Hierarchical Beta Process with Gaussian Process Prior for Hyperspectral Image Super Resolution

Naveed Akhtar[1(✉)], Faisal Shafait[2], and Ajmal Mian[1]

[1] School of Computer Science and Software Engineering,
The University of Western Australia, 35 Stirling Highway,
6009 Crawley, WA, Australia
naveed.akhtar@research.uwa.edu.au, ajmal.mian@uwa.edu.au
[2] School of Electrical Engineering and Computer Science,
National University of Sciences and Technology, H-12, Islamabad, Pakistan
faisal.shafait@seecs.edu.pk

Abstract. Hyperspectral cameras acquire precise spectral information, however, their resolution is very low due to hardware constraints. We propose an image fusion based hyperspectral super resolution approach that employes a Bayesian representation model. The proposed model accounts for spectral smoothness and spatial consistency of the representation by using Gaussian Processes and a spatial kernel in a hierarchical formulation of the Beta Process. The model is employed by our approach to first infer Gaussian Processes for the spectra present in the hyperspectral image. Then, it is used to estimate the activity level of the inferred processes in a sparse representation of a high resolution image of the same scene. Finally, we use the model to compute multiple sparse codes of the high resolution image, that are merged with the samples of the Gaussian Processes for an accurate estimate of the high resolution hyperspectral image. We perform experiments with remotely sensed and ground-based hyperspectral images to establish the effectiveness of our approach.

Keywords: Hyperspectral · Super-resolution · Beta/Gaussian Process

1 Introduction

Spectral characteristics of materials are considered vital in remote sensing, medical imaging and forensics [1–6]. Recently, they have also shown improved performance in various computer vision tasks, e.g. recognition [7–9], document analysis [10,11], tracking [12], pedestrian detection [13] and segmentation [14]. Hyperspectral imaging is an emerging modality that can efficiently obtain high-fidelity spectral representations of a scene. Nevertheless, the low resolution of contemporary hyperspectral cameras is currently a bottleneck in its ubiquitous use [4,15,16].

Reflectance spectra are characterized by their intensity distributions over continuous wavelength ranges. Hence, hyperspectral cameras integrate scene

B. Leibe et al. (Eds.): ECCV 2016, Part III, LNCS 9907, pp. 103–120, 2016.
DOI: 10.1007/978-3-319-46487-9_7

radiance with hundreds of spectrally sharp bases, thereby requiring longer exposures. This results in a reduced resolution image. Moreover, it is not straightforward to use high resolution sensors in hyperspectral cameras because they

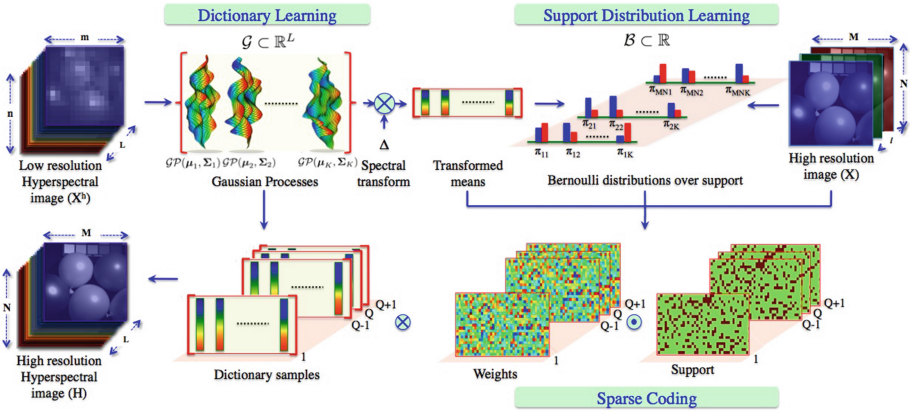

Fig. 1. Schematics: Using the proposed model, a set of Gaussian Processes ($\mathcal{GP}$s) is inferred for the spectra in hyperspectral image. The means of the $\mathcal{GP}$s are transformed according to the spectral channels of a high resolution image $\mathbf{X}$ of the same scene. The transformed means and $\mathbf{X}$ are used to compute a set $\mathcal{B}$ of Bernoulli distributions, signifying the activity level of $\mathcal{GP}$s in the sparse codes of $\mathbf{X}$. Multiple sparse codes of $\mathbf{X}$ are computed using the proposed model, each satisfying $\mathcal{B}$. The computed codes are used with the samples of $\mathcal{GP}$s to estimate the high resolution hyperspectral image.

further reduce the photon density that is already confined by the spectral filters. These constraints make hyperspectral image super resolution a particularly interesting research problem [17].

