdots,m_{\mathrm{x}}m_{\mathrm{y}}$). Since adjacent HS bands are known to be highly correlated, the HS vector $\boldsymbol{x}_i$ usually lives in a subspace whose dimension is much smaller than the number of bands m_λ [42], i.e.,

$$\boldsymbol{x}_i=\mathbf{V}^T\boldsymbol{u}_i \tag{7}$$

where $\boldsymbol{u}_i$ is the projection of the vector $\boldsymbol{x}_i$ onto the subspace spanned by the columns of $\mathbf{V}^T\in\mathbb{R}^{m_\lambda\times\widetilde{m}_\lambda}$. Note that $\mathbf{V}^T$ is possibly known *a priori* from the scene we are interested in or can be learned from the HS data. In the proposed framework, we exploit the dimensionality reduction (DR) as prior information instead of reducing the dimensionality of HS data directly. Another motivation for DR is that the dimension of the subspace $\widetilde{m}_\lambda$ is generally much smaller

than the number of bands, i.e., $\widetilde{m}_\lambda \ll m_\lambda$. As a consequence, inferring in the subspace $\mathbb{R}^{\widetilde{m}_\lambda \times 1}$ greatly decreases the computational burden of the fusion algorithm. Note that the DR technique defined by (7) has been used in some related HS analysis references, e.g., [32], [42]. More experimental justifications for the necessity of DR can be found in Section V-F. Using the notation $\mathbf{u} = \left[\boldsymbol{u}_1^T, \boldsymbol{u}_2^T, \cdots, \boldsymbol{u}_{m_\mathrm{x} m_\mathrm{y}}^T\right]^T$, we have $\mathbf{u} = \mathfrak{V}\mathbf{x}$, where $\mathfrak{V}$ is an $\widetilde{M} \times M$ block-diagonal matrix whose blocks are equal to $\mathbf{V}$ and $\widetilde{M} = m_\mathrm{x} m_\mathrm{y} \widetilde{m}_\lambda$. Instead of assigning a prior distribution to the vectors $\boldsymbol{x}_i$, we propose to define a prior for the projected vectors $\boldsymbol{u}_i$ $(i = 1, \cdots, m_\mathrm{x} m_\mathrm{y})$

$$\boldsymbol{u}_i | \boldsymbol{\mu}_{\boldsymbol{u}_i}, \boldsymbol{\Sigma}_{\boldsymbol{u}_i} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{u}_i}, \boldsymbol{\Sigma}_{\boldsymbol{u}_i}\right). \tag{8}$$

As $\boldsymbol{u}_i$ is a linear transformation of $\boldsymbol{x}_i$, the Gaussian prior assigned to $\boldsymbol{u}_i$ leads to a Gaussian prior for $\boldsymbol{x}_i$, which allows the ill-posed problem (2) to be regularized. The covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{u}_i}$ is designed to explore the correlations between the different spectral bands after projection in the subspace of interest. Also, the mean $\bar{\boldsymbol{\mu}}_\mathbf{u}$ of the whole image $\mathbf{u}$ as well as its covariance matrix $\bar{\boldsymbol{\Sigma}}_\mathbf{u}$ can be constructed from $\boldsymbol{\mu}_{\boldsymbol{u}_i}$ and $\boldsymbol{\Sigma}_{\boldsymbol{u}_i}$ as follows

$$\begin{aligned} \bar{\boldsymbol{\mu}}_\mathbf{u} &= \left[\boldsymbol{\mu}_{\boldsymbol{u}_1}^T, \cdots, \boldsymbol{\mu}_{\boldsymbol{u}_{m_\mathrm{x} m_\mathrm{y}}}^T\right]^T \\ \bar{\boldsymbol{\Sigma}}_\mathbf{u} &= \mathrm{diag}\left[\boldsymbol{\Sigma}_{\boldsymbol{u}_1}, \cdots, \boldsymbol{\Sigma}_{\boldsymbol{u}_{m_\mathrm{x} m_\mathrm{y}}}\right]. \end{aligned} \tag{9}$$

The Gaussian prior assigned to $\mathbf{u}$ implies that the target image $\mathbf{u}$ is *a priori* not too far from the mean vector $\bar{\boldsymbol{\mu}}_\mathbf{u}$, whereas the covariance matrix $\bar{\boldsymbol{\Sigma}}_\mathbf{u}$ tells us how much confidence we have for the prior (the choice of the hyperparameters $\bar{\boldsymbol{\mu}}_\mathbf{u}$ and $\bar{\boldsymbol{\Sigma}}_\mathbf{u}$ will be discussed later in Section III-C). Choosing a Gaussian prior for the vectors $\boldsymbol{u}_i$ is also motivated by the fact this kind of prior has been used successfully in several works related to the fusion of multiple degraded images, including [29], [43], [44]. Note finally that the Gaussian prior has the interest of being a conjugate distribution relative to the statistical model (6). As it will be shown in Section IV, coupling this Gaussian prior distribution with the Gaussian likelihood function leads to simpler estimators constructed from the posterior distribution $f(\mathbf{u}|\mathbf{z})$. Finally, it is interesting to mention that the proposed method is quite robust to the non-Gaussianity of the image. Some additional results obtained for synthetic non-Gaussian images as well as related discussions are available in Section V-I.

2) Noise variance priors: Inverse-gamma distributions are chosen as prior distributions for the noise variances $s_{p,i}^2$ $(i = 1, \ldots, n_{\lambda,p}, p = 1, \ldots, P)$

$$s_{p,i}^2|\nu, \gamma \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\gamma}{2}\right). \tag{10}$$

The inverse-gamma distribution is a very flexible distribution whose shape can be adjusted by its two parameters. For simplicity, we propose to fix the hyperparameter ν whereas the hyperparameter γ will be estimated from the data. This strategy is very classical for scale parameters (e.g., see [45]). Note that the inverse-gamma distribution (10) is conjugate for the statistical model (6), which will allow closed-form expressions to be obtained for the conditional distributions $f\left(s_{p,i}^2|\mathbf{z}\right)$ of the noise variances. By assuming the variances $\boldsymbol{s}^2 = \left\{s_{p,i}^2\right\}(\forall p, i)$ are *a priori* independent, the joint prior distribution of the noise variance vector $\boldsymbol{s}^2$ is

$$f\left(\boldsymbol{s}^2|\nu, \gamma\right) = \prod_{p=1}^{P} \prod_{i=1}^{n_{\lambda,p}} f\left(s_{p,i}^2|\nu, \gamma\right). \tag{11}$$

C. Hyperparameter priors

The hyperparameter vector associated with the parameter priors defined above includes $\bar{\boldsymbol{\mu}}_{\mathbf{u}}$, $\bar{\boldsymbol{\Sigma}}_{\mathbf{u}}$ and γ. The quality of the fusion algorithm investigated in this paper depends on the values of the hyperparameters that need to be adjusted carefully. Instead of fixing all these hyperparameters *a priori*, we propose to estimate some of them from the data using a hierarchical Bayesian algorithm [46, Chap. 8]. Specifically, we propose to fix $\bar{\boldsymbol{\mu}}_{\mathbf{u}}$ as the interpolated HS image in the subspace of interest following the strategy in [25]. Similarly, to reduce the number of statistical parameters to be estimated, all the covariance matrix are assumed to equal, i.e., $\boldsymbol{\Sigma}_{\boldsymbol{u}_i} = \boldsymbol{\Sigma}_{\boldsymbol{u}}(\forall i)$. Thus, the hyperparameter vector to be estimated jointly with the parameters of interest is $\boldsymbol{\Phi} = \{\boldsymbol{\Sigma}_{\boldsymbol{u}}, \gamma\}$. The prior distributions for these two hyperparameters are defined below.

1) Hyperparameter $\boldsymbol{\Sigma}_{\boldsymbol{u}}$: Assigning a conjugate *a priori* inverse-Wishart distribution to the covariance matrix of a Gaussian vector has provided interesting results in the signal and image processing literature [47], [48]. Following these works, we have chosen the following prior for $\boldsymbol{\Sigma}_{\boldsymbol{u}}$

$$\boldsymbol{\Sigma}_{\boldsymbol{u}} \sim \mathcal{W}^{-1}(\boldsymbol{\Psi}, \eta) \tag{12}$$

whose density is

$$f(\boldsymbol{\Sigma}_{\boldsymbol{u}}|\boldsymbol{\Psi},\eta) = \frac{|\boldsymbol{\Psi}|^{\frac{\eta}{2}}}{2^{\frac{\eta\widetilde{m}_{\lambda}}{2}}\Gamma_{\widetilde{m}_{\lambda}}(\frac{\eta}{2})}\,|\boldsymbol{\Sigma}_{\boldsymbol{u}}|^{-\frac{\eta+\widetilde{m}_{\lambda}+1}{2}}\,e^{-\frac{1}{2}\operatorname{tr}(\boldsymbol{\Psi}\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1})}.$$

Again, the hyper-hyperparameters $\boldsymbol{\Psi}$ and η will be fixed to provide a non-informative prior.

2) Hyperparameter γ: To reflect the absence of prior knowledge regarding the mean noise level, a non-informative Jeffreys' prior is assigned to the hyperparameter γ

$$f(\gamma) \propto \frac{1}{\gamma}\mathbf{1}_{\mathbb{R}^+}(\gamma) \tag{13}$$

where $\mathbf{1}_{\mathbb{R}^+}(\cdot)$ is the indicator function defined on $\mathbb{R}^+$

$$\mathbf{1}_{\mathbb{R}^+}(u) = \begin{cases} 1, & \text{if } u \in \mathbb{R}^+, \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

The use of the improper distribution (13) is classical and can be justified by different means (e.g., see [49]), providing that the corresponding full posterior distribution is statistically well defined, which is the case for the proposed fusion model.

D. Inferring the highly-resolved HS image from the posterior distribution of its projection $\mathbf{u}$

Following the parametrization in the prior model (7), the unknown parameter vector $\boldsymbol{\theta} = \{\mathbf{u}, \boldsymbol{s}^2\}$ is composed of the projected scene $\mathbf{u}$ and the noise variance vector $\boldsymbol{s}^2$. The joint posterior distribution of the unknown parameters and hyperparameters can be computed following the hierarchical model $f(\boldsymbol{\theta},\boldsymbol{\Phi}|\mathtt{z}) \propto f(\mathtt{z}|\boldsymbol{\theta})\,f(\boldsymbol{\theta}|\boldsymbol{\Phi})\,f(\boldsymbol{\Phi})$. By assuming prior independence between the hyperparameters $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ and γ and the parameters $\mathbf{u}$ and $\boldsymbol{s}^2$ conditionally upon $(\boldsymbol{\Sigma}_{\boldsymbol{u}},\gamma)$, the following results can be obtained $f(\boldsymbol{\theta}|\boldsymbol{\Phi}) = f(\mathbf{u}|\boldsymbol{\Sigma}_{\boldsymbol{u}})\,f(\boldsymbol{s}^2|\gamma)$ and $f(\boldsymbol{\Phi}) = f(\boldsymbol{\Sigma}_{\boldsymbol{u}})\,f(\gamma)$. Note that $f(\mathtt{z}|\boldsymbol{\theta})$, $f(\mathbf{u}|\boldsymbol{\Sigma}_{\boldsymbol{u}})$ and $f(\boldsymbol{s}^2|\gamma)$ have been defined in (6), (8) and (11), respectively.

The posterior distribution of the projected target image $\mathbf{u}$, required to compute the Bayesian estimators (4), is obtained by marginalizing out the hyperparameter vector $\boldsymbol{\Phi}$ and the noise variances $\boldsymbol{s}^2$ from the joint posterior distribution $f(\boldsymbol{\theta},\boldsymbol{\Phi}|\mathtt{z})$

$$f(\mathbf{u}|\mathtt{z}) \propto \int f(\boldsymbol{\theta},\boldsymbol{\Phi}|\mathtt{z})\,d\boldsymbol{\Phi}ds^2_{1,1},\ldots,ds^2_{P,n_{\lambda,P}}. \tag{15}$$

The posterior distribution (15) is too complex to obtain closed-form expressions of the MMSE and MAP estimators $\hat{\mathbf{u}}_{\mathrm{MMSE}}$ and $\hat{\mathbf{u}}_{\mathrm{MAP}}$. As an alternative, this paper proposes to use an MCMC algorithm to generate a collection of N_{MC} samples $\mathcal{U} = \left\{\tilde{\mathbf{u}}^{(1)}, \ldots, \tilde{\mathbf{u}}^{(N_{\mathrm{MC}})}\right\}$ that are asymptotically distributed according to the posterior of interest $f\left(\mathbf{u}|\mathbf{z}\right)$. These samples will be used to compute the Bayesian estimators of $\mathbf{u}$. More precisely, the MMSE estimator of $\mathbf{u}$ will be approximated by an empirical average of the generated samples

$$\hat{\mathbf{u}}_{\mathrm{MMSE}} \approx \frac{1}{N_{\mathrm{MC}} - N_{\mathrm{bi}}} \sum_{t=N_{\mathrm{bi}}+1}^{N_{\mathrm{MC}}} \tilde{\mathbf{u}}^{(t)} \tag{16}$$

where N_{bi} is the number of burn-in iterations. Once the MMSE estimate $\hat{\mathbf{u}}_{\mathrm{MMSE}}$ has been computed, the highly-resolved HS image can be computed as $\hat{\mathbf{x}}_{\mathrm{MMSE}} = \mathfrak{V}^T \hat{\mathbf{u}}_{\mathrm{MMSE}}$. Sampling directly according to the marginal posterior distribution $f\left(\mathbf{u}|\mathbf{z}\right)$ is not straightforward. Instead, we propose to sample according to the joint posterior $f\left(\mathbf{u}, \boldsymbol{s}^2, \boldsymbol{\Sigma}_{\boldsymbol{u}}|\mathbf{z}\right)$ (hyperparameter γ has been marginalized) by using a Metropolis-within-Gibbs sampler, which can be easily implemented since all the conditional distributions associated with $f\left(\mathbf{u}, \boldsymbol{s}^2, \boldsymbol{\Sigma}_{\boldsymbol{u}}|\mathbf{z}\right)$ are relatively simple. The resulting hybrid Gibbs sampler is detailed in the following section.

IV. Hybrid Gibbs Sampler

The Gibbs sampler has received a considerable attention in the statistical community (see [34], [46]) to solve Bayesian estimation problems. The interesting property of this Monte Carlo algorithm is that it only requires to determine the conditional distributions associated with the distribution of interest. These conditional distributions are generally easier to simulate than the joint target distribution. The block Gibbs sampler that we propose to sample according to $f\left(\mathbf{u}, \boldsymbol{s}^2, \boldsymbol{\Sigma}_{\boldsymbol{u}}|\mathbf{z}\right)$ is defined by a 3-step procedure reported in Algorithm 1. The distribution involved in this algorithm are detailed below.

A. Sampling $\boldsymbol{\Sigma_u}$ according to $f\left(\boldsymbol{\Sigma_u}|\mathbf{u}, \boldsymbol{s}^2, \mathbf{z}\right)$

Standard computations yield the following inverse-Wishart distribution as conditional distribution for the covariance matrix $\boldsymbol{\Sigma_u}$ (of the scene to be recovered)

$$\begin{aligned} &\boldsymbol{\Sigma_u}|\mathbf{u}, \boldsymbol{s}^2, \mathbf{z} \sim \\ &\mathcal{W}^{-1}\left(\boldsymbol{\Psi} + \sum_{i=1}^{m_{\mathrm{x}} m_{\mathrm{y}}} (\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i})^T (\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}_i}), m_{\mathrm{x}} m_{\mathrm{y}} + \eta\right). \end{aligned} \tag{17}$$

B. Sampling $\mathbf{u}$ according to $f\left(\mathbf{u}|\boldsymbol{\Sigma_u}, \boldsymbol{s}^2, \mathbf{z}\right)$

Choosing the conjugate distribution (8) as prior distribution for the projected unknown image $\mathbf{u}$ leads to the following conditional posterior distribution for $\mathbf{u}$

$$\mathbf{u}|\boldsymbol{\Sigma_u}, \boldsymbol{s}^2, \mathbf{z} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{u}|\mathbf{z}}, \boldsymbol{\Sigma}_{\mathbf{u}|\mathbf{z}}\right) \tag{18}$$

with

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{u}|\mathbf{z}} &= \left[\bar{\boldsymbol{\Sigma}}_{\mathbf{u}}^{-1} + \textstyle\sum_{p=1}^{P} \mathfrak{V}\mathbf{F}_p^T \boldsymbol{\Lambda}_p^{-1} \mathbf{F}_p \mathfrak{V}^T\right]^{-1} \\ \boldsymbol{\mu}_{\mathbf{u}|\mathbf{z}} &= \boldsymbol{\Sigma}_{\mathbf{u}|\mathbf{z}} \left[\textstyle\sum_{p=1}^{P} \mathfrak{V}\mathbf{F}_p^T \boldsymbol{\Lambda}_p^{-1} \mathbf{z}_p + \bar{\boldsymbol{\Sigma}}_{\mathbf{u}}^{-1} \bar{\boldsymbol{\mu}}_{\mathbf{u}}\right]. \end{aligned} \tag{19}$$

Sampling directly according to this multivariate Gaussian distribution requires the inversion of an $\widetilde{M} \times \widetilde{M}$ matrix, which is impossible in most fusion problems. An alternative would consist of sampling each element u_i ($i = 1, \ldots, \widetilde{M}$) of $\mathbf{u}$ conditionally upon the others according to $f\left(u_i|\mathbf{u}_{-i}, \boldsymbol{s}^2, \boldsymbol{\Sigma_u}, \mathbf{z}\right)$, where $\mathbf{u}_{-i}$ is the vector $\mathbf{u}$ whose ith component has been removed. However, this alternative would require to sample $\mathbf{u}$ by using $\widetilde{M}$ Gibbs moves, which is time demanding and leads to poor mixing properties.