Currently, the resolution of the cameras that perform a gross quantization of the scene radiance (e.g. RGB and RGB-NIR), is orders of magnitude higher than that of hyperspectral cameras [4]. We collectively term the images acquired by these cameras as the *multi-spectral* images. In this work, we propose to take advantage of the high resolution of a multi-spectral image by merging its spatial patterns with the samples of Gaussian Processes [18], learned to represent the hyperspectral image of the same scene. Gaussian Processes provide an excellent tool for modeling natural spectra [19] because they can easily incorporate the regularly occurring smoothness of spectra. To learn the Gaussian Processes, we propose a novel Bayesian representation model. Our model also incorporates spatial consistency in the representation by employing a kernel in the hierarchical formulation of the Beta Process [20]. We provide a detailed Markov Chain Monte Carlo (MCMC) analysis [21] for the Bayesian inference using our model.

Employing the proposed model, we develop an approach for hyperspectral image super resolution, shown in Fig. 1. The approach first uses the model to infer Gaussian Processes to represent the low resolution hyperspectral image. Then, the mean parameters of the processes are transformed to match the spectral quantization of the high resolution multi-spectral image. The model is then

employed with the transformed means and the multi-spectral image to infer a set of Bernoulli distributions. These distributions record the activity level of the Gaussian Processes in the representation of the multi-spectral image. Exploiting these distributions, the model is later applied to infer a set of sparse codes for the multi-spectral image, that are used with the samples of the Gaussian Processes to obtain a high resolution hyperspectral image. Experiments on remotely sensed and ground-based hyperspectral images show that our approach obtains higher fidelity super resolution images as compared to the state-of-the-art approaches.

2 Related Work

Hyperspectral imaging has been used in remote sensing for over three decades [22]. However, hyperspectral instruments installed in contemporary remote sensing platforms still lack in spatial resolution [5,16]. High cost of replacing these instruments and hardware limitations have motivated significant research in signal processing based resolution enhancement of remote sensing imagery [16]. Pan-sharpening [23] is one of the common techniques used for this purpose. It fuses a pan-chromatic high resolution image with a hyperspectral image to improve its spatial resolution. Wavelet based pan-sharpening [24], Intensity-Hue-Saturation transform based methods [25,26] and pan-sharpening with principal component analysis [27] are a few representative examples of this category.

Whereas sharp spatial patterns are apparent in pan-sharpened images, the resulting images often suffer from significant spectral distortions [28]. This can be attributed to the spectral limitations of the pan-chromatic images [29]. Therefore, [30] and [31] fused multi-spectral images with hyperspectral images. They used hyperspectral unmixing [32] for the image fusion. However, these methods assume relatively small difference between the spectral resolutions of the images being fused. Moreover, they also under-perform when the imaged scenes are highly mixed [33]. For such cases, their performance have been improved by Zurita-Milla et al. [34] by employing a sliding window technique.

Recently, matrix factorization based approaches have consistently shown state-of-the-art performance in hyperspectral image super resolution. These approaches are divided into three categories based on their underlying assumptions. The methods in the **first category** [4,16,17,33] assume that only the spectral transform between the images being fused is known beforehand. Kawakami et al. [4] factored a hyperspectral image and an RGB image into their respective bases and used the sparse codes of the RGB image with the basis of the hyperspectral image. Huang et al. [33] applied a similar approach to remote sensing imagery, using singular value decomposition to learn the bases, whereas a sparsity controlled approach was employed in [17] for non-negative matrix factorization. Motivated by the success of Bayesian matrix factorization in RGB and gray scale image super resolution [35], Akhtar et al. [16] developed a Bayesian approach for hyperspectral image super resolution.

The approaches in the **second category** [36–40] additionally assume priori knowledge of the spatial transform between the images. Lanaras et al. [36,37] formulated hyperspectral image super resolution as a coupled unmixing problem and proposed a matrix factorization algorithm to solve it. Their approach

exploits the physical constraints generally followed by the spectral signatures. Wycoff et al. [38] proposed to use the ADMM algorithm [41] for the factorization procedure. A variational based approach is also proposed by Wei et al. [39] for the same purpose. In [40], Yokoya et al. exploited the coupling between the images being fused and developed a coupled matrix factorization technique for hyperspectral image super resolution. They also noted that in practice, it is challenging to obtain an accurate estimate of the spatial transform between the images being fused. Nevertheless, accurate transform is generally assumed and exploited by the approaches in this category. The methods in the **third category** [29] assume availability of high resolution hyperspectral training data. This implicitly imposes both of the above assumptions in addition to the requirement that abundant training data is available despite hardware limitations.

Overall, the least restrictive methods are those belonging to the first category, in which our approach also falls. Akhtar et al. [16] have established the usefulness of Bayesian sparse representation for hyperspectral image super resolution. Nevertheless, their approach has two major limitations. That is, it neither considers the spectral smoothness nor the spatial consistency of the representation of the nearby image pixels. In this paper, we address these limitations by (1) employing Gaussian Processes for the spectral signatures to incorporate smoothness into their representation; and (2) enforcing spatial consistency of the representation with a suitable kernel in the hierarchical formulation of the Beta Process.