The efficient strategy adopted in this work relies on a particular MCMC method, called Hamiltonian Monte Carlo (HMC) method (sometimes referred to as hybrid Monte Carlo method), which is considered to generate vectors $\mathbf{u}$ directly. More precisely, we consider the HMC algorithm initially proposed by Duane *et al.* for simulating the lattice field theory in [36]. As detailed in [50], this technique allows mixing property of the sampler to be improved, especially in a high-dimensional problem. It exploits the gradient of the distribution to be sampled by introducing auxiliary "momentum" variables $\mathbf{m} \in \mathbb{R}^{\widetilde{M}}$. The joint distribution of

the unknown parameter vector $\mathbf{u}$ and the momentum is defined as

$$f\left(\mathbf{u},\mathbf{m}|\boldsymbol{s}^2,\boldsymbol{\Sigma}_{\boldsymbol{u}},\mathtt{z}\right)=f\left(\mathbf{u}|\boldsymbol{s}^2,\boldsymbol{\Sigma}_{\boldsymbol{u}},\mathtt{z}\right)f\left(\mathbf{m}\right)$$

where $f\left(\mathbf{m}\right)$ is the normal probability density function (pdf) with zero mean and identity covariance matrix. The Hamiltonian of the considered system is defined by taking the negative logarithm of the posterior distribution $f\left(\mathbf{u},\mathbf{m}|\boldsymbol{s}^2,\boldsymbol{\mu}_{\boldsymbol{u}},\boldsymbol{\Sigma}_{\boldsymbol{u}},\mathtt{z}\right)$ to be sampled, i.e.,

$$\begin{aligned} H\left(\mathbf{u},\mathbf{m}\right) &= -\log f\left(\mathbf{u},\mathbf{m}|\boldsymbol{s}^2,\boldsymbol{\mu}_{\boldsymbol{u}},\boldsymbol{\Sigma}_{\boldsymbol{u}},\mathtt{z}\right) \\ &= U\left(\mathbf{u}\right)+K\left(\mathbf{m}\right) \end{aligned} \tag{20}$$

where $U\left(\mathbf{u}\right)$ is the potential energy function defined by the negative logarithm of $f\left(\mathbf{u}|\boldsymbol{s}^2,\boldsymbol{\Sigma}_{\boldsymbol{u}},\mathtt{z}\right)$ and $K\left(\mathbf{m}\right)$ is the corresponding kinetic energy

$$\begin{aligned} U\left(\mathbf{u}\right) &= -\log f\left(\mathbf{u}|\boldsymbol{s}^2,\boldsymbol{\Sigma}_{\boldsymbol{u}},\mathtt{z}\right) \\ K\left(\mathbf{m}\right) &= \tfrac{1}{2}\mathbf{m}^T\mathbf{m}. \end{aligned} \tag{21}$$

The parameter space where $(\mathbf{u},\mathbf{m})$ lives is explored following the scheme detailed in Algo 2. At iteration t of the Gibbs sampler, a so-called *leap-frogging* procedure composed of N_{leapfrog} iterations is achieved to propose a move from the current state $\left\{\tilde{\mathbf{u}}^{(t)},\tilde{\mathbf{m}}^{(t)}\right\}$ to the state $\left\{\tilde{\mathbf{u}}^{(\star)},\tilde{\mathbf{m}}^{(\star)}\right\}$ with step size ε. This move is operated in $\mathbb{R}^{\widetilde{M}}\times\mathbb{R}^{\widetilde{M}}$ in a direction given by the gradient of the energy function

$$\nabla_{\mathbf{u}}U\left(\mathbf{u}\right)=-\sum_{p=1}^{P}\mathfrak{V}\mathbf{F}_p^T\boldsymbol{\Lambda}_p^{-1}\left(\mathbf{z}_p-\mathbf{F}_p\mathfrak{V}^T\mathbf{u}\right)+\boldsymbol{\Sigma}_{\mathbf{u}}^{-1}(\mathbf{u}-\bar{\boldsymbol{\mu}}_{\mathbf{u}}). \tag{22}$$

Then, the new state is accepted with probability $\rho_t=\min\left\{1,A_t\right\}$ where

$$\begin{aligned} A_t &= \frac{f\left(\tilde{\mathbf{u}}^{(\star)},\tilde{\mathbf{m}}^{(\star)}|\boldsymbol{s}^2,\boldsymbol{\Sigma}_{\boldsymbol{u}},\mathtt{z}\right)}{f\left(\tilde{\mathbf{u}}^{(t)},\tilde{\mathbf{m}}^{(t)}|\boldsymbol{s}^2,\boldsymbol{\Sigma}_{\boldsymbol{u}},\mathtt{z}\right)} \\ &= \exp\left[H\left(\tilde{\mathbf{u}}^{(t)},\tilde{\mathbf{m}}^{(t)}\right)-H\left(\tilde{\mathbf{u}}^{(\star)},\tilde{\mathbf{m}}^{(\star)}\right)\right]. \end{aligned} \tag{23}$$

This accept/reject procedure ensures that the simulated vectors $(\tilde{\mathbf{u}}^{(t)},\tilde{\mathbf{m}}^{(t)})$ are asymptotically distributed according to the distribution of interest. The way the parameters ε and N_{L} have been adjusted will be detailed in Section V.

To sample according to a high-dimension Gaussian distribution such as $f\left(\mathbf{u}|\boldsymbol{\Sigma}_{\boldsymbol{u}},\boldsymbol{s}^2,\mathtt{z}\right)$, one

might think of using other simulation techniques such as the method proposed in [51] to solve super resolution problems. Similarly, Orieux *et al.* have proposed a perturbation approach to sample high-dimensional Gaussian distributions for general linear inverse problems [52]. However, these techniques rely on additional optimization schemes included within the Monte Carlo algorithm, which implies that the generated samples are only approximately distributed according to the target distribution. Conversely, the HMC strategy proposed here ensures asymptotic convergence of the generated samples to the posterior distribution. Moreover, the HMC method is very flexible and can be easily extended to handle non-Gaussian posterior distributions contrary to the methods investigated in [51], [52].

C. Sampling $\boldsymbol{s}^2$ according to $f\left(\boldsymbol{s}^2|\mathbf{u}, \boldsymbol{\Sigma}_{\boldsymbol{u}}, \mathbf{z}\right)$

The conditional pdf of the noise variance $s_{p,i}^2$ $(i = 1, \ldots, n_{\lambda,p}, p = 1, \ldots, P)$ is

$$\begin{aligned} &f\left(s_{p,i}^2|\mathbf{u}, \boldsymbol{\Sigma}_{\boldsymbol{u}}, \mathbf{z}\right) \propto \\ &\left(\frac{1}{s_{p,i}^2}\right)^{\frac{n_{\mathrm{x},p} n_{\mathrm{y},p}}{2}+1} \exp\left(-\frac{\left\|(\mathbf{z}_p - \mathbf{F}_p \mathfrak{V}^T \mathbf{u})_i\right\|^2}{2 s_{p,i}^2}\right) \end{aligned} \tag{24}$$

where $(\mathbf{z}_p - \mathbf{F}_p \mathfrak{V}^T \mathbf{u})_i$ contains the elements of the ith band. Generating samples $s_{p,i}^2$ distributed according to $f\left(s_{p,i}^2|\mathbf{u}, \boldsymbol{\Sigma}_{\boldsymbol{u}}, \mathbf{z}\right)$ is classically achieved by drawing samples from the following inverse-gamma distribution

$$s_{p,i}^2|\mathbf{u}, \mathbf{z} \sim \mathcal{IG}\left(\frac{n_{\mathrm{x},p} n_{\mathrm{y},p}}{2}, \frac{\left\|(\mathbf{z}_p - \mathbf{F}_p \mathfrak{V}^T \mathbf{u})_i\right\|^2}{2}\right). \tag{25}$$

In practice, if the noise variances are known a prior, we simply assign the noise variances to be known values and remove the sampling of the noise variances.

D. Complexity Analysis

The MCMC method can be computationally costly compared with optimization methods [53]. The complexity of the proposed Gibbs sampler is mainly due to the Hamiltonian Monte Carlo method. The complexity of the Hamiltonian MCMC method is $\mathcal{O}((\widetilde{m_\lambda})^3) + \mathcal{O}((\widetilde{m_\lambda} m_{\mathrm{x}} m_{\mathrm{y}})^2)$, which is highly expensive as m_λ increases. Generally the number of pixels $m_{\mathrm{x}} m_{\mathrm{y}}$ cannot be reduced significantly. Thus, projecting the high-dimensional $m_\lambda \times 1$ vectors

to a low-dimension space to form $\widetilde{m_\lambda} \times 1$ vectors decreases the complexity while keeping most important information.

V. Simulation Results

This section studies the performance of the proposed Bayesian fusion algorithm. 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) [54]. This image was initially composed of 224 bands that have been reduced to 177 bands ($m_\lambda = n_{\lambda,1} = 177$) after removing the water vapor absorption bands.