3 Problem Formulation

Let us denote an acquired hyperspectral image with L spectral bands by $\mathbf{X}^h \in \mathbb{R}^{m \times n \times L}$. Let $\mathbf{X} \in \mathbb{R}^{M \times N \times l}$ be the high resolution image of the same scene obtained by a multi-spectral sensor. We aim at estimating a high resolution hyperspectral image $\mathbf{H} \in \mathbb{R}^{M \times N \times L}$ by merging $\mathbf{X}$ with $\mathbf{X}^h$. The two available images are considered to be linear mappings of the target image. Formally, $\mathbf{X}^h = \boldsymbol{\Omega}^h(\mathbf{H})$ and $\mathbf{X} = \boldsymbol{\Omega}(\mathbf{H})$, where $\boldsymbol{\Omega}^h : \mathbb{R}^{M \times N \times L} \rightarrow \mathbb{R}^{m \times n \times L}$ and $\boldsymbol{\Omega} : \mathbb{R}^{M \times N \times L} \rightarrow \mathbb{R}^{M \times N \times l}$. Moreover, we consider $M \gg m, N \gg n$ and $L \gg l$, to appropriately model the practical conditions. We neither assume prior knowledge of the spatial transform between the images being fused nor the availability of high resolution hyperspectral training data. Following [4,16,17,33,37], we assume aligned $\mathbf{X}$ and $\mathbf{X}^{\mathbf{h}}$. In practice, accurate alignment is possible using beam-splitting mechanism [13].

We denote the number of pure spectral signatures (a.k.a endmembers) in an imaged scene by K and the k^{th} signature by $\boldsymbol{\psi}_k \in \mathbb{R}^L$. These signatures represent the reflectances of spectrally distinct materials in the scene Let $\boldsymbol{\Psi} \in \mathbb{R}^{L \times K}$ be the matrix comprising these spectral signatures. Thus, a pixel $\mathbf{x}^h \in \mathbb{R}^L$ of the hyperspectral image can be represented as $\mathbf{x}^h = \boldsymbol{\Psi}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} \in \mathbb{R}^K$ is a coefficient vector. A pixel of the multi-spectral image can similarly be represented as $\mathbf{x} = \widetilde{\boldsymbol{\Psi}}\boldsymbol{\beta}$, where $\widetilde{\boldsymbol{\Psi}} \in \mathbb{R}^{l \times K}$ is obtained by transforming $\boldsymbol{\Psi}$, such that $\widetilde{\boldsymbol{\Psi}} = \boldsymbol{\Delta}\boldsymbol{\Psi}$. Following the literature (see Sect. 2), we assume *a priori* knowledge of the spectral transformation operator $\boldsymbol{\Delta} \in \mathbb{R}^{l \times L}$. Since the exact value of K is generally

unknown, we allow for $K > L$. Hence, we expect the coefficient vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to be sparse. Adopting the naming conventions from the sparse representation literature [42], we refer to $\boldsymbol{\Psi}$ and $\widetilde{\boldsymbol{\Psi}}$ as *dictionaries*; to their columns as *dictionary atoms*; and the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as *sparse codes*.

4 Proposed Approach

Our approach utilizes a hierarchical Bayesian sparse representation model, that we propose to represent hyperspectral images. The model is used to infer an ensemble of dictionaries for the hyperspectral image in the form of Gaussian Processes ($\mathcal{GP}$s) [18]. $\mathcal{GP}$s are particularly well-suited for forming dictionaries for hyperspectral images because they can easily incorporate the relative smoothness of the natural spectra with appropriate kernels. The approach transforms the mean parameters of the $\mathcal{GP}$s and uses them with the multi-spectral image $\mathbf{X}$ to estimate the activity level of the $\mathcal{GP}$s in the sparse representation of $\mathbf{X}$. We again utilize the proposed model for this estimation. Lastly, the model is used to learn multiple sparse codes of $\mathbf{X}$ that are combined with the samples from the $\mathcal{GP}$s to compute the high resolution hyperspectral image. The proposed approach is summarized in Fig. 1.

4.1 Dictionary Learning

Below, we describe the proposed representation model and its Bayesian inference process that results in the learning of the $\mathcal{GP}$s (i.e. dictionary atoms). Other stages of our approach also exploit the same model, with minor variations in the inference process. We explain those variations in Sects. 4.2 and 4.3.

Representation model: We model the i^{th} pixel of a hyperspectral image as $\mathbf{x}_i^h = \boldsymbol{\Psi}\boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_i^h$, where $\boldsymbol{\epsilon}_i^h \in \mathbb{R}^L$ represents noise. The following hierarchical Bayesian representation model is proposed to compute the probability distributions over the dictionary atoms and the sparse codes:

$$\mathbf{x}_i^h = \boldsymbol{\Psi}\boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_i^h \tag{1}$$