A. Fusion of HS and MS images

We propose to reconstruct the reference HS image from two lower resolved images. First, a high-spectral low-spatial resolution image $\mathbf{z}_1$, denoted as HS image, has been generated by applying a 5×5 averaging filter on each band of the reference image. Besides, an MS image $\mathbf{z}_2$ is obtained by successively averaging the adjacent bands according to realistic spectral responses. More precisely, the reference image is filtered using the LANDSAT-like spectral responses depicted in the top of Fig. 1, to obtain a 7-band ($n_{\lambda,2} = 7$) MS image. Note here that the observation models $\mathbf{F}_1$ and $\mathbf{F}_2$ corresponding to the HS and MS images are perfectly known. In addition to the blurring and spectral mixing, the HS and MS images have been both contaminated by zero-mean additive Gaussian noise. The noise power $s_{p,i}^2$ depends on the signal to noise ratio $\text{SNR}_{p,i}$ $(i = 1, \cdots, n_{\lambda,p}, p = 1, 2)$ defined by $\text{SNR}_{p,i} = 10 \log_{10} \left(\frac{\|(\mathbf{F}_p \mathbf{x})_i\|_F^2}{n_{\text{x},p} n_{\text{y},p} s_{p,i}^2} \right)$, where $\|.\|_F$ is the Frobenius norm. Our simulations have been conducted with $\text{SNR}_{1,\cdot} = 35$dB for the first 127 bands and $\text{SNR}_{1,\cdot} = 30$dB for the remaining 50 bands of the HS image. For the MS image, $\text{SNR}_{2,\cdot}$ is 30dB for all bands. A composite color image, formed by selecting the red, green and blue bands of the high-spatial resolution HS image (the reference image) is shown in the right bottom of Fig. 2. The noise-contaminated HS and MS images are depicted in the top left and top right of Fig. 2.

1) Subspace learning: Learning the matrix $\mathbf{V}$ in (7) is a preprocessing step, which can be solved by different strategies. A lot of DR methods might be exploited, such as locally linear

embedding (LLE) [55], independent component analysis (ICA) [56], hyperspectral signal subspace identification by minimum error (HySime) [42], minimum change rate deviation (MCRD) [57] and so on. In this work, we propose to use the principal component analysis (PCA), which is a classical DR technique used in HS imagery. It maps the original data into a lower dimensional subspace while preserving most information about the original data. Note that the bases of this subspace are the columns of the transformation matrix $\mathbf{V}^T$, which are exactly the same for all pixels (or spectral vectors). As in paragraph III-B1, the vectorized HS image $\mathbf{z}_1$ can be written as $\mathbf{z}_1 = \left[\boldsymbol{z}_{1,1}^T, \boldsymbol{z}_{1,2}^T, \cdots, \boldsymbol{z}_{1,n_{\mathrm{x},1}n_{\mathrm{y},1}}^T\right]^T$, where $\boldsymbol{z}_{1,i} = \left[z_{1,i,1}, z_{1,i,2}, \cdots, z_{1,i,n_{\lambda,1}}\right]^T$. Then, the sample covariance matrix of the HS image $\mathbf{z}_1$ is diagonalized leading to

$$\mathbf{W}^T \boldsymbol{\Upsilon} \mathbf{W} = \mathbf{D} \tag{26}$$

where $\mathbf{W}$ is an $m_\lambda \times m_\lambda$ orthogonal matrix ($\mathbf{W}^T = \mathbf{W}^{-1}$) and $\mathbf{D}$ is a diagonal matrix whose diagonal elements are the ordered eigenvalues of $\boldsymbol{\Upsilon}$ denoted as $d_1 \geq d_2 \geq ... \geq d_{m_\lambda}$. The dimension of the projection subspace $\widetilde{m}_\lambda$ is defined as the minimum integer satisfying the condition $\sum_{i=1}^{\widetilde{m}_\lambda} d_i / \sum_{i=1}^{m_\lambda} d_i \geq 0.99$. The matrix $\mathbf{V}$ is then constructed as the eigenvectors associated with the $\widetilde{m}_\lambda$ largest eigenvalues of $\boldsymbol{\Upsilon}$. As an illustration, the eigenvalues of the sample covariance matrix $\boldsymbol{\Upsilon}$ for the Moffett field image are displayed in Fig. 3. For this example, the $\widetilde{m}_\lambda = 10$ eigenvectors contain 99.93% of the information. An experimental justification is reported in Section V-F, which contains additional results obtained for fusion with and without PCA.

2) Hyper-hyperparameter selection: In our experiments, fixed hyper-hyperparameters have been chosen as follows: $\boldsymbol{\Psi} = \mathbf{I}_{\widetilde{m}_\lambda}, \eta = \widetilde{m}_\lambda + 3$.

These choices can be motivated by the following arguments:

- The identity matrix assigned to $\boldsymbol{\Psi}$ ensures a non-informative prior.
- Setting the inverse gamma parameters to $\eta = \widetilde{m}_\lambda + 3$ also leads to a non-informative prior [45].
- Note that parameter ν disappears when the joint posterior is integrated out with respect to parameter γ.

B. Stepsize and Leapfrog Steps

The performance of the HMC method is mainly governed by the stepsize ε and the number of leapfrog steps N_{L}. As pointed out in [37], a too large stepsize will result in a very low acceptance rate and a too small stepsize yields high computational complexity. In order to adjust the stepsize parameter ε, we propose to monitor the statistical acceptance ratio $\hat{\rho}_t$ defined as $\hat{\rho}_t = \frac{N_{\mathrm{a,t}}}{N_{\mathrm{W}}}$ where N_{W} is the length of the counting window (in our experiment, the counting window at time t contains the vectors $\tilde{\mathbf{x}}^{(t-N_{\mathrm{W}}+1)}, \tilde{\mathbf{x}}^{(t-N_{\mathrm{W}})}, \cdots, \tilde{\mathbf{x}}^{(t)}$ with $N_{\mathrm{W}} = 50$) and $N_{\mathrm{a,t}}$ is the number of accepted samples in this window at time t. As explained in [58], the adaptive tuning should adapt less and less as the algorithm proceeds to guarantee that the generated samples form a stationary Markov chain. In the proposed implementation, the parameter ε is adjusted as in Algorithm 3. The thresholds have been fixed to $(\alpha_d, \alpha_u) = (0.3, 0.9)$ and the scale parameters are $(\beta_d, \beta_u) = (1.1, 0.9)$ (these parameters were adjusted by cross-validation). Note that the initial value of ε should not be too large to 'blow up' the leapfrog trajectory [37]. Generally, the stepsize converges after some iterations of Algorithm 3.

Regarding the number of leapfrogs, setting the trajectory length N_{L} by trial and error is necessary [37]. To avoid the potential resonance, N_{L} is randomly chosen from a uniform distribution from N_{min} to N_{max}. After some preliminary runs and tests, $N_{\mathrm{min}} = 50$ and $N_{\mathrm{max}} = 55$ have been selected.

C. Evaluation of the Fusion Quality

To evaluate the quality of the proposed fusion strategy, different image quality measures can be investigated. Referring to [26], we propose to use RSNR, SAM, UIQI, ERGAS and DD as defined below. These measures have been widely used in the HS image processing community and are appropriate for evaluating the quality of the fusion in terms of spectral and spatial resolutions [25], [27], [28].

RSNR: The reconstruction SNR (RSNR) is related to the difference between the actual and fused images $\mathrm{RSNR}(\mathbf{x}, \hat{\mathbf{x}}) = 10 \log_{10} \left(\frac{\|\mathbf{x}\|^2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2} \right)$. The larger RSNR, the better the fusion quality and vice versa.

SAM: The spectral angle mapper (SAM) measures the spectral distortion between the actual and estimated images. The SAM of two spectral vectors $\boldsymbol{x}_n$ and $\hat{\boldsymbol{x}}_n$ is defined as $\text{SAM}(\boldsymbol{x}_n, \hat{\boldsymbol{x}}_n) = \arccos\left(\frac{\langle \boldsymbol{x}_n, \hat{\boldsymbol{x}}_n \rangle}{\|\boldsymbol{x}_n\|_2 \|\hat{\boldsymbol{x}}_n\|_2}\right)$. The average SAM is finally obtained by averaging the SAMs of all image pixels. Note that SAM value is expressed in radians and thus belongs to $[-\frac{\pi}{2}, \frac{\pi}{2}]$. The smaller the absolute value of SAM, the less important the spectral distortion.

UIQI: The universal image quality index (UIQI) was proposed in [59] for evaluating the similarity between two single band images. It is related to the correlation, luminance distortion and contrast distortion of the estimated image to the reference image. The UIQI between $\mathbf{a} = [a_1, a_2, \cdots, a_N]$ and $\hat{\mathbf{a}} = [\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_N]$ is defined as $\text{UIQI}(\mathbf{a}, \hat{\mathbf{a}}) = \frac{4\sigma^2_{a\hat{a}} \mu_a \mu_{\hat{a}}}{(\sigma^2_a + \sigma^2_{\hat{a}})(\mu^2_a + \mu^2_{\hat{a}})}$, where $(\mu_a, \mu_{\hat{a}}, \sigma^2_a, \sigma^2_{\hat{a}})$ are the sample means and variances of a and $\hat{a}$, and $\sigma^2_{a\hat{a}}$ is the sample covariance of $(a, \hat{a})$. The range of UIQI is $[-1, 1]$ and UIQI$= 1$ when $\mathbf{a} = \hat{\mathbf{a}}$. For multi-band image, the UIQI is obtained band-by-band and averaged over all bands.

ERGAS: The relative dimensionless global error in synthesis (ERGAS) calculates the amount of spectral distortion in the image [60]. This measure of fusion quality is defined as $\text{ERGAS} = 100 \times \frac{1}{d^2} \sqrt{\frac{1}{m_\lambda} \sum_{i=1}^{m_\lambda} \left(\frac{\text{RMSE}(i)}{\mu_i}\right)}$, where $1/d^2$ is the ratio between the pixel sizes of the MS and HS images, μ_i is the mean of the ith band of the HS image, and m_λ is the number of HS bands. The smaller ERGAS, the smaller the spectral distortion.

DD: The degree of distortion (DD) between two images $\mathbf{X}$ and $\hat{\mathbf{X}}$ is defined as $\text{DD}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{M} \|\text{vec}(\mathbf{X}) - \text{vec}(\hat{\mathbf{X}})\|_1$. The smaller DD, the better the fusion.

D. Comparison with other Bayesian models

The Bayesian model proposed here differs from previous Bayesian models [25], [26] in three-fold. First, in addition to the target image $\mathbf{x}$, the hierarchical Bayesian model allows the distributions of the noise variances $\boldsymbol{s}^2$ and the hyperparameter $\boldsymbol{\Sigma_u}$ to be inferred. The hierarchical inference structure makes this Bayesian model more general and flexible [61]. Second, the covariance matrix $\boldsymbol{\Sigma_u}$ is assumed to be block diagonal, which allows us to exploit the correlations between spectral bands. Third, the proposed method takes advantage of the relation between the MS image and the target image by introducing a forward model $\mathbf{F}_2$. This paragraph compares the proposed Bayesian fusion method with these two state-of-the-art fusion algorithms [25] [26] for HS+MS fusion. The MMSE estimator of the image

using the proposed Bayesian method is obtained from (16). In this simulation, $N_{\text{MC}} = 500$ and $N_{\text{bi}} = 500$. The fusion results obtained with different algorithms are depicted in Fig. 2. Graphically, the proposed algorithm performs competitively with the state-of-the-art methods. This result is confirmed quantitatively in Table II which shows the RSNR, UIQI, SAM, ERGAS and DD for the three methods. It can be seen that the HMC method provides slightly better results in terms of image restoration than the other methods. However, the proposed method allows the image covariance matrix and the noise variances to be estimated. The samples generated by the MCMC method can also be used to compute confidence intervals for the estimators (e.g., see error bars in Fig. 4).

E. Estimation of the noise variances

The proposed Bayesian method allows noise variances $s_{p,i}^2$ $(i = 1, \cdots, n_{\lambda,p}, p = 1, \cdots, P)$ to be estimated from the samples generated by the Gibbs sampler. The MMSE estimators of $s_{1,(\cdot)}^2$ and $s_{2,(\cdot)}^2$ are illustrated in Fig. 4. Graphically, the estimations can track the variations of the noise powers within tolerable discrepancy.

F. Fusion with and without DR

To illustrate the necessity of DR, the proposed method for fusing part of the Moffett field dataset ($64 \times 64 \times 176$) with $N_{\text{bi}} = 1000$ and $N_{\text{MC}} = 1000$ has been implemented with ($\widetilde{m}_\lambda = 10$) and without ($m_\lambda = 176$) DR. The qualitative results are displayed in Figure 5 and the quantitative results are reported in Table III. It is found that without reducing dimensionality, the fusion suffers from a substantial performance degradation and requires a more computationally expensive cost.

G. Robustness with respect to the knowledge of $\mathbf{F}_2$

The sampling algorithm summarized in Algorithm 2 requires the knowledge of the spectral response $\mathbf{F}_2$. However, this knowledge can be partially known in some practical applications. As the spectral response is the same for each vector $\boldsymbol{x}_i$ $(i = 1, \cdots, m_{\text{x}} m_{\text{y}})$, $\mathbf{F}_2$ is a block diagonal matrix whose blocks are f_2 of size $n_{\lambda,2} \times m_\lambda$, i.e., $\mathbf{F}_2 = \text{diag}[f_2, \cdots, f_2]$. This paragraph is devoted to testing the robustness of the proposed algorithm to the imperfect

knowledge of f_2. In order to analyze this robustness, a zero-mean white Gaussian error has been added to any non-zero component of f_2 as shown in the bottom of Fig. 1. Of course, the level of uncertainty regarding f_2 is controlled by the variance of the error denoted as σ_2^2. The corresponding FSNR is defined as FSNR $= 10\log_{10}\left(\frac{\|f_2\|_F^2}{m_\lambda n_{\lambda,2} s_2^2}\right)$ to adjust the knowledge of f_2. The larger FSNR, the more knowledge we have about f_2. The RSNRs between the reference and estimated images are displayed in Fig. 6 as a function of FSNR. Obviously, the performance of the proposed Bayesian fusion algorithm decreases as the uncertainty about f_2 increases. However, as long as the FSNR is above 8dB, the performance of the proposed method always outperforms the MAP and wavelet-based MAP methods. Thus, the proposed method is quite robust with respect to the imperfect knowledge of f_2.

H. Test on additional dataset

This section considers another reference image (the high spatial and high spectral image is a $128 \times 64 \times 93$ HS image with very high spatial resolution of 1.3 m/pixel) acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) optical sensor over the urban area of the University of Pavia, Italy. The flight was operated by the Deutsches Zentrum für Luft- und Raumfahrt (DLR, the German Aerospace Agency) in the framework of the HySens project, managed and sponsored by the European Union. This image was initially composed of 115 bands that have been reduced to 93 bands after removing the water vapor absorption bands (with spectral range from 0.43 to 0.86 μm). This image has received a lot of attention in the remote sensing literature [62]. The HS blurring kernel is the same as in paragraph V-A and the MS spectral response is a 4-band IKONOS-like reflectance spectral response. The noise level is defined by $\text{SNR}_{1,\cdot} = 35$dB for the first 43 bands and $\text{SNR}_{1,\cdot} = 30$dB for the remaining 50 bands of the HS image. For MS image, $\text{SNR}_{2,\cdot}$ is 30dB for all bands. The ground-truth, HS, MS and fusion results obtained with different algorithms are displayed in Fig. 7. The corresponding image quality measures are reported in Table IV. The estimates of the noise variances are shown in Fig. 8. These results are in good agreement with the performance obtained before.

I. Test on the synthetic non-Gaussian data

Some simulations with non-Gaussian data distributed according to a mixture of two Gaussian distributions (with different means and the same covariance matrix) have been conducted as well as the simulation with Gaussian distributed data. Tables V and VI show that the proposed method always provides the best results for Gaussian and non-Gaussian data which means the proposed method is quite robust to the non-Gaussianity of the image.

J. Application to pansharpening

The proposed algorithm can also be used for pansharpening, which is a quite important and popular application in the area of remote sensing. In this section, we focus on fusing panchromatic and hyperspectral images (HS+PAN), which is the extension of conventional pansharpening (MS+PAN). The HS image considered in this section was used in paragraph V-H whereas the PAN image was obtained by averaging all the high resolution HS bands. The SNR of the PAN image is 30dB. Apart from [25], [26], we also compare the results with the method of [63], which proposes a popular pansharpening method. The results are displayed in Fig. 9 and the quantitative results are reported in Table VII. The proposed Bayesian method still provides interesting results.

VI. Conclusions

This paper proposed a hierarchical Bayesian model to fuse multiple multi-band images with various spectral and spatial resolutions. The image to be recovered was assumed to be degraded according to physical transformations included within a forward model. An appropriate prior distribution, that exploited geometrical concepts encountered in spectral unmixing problems was proposed. The resulting posterior distribution was efficiently sampled thanks to a Hamiltonian Monte Carlo algorithm. Simulations conducted on pseudo-real data showed that the proposed method competed with the state-of-the-art techniques to fuse MS and HS images. These experiments also illustrated the robustness of the proposed method with respect to the misspecification of the forward model. Future work includes the estimation of the parameters involved in the forward model (e.g., the spatial and spectral responses of the sensors) to obtain a fully unsupervised fusion algorithm. The incorporation of spectral mixing

constraints for a possible improved spectral accuracy and the generalization to nonlinear degradations would also deserve some attention.

ACKNOWLEDGMENTS

The authors thank Dr. Paul Scheunders and Dr. Yifan Zhang for sharing the codes of [26] and Jordi Inglada, from Centre National d'Études Spatiales (CNES), for providing the LANDSAT spectral responses used in the experiments. The authors also acknowledge Prof. José M. Bioucas Dias for valuable discussions about this work that were handled during his visit in Toulouse within the CIMI Labex.

REFERENCES

[1] 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.

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

[3] T.-M. Tu, P. S. Huang, C.-L. Hung, and C.-P. Chang, "A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery," *IEEE Geosci. and Remote Sensing Lett.*, vol. 1, no. 4, pp. 309–312, 2004.

[4] H. Aanaes, J. Sveinsson, A. Nielsen, T. Bovith, and J. Benediktsson, "Model-based satellite image fusion," *IEEE Trans. Geosci. and Remote Sens.*, vol. 46, no. 5, pp. 1336–1346, May 2008.

[5] M. Joshi and A. Jalobeanu, "MAP estimation for multiresolution fusion in remotely sensed images using an IGMRF prior model," *IEEE Trans. Geosci. and Remote Sens.*, vol. 48, no. 3, pp. 1245–1255, March 2010.

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

[7] X. Ding, Y. Jiang, Y. Huang, and J. Paisley, "Pan-sharpening with a Bayesian nonparametric dictionary learning model," in *Proc. Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, Reykjavik, Iceland, 2014, pp. 176–184.

[8] C.-I Chang, *Hyperspectral Imaging: Techniques for Spectral detection and classification*. New York: Kluwer, 2003.

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

[10] 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.