$$\begin{aligned}
\boldsymbol{\psi}_k &\sim \mathcal{GP}(\boldsymbol{\psi}_k|\mathbf{0}, \boldsymbol{\Sigma}_k) & z_{ik} &\sim \text{Bern}(z_{ik}|\pi_{ik}) \\
\boldsymbol{\Sigma}_k(\theta_a, \theta_b) &= \frac{1}{\eta_k}\exp\left(\frac{-|\theta_b - \theta_a|}{\eta_o}\right) & \pi_{ik} &= \mathbf{e}^{\mathrm{T}}(\boldsymbol{\kappa} \odot \boldsymbol{\Xi}_{ik})\mathbf{e} \\
\eta_k &\sim \text{Gam}(\eta_k|a_o, b_o) & \boldsymbol{\Xi}_{ik}(q,r) &\sim \text{Beta}\left(\boldsymbol{\Xi}_{ik}(q,r)|e_o\rho_k, f_o(1-\rho_k)\right) \\
\alpha_{ik} &= w_{ik}z_{ik} & \rho_k &\sim \text{Beta}\left(\rho_k|\frac{g_o}{K}, \frac{h_o(K-1)}{K}\right) \\
w_{ik} &\sim \mathcal{N}(w_{ik}|0, \lambda_w^{-1}) & \boldsymbol{\epsilon}_i^h &\sim \mathcal{N}(\boldsymbol{\epsilon}_i^h|\mathbf{0}, \lambda_\epsilon^{-1}\mathbf{I}_L) \\
\lambda_w &\sim \text{Gam}(\lambda_w|c_o, d_o) & \lambda_\epsilon &\sim \text{Gam}(\lambda_\epsilon|k_o, l_o).
\end{aligned}$$

In the above expressions, $\mathcal{N}$, Gam, Bern and Beta respectively denote the Normal, Gamma, Bernoulli and the Beta probability distributions. The symbol

$\odot$ denotes the element-wise product and the subscript 'o' signifies the hyper-parameters of the distributions whose values remain fixed during the inference process (discussed below). For reading convenience, we explain the remaining symbols along the relevant discussion on the representation model.

In the proposed model, we let the k^{th} dictionary atom $\boldsymbol{\psi}_k$ to be a sample drawn from a Gaussian Process. We define the kernel of a Gaussian Process, i.e. $\boldsymbol{\Sigma}_k \in \mathbb{R}^{L \times L}$, such that it promotes high correlations between the adjacent coefficients of $\boldsymbol{\psi}_k$. Recall that, $\boldsymbol{\psi}_k$ signifies a spectral signature in our formulation. Thus, the kernel incorporates the relative smoothness of the spectra in the proposed model. In the given definition of the kernel, θ_t denotes the wavelength at the t^{th} channel of the image, whereas $|.|$ represents the absolute value. Such an exponential form of the kernel is common for Gaussian Processes [18]. However, we also include an additional scaling parameter η_k in the kernel to allow it to adjust to the observed data. The value of this parameter is automatically inferred in our approach. We place a non-informative Gamma prior over η_k; such that $\text{Gam}(\eta_k | a_o, b_o) = \frac{b_o^{a_o} \eta_k^{(a_o - 1)}}{\Gamma(a_o)} \exp(-b_o \eta_k)$, where $\Gamma(.)$ is the well-known Gamma function. The remaining model also utilizes the same functional form of the Gamma prior. In our approach, the value of η_o is fixed to $1/L$.

We compute the k^{th} coefficient α_{ik} of $\boldsymbol{\alpha}_i$ as the product of a sample w_{ik} from a Normal distribution, with precision λ_w; and a sample $z_{ik} \in \{0, 1\}$ from a Bernoulli distribution, with parameter $0 \leq \pi_{ik} \leq 1$. Thus, according to our model, a pixel $\mathbf{x}_i^h$ selects the k^{th} dictionary atom in its representation with a probability π_{ik}. This statistical modeling of $\boldsymbol{\alpha}_i$ is inspired by the weighted Beta-Bernoulli Process [43]. Zhou et al. [43] showed that Bernoulli priors over the support of $\boldsymbol{\alpha}_i$, with conjugate Beta priors, successfully capture the intrinsic sparsity of the signal. On the other hand, Normal priors (over w_{ik}) take care of the coefficient weights. For a similar situation, Akhtar et al. [16] directly placed a Beta probability prior over the Bernoulli distribution parameter. However, their approach neither forces the atoms of the dictionary nor the sparse codes to be similar for the nearby pixels in the image. To enforce this spatial consistency in the representation model, we compute π_{ik} as a weighted sum of the samples from a Beta probability distribution. In our approach, these Beta distribution samples signify the probabilities of selection of $\boldsymbol{\psi}_k$ in the representations of the nearby pixels of $\mathbf{x}_i^h$. Hence, $\boldsymbol{\psi}_k$ has more chances to get selected for $\mathbf{x}_i^h$, if that dictionary atom is also used in the representations of the nearby pixels of $\mathbf{x}_i^h$.