[11] K. Kotwal and S. Chaudhuri, "A Bayesian approach to visualization-oriented hyperspectral image fusion," *Information Fusion*, vol. 14, no. 4, pp. 349–360, 2013.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] Z. Chen, H. Pu, B. Wang, and G.-M. Jiang, "Fusion of hyperspectral and multispectral images: A novel framework based on generalization of pan-sharpening methods," *IEEE Geosci. and Remote Sensing Lett.*, vol. 11, no. 8, pp. 1418–1422, Aug 2014.

[17] M. E. Winter and E. Winter, "Resolution enhancement of hyperspectral data," in *Proc. IEEE Aerospace Conference*, 2002, pp. 3–1523.

[18] G. Chen, S.-E. Qian, J.-P. Ardouin, and W. Xie, "Super-resolution of hyperspectral imagery using complex ridgelet transform," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 10, no. 03, 2012.

[19] D. Inc., "Worldview-3," http://www.satimagingcorp.com/satellite-sensors/WorldView3-DS-WV3-Web.pdf, Jan. 2013.

[20] N. Ohgi, A. Iwasaki, T. Kawashima, and H. Inada, "Japanese hyper-multi spectral mission," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Honolulu, Hawaii, USA, July 2010, pp. 3756–3759.

[21] 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.

[22] W. Dou, Y. Chen, X. Li, and D. Z. Sui, "A general framework for component substitution image fusion: An implementation using the fast image fusion method," *Comput. & Geosci.*, vol. 33, no. 2, pp. 219–228, 2007.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] ——, "Hyperspectral resolution enhancement using high-resolution multispectral imagery with arbitrary response functions," *IEEE Trans. Image Process.*, vol. 43, no. 3, pp. 455–465, March 2005.

[31] X. Otazu, M. Gonzalez-Audicana, O. Fors, and J. Nunez, "Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods," *IEEE Trans. Geosci. and Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, 2005.

[32] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4355–4368, 2009.

[33] M. V. Joshi, L. Bruzzone, and S. Chaudhuri, "A model-based approach to multiresolution fusion in remotely sensed images," *IEEE Trans. Geosci. and Remote Sens.*, vol. 44, no. 9, pp. 2549–2562, Sept. 2006.

[34] G. Casella and E. I. George, "Explaining the Gibbs sampler," *The American Statistician*, vol. 46, no. 3, pp. 167–174, 1992.

[35] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[36] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid Monte Carlo," *Physics Lett. B*, vol. 195, no. 2, pp. 216–222, Sept. 1987.

[37] R. M. Neal, "MCMC using Hamiltonian dynamics," *Handbook of Markov Chain Monte Carlo*, vol. 54, pp. 113–162, 2010.

[38] D. Fasbender, D. Tuia, P. Bogaert, and M. Kanevski, "Support-based implementation of Bayesian data fusion for spatial enhancement: Applications to ASTER thermal images," *IEEE Geosci. and Remote Sensing Lett.*, vol. 5, no. 4, pp. 598–602, Oct. 2008.

[39] M. Elbakary and M. Alam, "Superresolution construction of multispectral imagery based on local enhancement," *IEEE Geosci. and Remote Sensing Lett.*, vol. 5, no. 2, pp. 276–279, April 2008.

[40] J. B. Campbell, *Introduction to remote sensing*, 3rd ed. New-York, NY: Taylor & Francis, 2002.

[41] R. Schultz and R. Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Trans. Image Process.*, vol. 3, no. 3, pp. 233–242, May 1994.

[42] 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.

[43] 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.

[44] 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.

[45] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 747–758, March 2002.

[46] 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.

[47] S. Bidon, O. Besson, and J.-Y. Tourneret, "The adaptive coherence estimator is the generalized likelihood ratio test for a class of heterogeneous environments," *IEEE Signal Process. Lett.*, vol. 15, pp. 281–284, 2008.

[48] M. Bouriga and O. Féron, "Estimation of covariance matrices based on hierarchical inverse-Wishart priors," *J. of Stat. Planning and Inference*, 2012.

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

[50] R. M. Neal, "Probabilistic inference using Markov chain Monte Carlo methods," Dept. of Computer Science, University of Toronto, Tech. Rep. CRG-TR-93-1, Sept. 1993.

[51] H. Zhang, Y. Zhang, H. Li, and T. S. Huang, "Generative Bayesian image super resolution with natural image prior," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4054–4067, 2012.

[52] F. Orieux, O. Féron, and J.-F. Giovannelli, "Sampling high-dimensional Gaussian distributions for general linear inverse problems," *IEEE Signal Process. Lett.*, vol. 19, no. 5, pp. 251–254, 2012.

[53] D. Ceperley, Y. Chen, R. V. Craiu, X.-L. Meng, A. Mira, and J. Rosenthal, "Challenges and advances in high dimensional and high complexity monte carlo computation and theory," in *Banff International Research Station*, 2012.

[54] 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.

[55] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[56] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. and Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, 2006.

[57] R. Dianat and S. Kasaei, "Dimension reduction of optical remote sensing images via minimum change rate deviation method," *IEEE Trans. Geosci. and Remote Sens.*, vol. 48, no. 1, pp. 198–206, 2010.

[58] G. O. Roberts and J. S. Rosenthal, "Coupling and ergodicity of adaptive Markov Chain Monte Carlo algorithms," *J. of Appl. Probability*, vol. 44, no. 2, pp. pp. 458–475, 2007.

[59] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, 2002.

[60] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in *Proc. Int. Conf. Fusion of Earth Data*, Nice, France, Jan 2000, pp. 99–103.

[61] 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.

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

[63] S. Rahmani, M. Strait, D. Merkurjev, M. Moeller, and T. Wittman, "An adaptive IHS pan-sharpening method," *IEEE Geosci. and Remote Sensing Lett.*, vol. 7, no. 4, pp. 746–750, 2010.

ALGORITHM 1:

Hybrid Gibbs sampler

for $t = 1$ to N_{MC} **do**
% Sampling the image variances - see paragraph IV-A
Sample $\tilde{\boldsymbol{\Sigma}}_{\mathbf{u}}^{(t)}$ according to the conditional distribution (17),
% Sampling the high-resolved image - see paragraph IV-B
Sample $\tilde{\mathbf{u}}^{(t)}$ using an HMC algorithm detailed in Algorithm 2
% Sampling the noise variances - see paragraph IV-C
for $p = 1$ to P **do**
for $i = 1$ to $n_{\lambda,p}$ **do**
Sample $\tilde{s}_{p,i}^{2(t)}$ from the conditional distribution (25),
end for
end for
end for

ALGORITHM 2:

Hybrid Monte Carlo algorithm

% Momentum initialization
Sample $\tilde{\mathbf{m}}^{(\star)} \sim \mathcal{N}\left(\mathbf{0}_{\widetilde{M}}, \mathbf{I}_{\widetilde{M}}\right)$,
Set $\tilde{\mathbf{m}}^{(t)} \leftarrow \tilde{\mathbf{m}}^{(\star)}$,
% Leapfrogging
for $j = 1$ to N_{L} **do**
Set $\tilde{\mathbf{m}}^{(\star)} \leftarrow \tilde{\mathbf{m}}^{(\star)} - \frac{\varepsilon}{2}\nabla_{\mathbf{u}} U\left(\tilde{\mathbf{u}}^{(\star)}\right)$,
Set $\tilde{\mathbf{u}}^{(\star)} \leftarrow \tilde{\mathbf{u}}^{(\star)} + \varepsilon\tilde{\mathbf{m}}^{(\star)}$,
Set $\tilde{\mathbf{m}}^{(\star)} \leftarrow \tilde{\mathbf{m}}^{(\star)} - \frac{\varepsilon}{2}\nabla_{\mathbf{u}} U\left(\tilde{\mathbf{u}}^{(\star)}\right)$,
end for
% Accept/reject procedure, See (23)
Sample $w \sim \mathcal{U}\left([0,1]\right)$,
if $w < \rho_t$ **then**
$\tilde{\mathbf{u}}^{(t+1)} \leftarrow \tilde{\mathbf{u}}^{(\star)}$
else
$\tilde{\mathbf{u}}^{(t+1)} \leftarrow \tilde{\mathbf{u}}^{(t)}$
end if
Set $\tilde{\mathbf{x}}^{(t+1)} = \mathfrak{V}^T\tilde{\mathbf{u}}^{(t+1)}$
Run Algorithm 3 to update stepsize

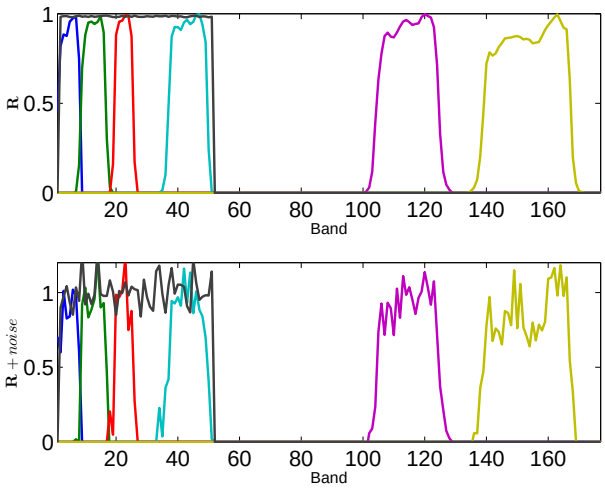

Fig. 1. LANDSAT spectral responses. (Top) without noise. (Bottom) with an additive Gaussian noise with FSNR = 8dB.

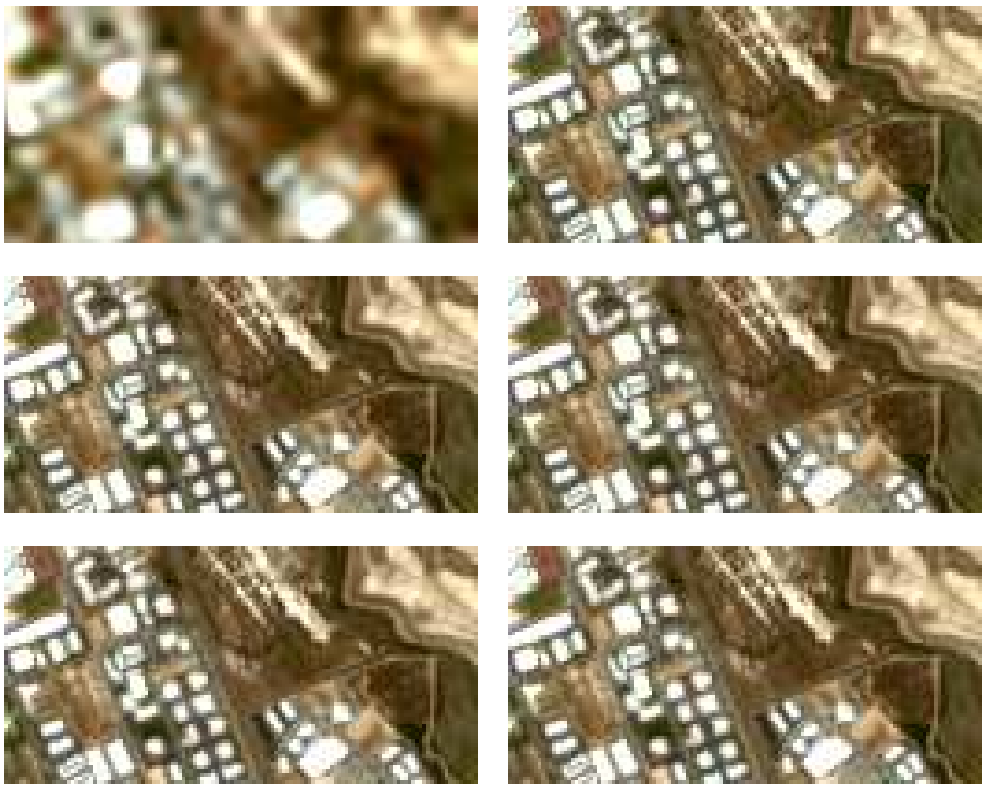

Fig. 2. AVIRIS dataset: (Top left) HS Image. (Top right) MS Image. (Middle left) MAP [25]. (Middle right) Wavelet MAP [26]. (Bottom left) Hamiltonian MCMC. (Bottom right) Reference.

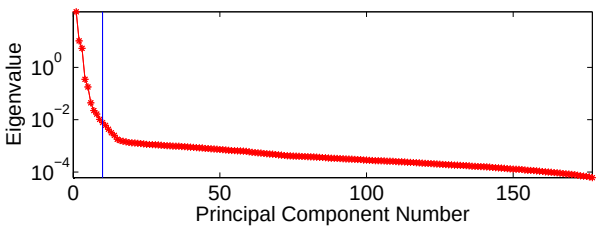

Fig. 3. Eigenvalues of $\boldsymbol{\Upsilon}$ for the HS image.

ALGORITHM 3:

Adjusting Stepsize

Update $\hat{\rho}_t$ with $N_{\text{a,t}}$: $\hat{\rho}_t = \frac{N_{\text{a,t}}}{N_{\text{W}}}$
% Burn-in ($t \leq N_{MC}$):
if $\hat{\rho}_t > \alpha_u$ **then**
  Set $\varepsilon = \beta_u \varepsilon$
else if $\hat{\rho}_t < \alpha_d$ **then**
  Set $\varepsilon = \beta_d \varepsilon$
end if
% After Burn in ($t > N_{MC}$):
if $\hat{\rho}_t > \alpha_u$ **then**
  Set $\varepsilon = [1 - (1 - \beta_u)\text{exp}(-0.01 \times (t - N_{\text{bi}}))]\varepsilon$,
else if $\hat{\rho}_t < \alpha_d$ **then**
  Set $\varepsilon = [1 - (1 - \beta_d)\text{exp}(-0.01 \times (t - N_{\text{bi}}))]\varepsilon$,
end if
$(t = N_{\text{bi}} + 1, \cdots, N_{\text{MC}})$

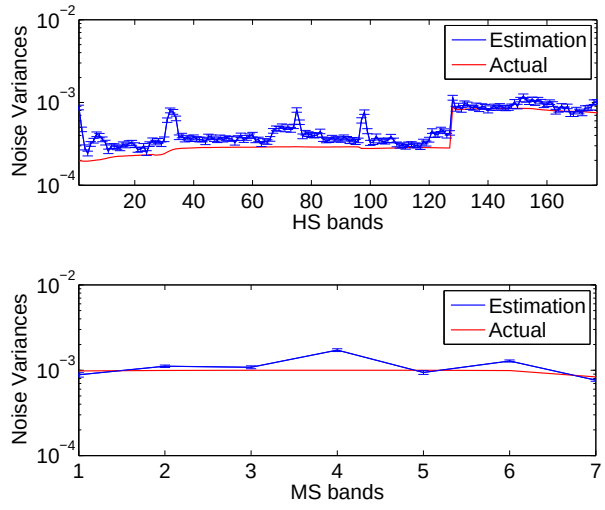

Fig. 4. Noise variances and their MMSE estimates. (Top) HS image. (Bottom) MS image.

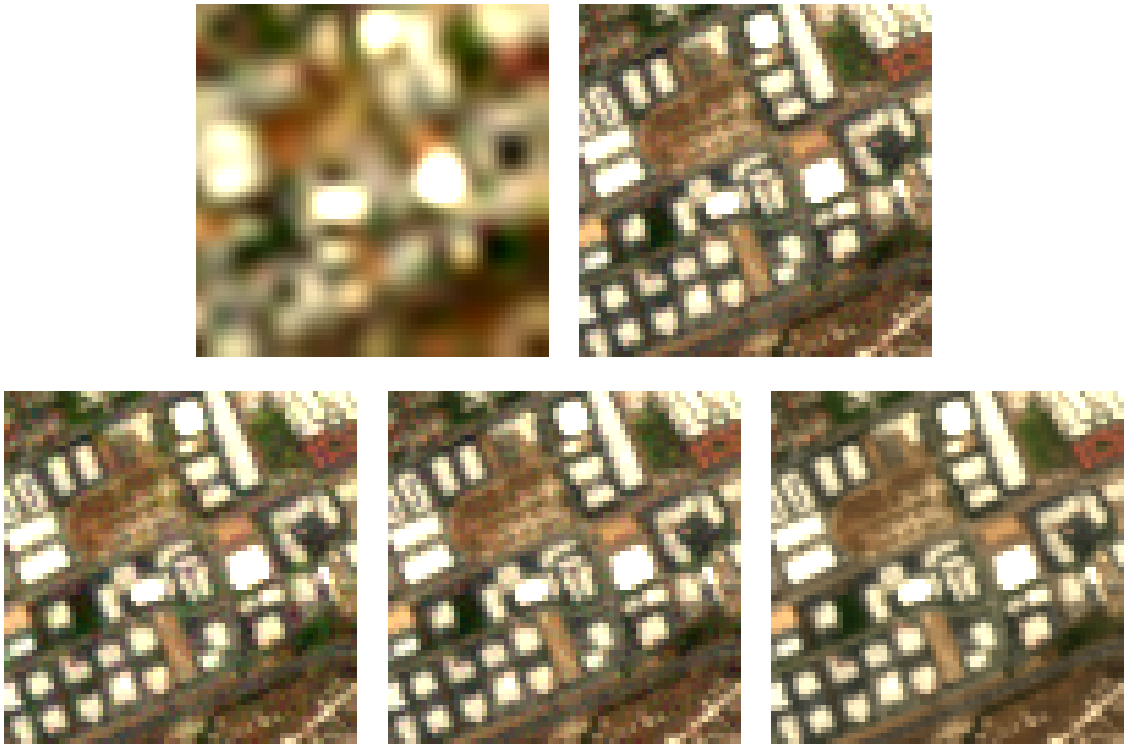

Fig. 5. Part of Moffett field dataset: (Top 1) HS. (Top 2) MS. (Bottom 1) Fusion without DR. (Bottom 2) Fusion with DR. (Bottom 3) Reference.

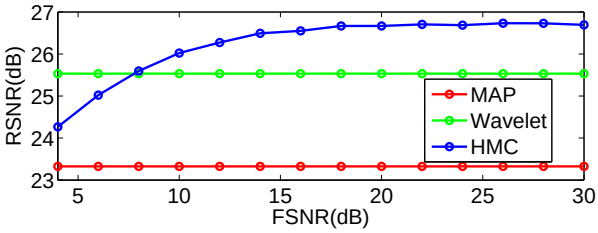

Fig. 6. Reconstruction errors of the different fusion methods versus FSNR.