Concretely, we let $\pi_{ik} = \mathbf{e}^{\text{T}}(\boldsymbol{\kappa} \odot \boldsymbol{\Xi}_{ik})\mathbf{e}$, where $\mathbf{e} \in \mathbb{R}^P$ is a vector of 1s; $\boldsymbol{\kappa} \in \mathbb{R}^{P \times P}$ is the spatial kernel and $\boldsymbol{\Xi}_{ik} \in \mathbb{R}^{P \times P}$ comprises the samples of a Beta probability distribution. Here, P is the size of the image patch that contains the neighborhood pixels centered around $\mathbf{x}_i^h$. We compute a coefficient of $\boldsymbol{\kappa}$ at index (q, r) as $\boldsymbol{\kappa}(q, r) = \exp(-\|\mathcal{I}_i - \mathcal{I}_j\|_2 / \sigma_o)$, where $\mathcal{I}_t$ denotes the index of the t^{th} pixel in the image and σ_o decides the kernel width. We sample $\boldsymbol{\Xi}_{ik}(q, r)$ from a Beta distribution and, keeping in view the physical significance of these samples, we place a second Beta prior over the parameter ρ_k of the distribution. The second prior plays the same role in our model that is played by the Beta prior in the model employed by Akhtar et al. [16]. However, the resulting stochastic

process uses $\mathcal{GP}$ as the base measure in our model, instead of a Multi-variate Gaussian, as in [16]. We also note that the notion of hierarchical construction of the Beta Process was first introduced by Thibaux and Jordan [20]. However, our model differs from their proposal, as they did not use a kernel for computing π_{ik} and employed a Normal distribution as the base measure.

Following the literature [4,16] we consider white noise in our model. In Eq. 1, the covariance matrix of the noise distribution is denoted as $\lambda_\epsilon^{-1}\mathbf{I}_L$, where $\mathbf{I}_L \in \mathbb{R}^{L\times L}$ is the identity matrix. We place a Gamma prior over the noise precision λ_ϵ. This allows a Bayesian model to automatically adjust to the noise level of the observed data [43].

Inference: We perform Markov Chain Monte Carlo (MCMC) analysis [21] to infer the posterior probability distributions over the dictionary atoms and the sparse codes using our model. Below, we derive the expressions of the probability distributions that are sampled sequentially to perform the MCMC analysis. The sampling process is carried out iteratively.

Sampling $\boldsymbol{\psi}_k$: For brevity, let us denote the contribution of $\boldsymbol{\psi}_k$ to $\mathbf{x}_i^h$ as $\mathbf{x}_{i_{\psi_k}}^h = \mathbf{x}_i^h - \boldsymbol{\Psi}(\mathbf{w}_i \odot \mathbf{z}_i) + \boldsymbol{\psi}_k(w_{ik}z_{ik})$, where $\mathbf{w}_i, \mathbf{z}_i \in \mathbb{R}^K$ are the vectors formed by concatenating w_{ik} and $z_{ik}, \forall_k$. According to our model, the posterior probability distribution over $\boldsymbol{\psi}_k$ can be expressed as:

$$p(\boldsymbol{\psi}_k|-) \propto \prod_{i=1}^{mn} \mathcal{N}(\mathbf{x}_{i_{\psi_k}}^h | \boldsymbol{\psi}_k(w_{ik}z_{ik}), \lambda_\epsilon^{-1}\mathbf{I}_L)\mathcal{GP}(\boldsymbol{\psi}_k|\mathbf{0}, \boldsymbol{\Sigma}_k).$$

Exploiting the linear Gaussian model [44], it can be shown that $\mathcal{GP}(\boldsymbol{\psi}_k|\boldsymbol{\mu}_k, \widehat{\boldsymbol{\Sigma}}_k)$ must be used to sample this posterior probability distribution, where

$$\widehat{\boldsymbol{\Sigma}}_k = \left(\boldsymbol{\Sigma}_k^{-1} + \lambda_\epsilon \sum_{i=1}^{mn}(w_{ik}z_{ik})^2\right)^{-1}, \quad \boldsymbol{\mu}_k = \lambda_\epsilon \widehat{\boldsymbol{\Sigma}}_k \sum_{i=1}^{mn}(w_{ik}z_{ik})\mathbf{x}_{i_{\psi_k}}^h.$$

Sampling η_k: According to the proposed model, the posterior probability distribution over η_k can be written as $p(\eta_k|-) \propto \mathcal{GP}\left(\boldsymbol{\Psi}|\mathbf{0}, \frac{1}{\eta_k}\widetilde{\boldsymbol{\Sigma}}_k\right) \mathrm{Gam}(\eta_k|a_o, b_o)$, where $\mathbf{0}$ is a vector of zeros in $\mathbb{R}^L$ and $\widetilde{\boldsymbol{\Sigma}}_k = \exp(-|\theta_b - \theta_a|/\eta_o)$. The right hand side of the proportionality can be further be expanded into the following expression: $\frac{b_o^{a_o}\left(\sqrt[L]{2\pi}\right)^{-1}\eta_k^{a_o-1}}{\Gamma(a_o)\sqrt{\det\left(\frac{1}{\eta_k}\widetilde{\boldsymbol{\Sigma}}_k\right)}} \exp\left\{-\frac{\eta_k}{2}\boldsymbol{\psi}_k^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}_k\boldsymbol{\psi}_k - b_o\eta_k\right\}$, which is proportional to the Gamma probability distribution $\mathrm{Gam}\left(\eta_k|a_o + \frac{L}{2}, b_o + \frac{1}{2}\boldsymbol{\psi}_k^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}_k\boldsymbol{\psi}_k\right)$, that we use to sample η_k.

Sampling w_{ik}: The posterior distribution over w_{ik} has the functional form $p(w_{ik}|-) \propto \mathcal{N}(\mathbf{x}_{i_{\psi_k}}^h|\boldsymbol{\psi}_k(w_{ik}z_{ik}), \lambda_\epsilon^{-1}\mathbf{I}_L)\mathcal{N}(w_{ik}|0, \lambda_w^{-1})$. Again, making the use of the linear Gaussian model, w_{ik} can be sampled from $\mathcal{N}(w_{ik}|\mu_w, \widehat{\lambda}_w^{-1})$, where $\widehat{\lambda}_w = \lambda_w + \lambda_\epsilon z_{ik}^2\boldsymbol{\psi}_k^{\mathrm{T}}\boldsymbol{\psi}_k$ and $\mu_w = z_{ik}\boldsymbol{\psi}^{\mathrm{T}}\mathbf{x}_{i_{\psi_k}}^h\lambda_\epsilon/\lambda_w$.

Sampling λ_w: According to our model, the posterior over λ_w can be written as: $p(\lambda_w|-) \propto \prod_{i=1}^{mn} \mathcal{N}(\mathbf{w}_i|\mathbf{0}, \lambda_w^{-1}\mathbf{I}_K)\text{Gam}(\lambda_w|c_o, d_o)$. Hence, employing the conjugacy between the Normal and the Gamma distributions, we sample λ_w from $\text{Gam}(\lambda_w|\frac{Kmn}{2} + c_o, \frac{1}{2}\sum_{i=1}^{mn} \mathbf{w}_i^{\text{T}}\mathbf{w}_i + d_o)$.

Sampling z_{ik}: We can write $p(z_{ik}|-) \propto \mathcal{N}(\mathbf{x}_{i_{\psi_k}}^h|\boldsymbol{\psi}_k(w_{ik}z_{ik}), \lambda_\epsilon^{-1}\mathbf{I}_L)\text{Bern}(z_{ik}|\pi_{ik})$. From here, it is easy to show that z_{ik} must be sampled from the Bernoulli distribution: $\text{Bern}\Big(z_{ik}|\frac{\gamma}{1-\pi_{ik}+\gamma}\Big)$, where $\gamma = \exp\Big(-\frac{\lambda_\epsilon}{2}(\boldsymbol{\psi}_k^{\text{T}}\boldsymbol{\psi}_k w_{ik}^2 - 2w_{ik}\mathbf{x}_{i_{\psi_k}}^{h\text{T}}\boldsymbol{\psi}_k)\Big)$.

Sampling $\boldsymbol{\Xi}_{ik}(q,r)$: Let $\boldsymbol{\Xi}_{ik}(q,r) = \xi_{ik}$ and $\pi_{ik}^{\sim(q,r)} = \pi_{ik} - \boldsymbol{\kappa}(q,r)\xi_{ik}$. Using these notations, we can express the posterior distribution over $\boldsymbol{\Xi}_{ik}(q,r)$ as:

$$p(\xi_{ik}|-) \propto \text{Beta}\left(\xi_{ik}|e_o\rho_k, f_o(1-\rho_k)\right) \prod_{q,r\in\{1,\ldots,P\}} \text{Bern}\left(z_{ik}|\boldsymbol{\kappa}(q,r)\xi_{ik} + \pi_{ik}^{\sim(q,r)}\right).$$

This distribution can not be directly sampled. However, considering its functional form, it is possible to associate the popularity of the k^{th} dictionary atom to the pixels in the neighborhood $\aleph$ of $\mathbf{x}_i^h$ as:

$$\xi_{ik} \sim \text{Beta}\Big(\xi_{ik}\Big|e_o\rho_k + \sum_{\{j:(p,q)\in\aleph\}} z_{jk}, f_o(1-\rho_k) + \sum_{\{j:(p,q)\in\aleph\}} (1-z_{jk})\Big).$$

We use the above distribution as the proposal distribution $\mathcal{Q}$ in the Metropolis Hastings (MH) algorithm [45] and in step τ of MH algorithm, we draw $\xi_{ik}^* \sim \mathcal{Q}(\xi_{ik}|\xi_{ik}^\tau)$, where ξ_{ik}^τ is the current ξ_{ik}, and accept the sample with a probability:

$$\varrho = \min\left\{1, \frac{p(\xi_{ik}^*)\mathcal{Q}(\xi_{ik}^\tau|\xi_{ik}^*)}{p(\xi_{ik}^\tau)\mathcal{Q}(\xi_{ik}^*|\xi_{ik}^\tau)}\right\}$$

It can be shown that the fraction in the brackets can be analytically computed as follows:

$$\left(\frac{\xi_{ik}^\tau}{\xi_{ik}^*}\right)^{\sum_{\{j:(p,q)\in\aleph\}} z_{jk}} \left(\frac{1-\xi_{ik}^\tau}{1-\xi_{ik}^*}\right)^{\sum_{\{j:(p,q)\in\aleph\}} (1-z_{jk})} \prod_{\{j:(p,q)\in\aleph\}} \left(1+\frac{\Upsilon}{\xi_{ik}^\tau}\right)^{z_{jk}} \left(1-\frac{\Upsilon}{1-\xi_{ik}^\tau}\right)^{1-z_{jk}},$$

where $\Upsilon = \kappa(p,q)(\xi_{ik}^* - \xi_{ik}^\tau)$.