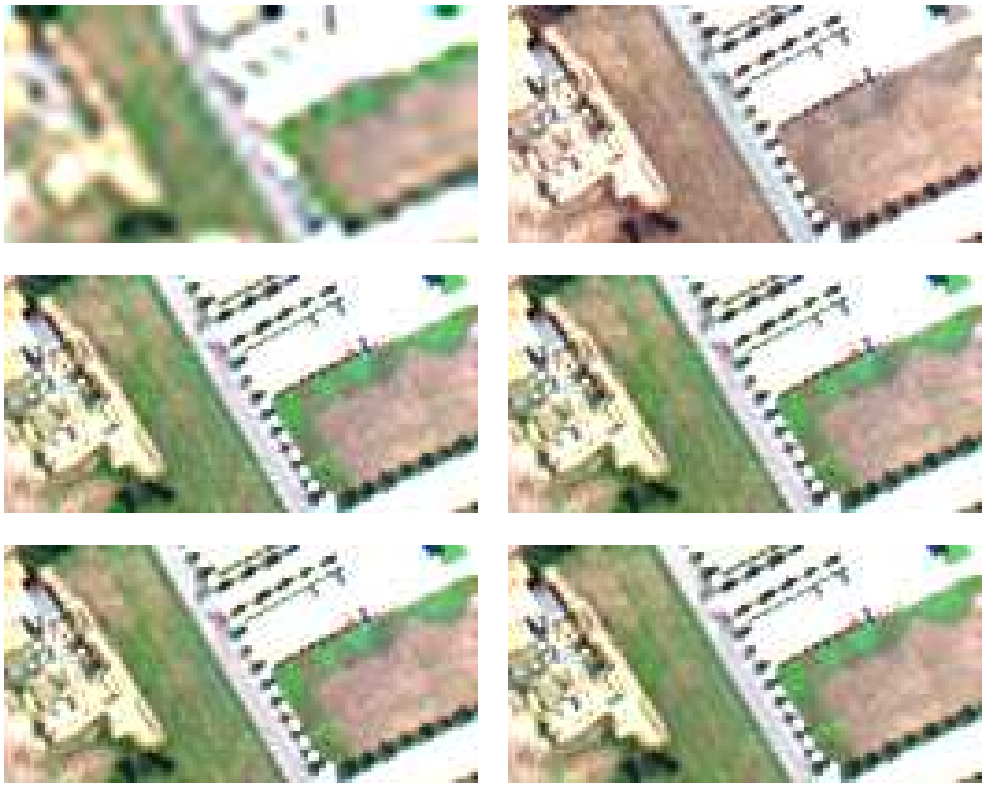

Fig. 7. ROSIS dataset: (Top left) HS Image. (Top right) MS Image. (Middle left) MAP [25]. (Middle right) Wavelet MAP [26]. (Bottom left) Hamiltonian MCMC. (Bottom right) Reference.

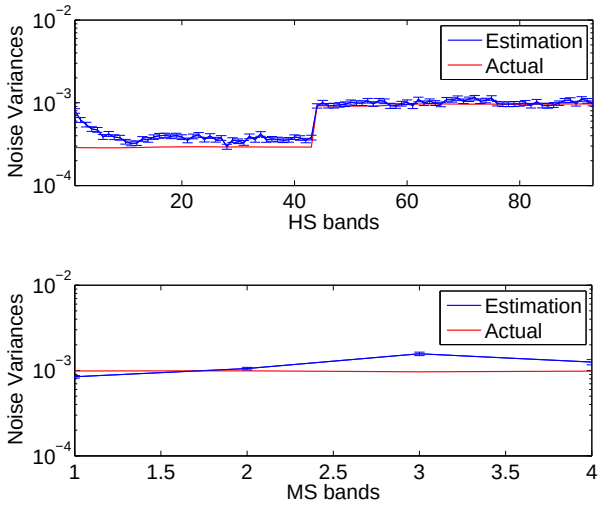

Fig. 8. Noise variances and their MMSE estimates (ROSIS dataset). (Top) HS image. (Bottom) MS image.

Fig. 9. ROSIS dataset: (Top left) Reference. (Top right) PAN Image. (Middle left) Adaptive IHS [63]. (Middle right) MAP [25]. (Bottom left) Wavelet MAP [26]. (Bottom right) Hamiltonian MCMC.

TABLE I
NOTATIONS

Notation	Definition	Size
$\mathbf{X}$	unobserved scene/target image	$m_{\mathrm{x}} \times m_{\mathrm{y}} \times m_{\lambda}$
$\mathbf{x}$	vectorization of $\mathbf{X}$	$m_{\mathrm{x}} m_{\mathrm{y}} m_{\lambda} \times 1$
$\boldsymbol{x}_i$	ith spectral vector of $\mathbf{x}$	$m_{\lambda} \times 1$
$\mathbf{u}$	vectorized image in subspace	$m_{\mathrm{x}} m_{\mathrm{y}} \widetilde{m}_{\lambda} \times 1$
$\boldsymbol{u}_i$	ith spectral vector of $\mathbf{u}$	$\widetilde{m}_{\lambda} \times 1$
$\bar{\boldsymbol{\mu}}_{\mathbf{u}}$	prior mean of $\mathbf{u}$	$m_{\mathrm{x}} m_{\mathrm{y}} \widetilde{m}_{\lambda} \times 1$
$\bar{\boldsymbol{\Sigma}}_{\mathbf{u}}$	prior covariance of $\mathbf{u}$	$m_{\mathrm{x}} m_{\mathrm{y}} \widetilde{m}_{\lambda} \times m_{\mathrm{x}} m_{\mathrm{y}} \widetilde{m}_{\lambda}$
$\boldsymbol{\mu}_{\boldsymbol{u}_i}$	prior mean of $\boldsymbol{u}_i$	$\widetilde{m}_{\lambda} \times 1$
$\boldsymbol{\Sigma}_{\boldsymbol{u}_i}$	prior covariance of $\boldsymbol{u}_i$	$\widetilde{m}_{\lambda} \times \widetilde{m}_{\lambda}$
P	number of multi-band images	1
$\mathbf{Z}_p$	pth remotely sensed images	$n_{\mathrm{x},p} \times n_{\mathrm{y},p} \times n_{\lambda,p}$
$\mathbf{z}_p$	vectorization of $\mathbf{Z}_p$	$n_{\mathrm{x},p} n_{\mathrm{y},p} n_{\lambda,p} \times 1$
$\mathbf{z}$	set of P observation $\mathbf{z}_p$	$n_{\mathrm{x},p} n_{\mathrm{y},p} n_{\lambda,p} P \times 1$

TABLE II
PERFORMANCE OF HS+MS FUSION METHODS IN TERMS OF: RSNR (DB), UIQI, SAM (DEG), ERGAS AND DD($\times 10^{-2}$) (AVIRIS DATASET).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time(s)
MAP	23.33	0.9913	5.05	4.21	4.87	**1.6**
Wavelet	25.53	0.9956	3.98	3.95	3.89	31
Proposed	**26.74**	**0.9966**	**3.40**	**3.77**	**3.33**	530

TABLE III
QUANTITATIVE PERFORMANCE OF MS + HS FUSION METHOD (FOR PART OF MOFFETT FIELD DATASET): RSNR (IN DB), UIQI, SAM (IN DEGREE), ERGAS, DD (IN 10^{-3}) AND TIME (IN SECOND).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
Fusion with DR	29.76	0.9944	1.603	0.878	6.007	1789
Fusion without DR	26.86	0.9893	2.573	1.255	8.598	21531

TABLE IV
PERFORMANCE OF HS+MS FUSION METHODS IN TERMS OF: RSNR (DB), UIQI, SAM (DEG), ERGAS AND DD($\times 10^{-2}$) (ROSIS DATASET).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time(s)
MAP [25]	26.58	0.9926	2.90	1.36	3.61	**1.5**
Wavelet [26]	26.62	0.9925	2.87	1.35	3.60	30
Proposed	**27.30**	**0.9933**	**2.60**	**1.24**	**3.27**	410

TABLE V
PERFORMANCE OF DIFFERENT FUSION METHODS FOR GAUSSIAN DATA.

Methods	RSNR(dB)	UIQI	SAM(degree)
Hardie	18.46	0.395	5.440
Zhang	19.17	0.429	5.374
MCMC	**19.40**	**0.470**	**5.218**

TABLE VI
PERFORMANCE OF DIFFERENT FUSION METHODS FOR NON-GAUSSIAN DATA.

Methods	RSNR(dB)	UIQI	SAM(degree)
Hardie	15.90	0.919	9.905
Zhang	16.23	0.924	9.968
MCMC	**16.40**	**0.926**	**9.809**

TABLE VII
PERFORMANCE OF HS+PAN FUSION METHODS IN TERMS OF: RSNR (DB), UIQI, SAM (DEG), ERGAS AND DD($\times 10^{-2}$) (ROSIS DATASET).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time(s)
AIHS [63]	16.69	0.9176	7.23	4.24	9.99	7.7
MAP [25]	17.54	0.9177	6.55	3.78	8.78	**1.4**
Wavelet [26]	18.03	0.9302	6.08	3.57	8.33	26
Proposed	**18.23**	**0.9341**	**6.05**	**3.49**	**8.20**	387