Sampling ρ_k: The posterior on ρ_k can be expressed as:

$$p(\rho_k|-) \propto \text{Beta}\left(\rho_k|g_o/K, h_o(K-1)/K\right) \prod_{i=1}^{mn} \text{Beta}\left(\xi_{ik}|e_o\rho_k, f_o(1-\rho_k)\right).$$

For analytical simplification, we let $e_o = f_o = 1$. By expanding the expressions for the distributions and neglecting the constant terms, we can show that $\forall i \in \{1,\ldots,mn\}$:

$$p(\rho_k|-) \propto \frac{\rho_k^{\left(\frac{g_o}{K}-1\right)}(1-\rho_k)^{\left(\frac{h_o(K-1)}{K}-1\right)}}{(\Gamma(\rho_k)\Gamma(1-\rho_k))^{mn}} \left(\frac{\xi_{ik}}{1-\xi_{ik}}\right)^{\rho_k mn}.$$

Here, the term depending on ξ_{ik} can be simplified to $\exp\left(\rho_k \sum_{i=1}^{mn} \log \frac{\xi_{ik}}{1-\xi_{ik}}\right)$. Therefore, following Zhou et al. [46], we use slice sampling algorithm [47] to sample ρ_k from the following exponential distribution:

$$\rho_k \sim \text{Exp}\left(\sum_{i=1}^{mn} \log \frac{\xi_{ik}}{1-\xi_{ik}}\right) \mathcal{R}(\varsigma, \upsilon, \omega),$$

where $\mathcal{R}(\varsigma, \upsilon, \omega)$ is the range of ρ_k and $\upsilon \sim \text{Unif}(0, (1-\rho_k)^{\left(\frac{h_o(K-1)}{K}-1\right)})$, $\varsigma \sim \text{Unif}(0, \rho_k^{(\frac{g_o}{K}-1)})$ and $\omega \sim \text{Unif}(0, \sin^{mn}(\pi\rho_k))$. Here, Unif denotes the uniform distribution. We restrict $0 < \rho_k < 1$ and exploit the fact that $\Gamma(\rho_k)\Gamma(1-\rho_k) \propto 1/\sin(\pi\rho_k)$ [48] to arrive at the expression for sampling ω.

Sampling λ_ϵ: We have $p(\lambda_\epsilon|-) \propto \prod_{i=1}^{mn} \mathcal{N}(\mathbf{x}_i^h|\boldsymbol{\Psi}(\mathbf{w}_i \odot \mathbf{z}_i), \lambda_\epsilon^{-1}\mathbf{I}_L)\text{Gam}(\lambda_\epsilon|k_o, l_o)$. Again, by employing the conjugacy of the probability distributions we can sample λ_ϵ from $\text{Gam}\left(\lambda_\epsilon|\frac{Lmn}{2} + k_o, \frac{1}{2}\sum_{i=1}^{mn}||\mathbf{x}_i^h - \boldsymbol{\Psi}(\mathbf{w}_i \odot \mathbf{z}_i)||_2^2 + l_o\right)$, where $||.||_2$ denotes the vector ℓ-2 norm.

4.2 Support Distribution Learning

Once the inference is complete, we get a set $\mathcal{G} \subset \mathbb{R}^L$ of K Gaussian Processes, where each process represents a probability distribution over a dictionary atom. Probability distributions over other model parameters (e.g. z_{ik}) are also inferred as by-products, however they are not required by our approach. We transform the means of the $\mathcal{GP}$s using the spectral transformation operator $\boldsymbol{\Delta}$ and use them to represent the high resolution multi-spectral image $\mathbf{X}$. To learn the representation, we again use the proposed model. However, during inference, instead of sampling for the dictionary atoms, we keep them fixed to the transformed means of the $\mathcal{GP}$s. For our model, this implies $\eta_k \to \infty$. Therefore, we also do not sample for η_k. These modifications in the sampling process effectively reduce our dictionary learning process to a sparse coding process. We defer further discussion on sparse coding to Sect. 4.3. Here, we are interested in the set $\mathcal{B} \subset \mathbb{R}$ of $K \times MN$ Bernoulli distributions (i.e. parameters $\pi_{ik}, \forall i, k$) computed by the inference process. This set contains K distributions for each pixel of $\mathbf{X}$ that determines the support (indices of non-zero elements) of the sparse codes for that pixel. Since the basis vectors for the sparse codes are the transformed means of the $\mathcal{GP}$s, $\mathcal{B}$ encodes the activity level of the $\mathcal{GP}$s in the sparse representation of $\mathbf{X}$. We store $\mathcal{B}$ to later exploit it in an accurate reconstruction of the high resolution hyperspectral image.

We emphasize that the association of K Bernoulli distributions with a single pixel in our approach is different from the affiliation of K such distributions with the complete image, used by Akhtar et al. [16]. Our approach computes more distributions to promote spatial consistency in the representations.

4.3 Sparse Coding

Let us briefly consider the sparse coding process for a pixel $\mathbf{x}$ of $\mathbf{X}$, discussed in the previous section. That process computes the codes $\boldsymbol{\beta}$ of $\mathbf{x}$, such that their support follows the Bernoulli distributions in $\mathcal{B}$. Let $\mathbf{z} \in \mathbb{R}^K$ be the binary vector indexing that support. It is easy to imagine that sampling the same distributions in $\mathcal{B}$ multiple times, can often result in different $\mathbf{z}$. Being a sample of the *learned* Bernoulli distributions, each such $\mathbf{z}$ is a useful support of the sparse codes for our approach. This entails the existence of multiple useful sparse codes and hence, multiple useful reconstructions of $\mathbf{x}$ in our probabilistic settings.

Let $\widetilde{\boldsymbol{\Psi}}$ denote the dictionary formed by the transformed means of the $\mathcal{GP}$s. We can write, $\widetilde{\mathbf{x}} = \widetilde{\boldsymbol{\Psi}}(\mathbf{w} \odot \mathbf{z})$, where $\boldsymbol{\beta} = \mathbf{w} \odot \mathbf{z}$ and $\widetilde{\mathbf{x}}$ is the reconstructed $\mathbf{x}$. We propose the following lemma regarding $\widetilde{\mathbf{x}}$:

Lemma 1. $\left\|\mathbb{E}\left[\widetilde{\mathbf{x}}\right] - \mathbf{x}\right\|_2^2 \leq \left\|\widetilde{\mathbf{x}} - \mathbf{x}\right\|_2^2$, *where* $\mathbb{E}[.]$ *is the expectation operator.*

Proof: We can write $\left\|\mathbb{E}\left[\widetilde{\mathbf{x}}\right] - \mathbf{x}\right\|_2^2 = \left\|\widetilde{\boldsymbol{\Psi}}\mathbb{E}\left[\boldsymbol{\beta}\right] - \mathbf{x}\right\|_2^2$. Since $\boldsymbol{\beta} = \mathbf{w} \odot \mathbf{z}$ and multiple $\mathbf{z}$ exist, we can exploit the conditional expectation of the discrete random variables to write $\mathbb{E}\left[\boldsymbol{\beta}\right] = \mathbb{E}\left[\mathbb{E}[\boldsymbol{\beta}|\mathbf{z}]\right]$ [49]. From the results in [16], we already know that $\mathbb{E}\left[\mathbb{E}[\boldsymbol{\beta}|\mathbf{z}]\right] = \boldsymbol{\beta}_{opt}$, where $\boldsymbol{\beta}_{opt}$ is the optimal $\boldsymbol{\beta}$ with respect to the squared error. Since, $\mathbb{E}\left[\widetilde{\mathbf{x}}\right] = \widetilde{\boldsymbol{\Psi}}\boldsymbol{\beta}_{opt}$ and $\widetilde{\mathbf{x}} = \widetilde{\boldsymbol{\Psi}}\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is not guaranteed to be optimal, $\left\|\mathbb{E}\left[\widetilde{\mathbf{x}}\right] - \mathbf{x}\right\|_2^2 \leq \left\|\widetilde{\mathbf{x}} - \mathbf{x}\right\|_2^2$.

Lemma 1 shows that the expected value of multiple reconstructions of $\mathbf{x}$ can be superior to its single reconstruction. Moreover, by using $\widetilde{\boldsymbol{\Psi}} = \boldsymbol{\Delta}\boldsymbol{\Psi}$ in the above proof, we can extend this result to show that $\left\|\mathbb{E}[\widetilde{\mathbf{h}}] - \mathbf{h}\right\|_2^2 \leq \left\|\widetilde{\mathbf{h}} - \mathbf{h}\right\|_2^2$, where $\mathbf{h}$ and $\widetilde{\mathbf{h}}$ denote the pixels of the target high resolution hyperspectral image and its reconstruction, respectively. To exploit this finding, we adopted the following strategy for the reconstruction of the super resolution image $\mathbf{H}$. First, we compute Q sparse codes for $\mathbf{X}$ using our model. For these computations, we fixed both $\widetilde{\boldsymbol{\Psi}}$ and $\mathcal{B}$ during the Bayesian inference. Since the sparse codes in Sect. 4.2 were also computed using the same $\mathcal{B}$ and $\widetilde{\boldsymbol{\Psi}}$, we also use them in the upcoming computations. Second, we draw $Q+1$ samples from the already inferred Gaussian Processes and use them with the available $Q+1$ sparse codes to construct the same number of reconstructions of $\mathbf{H}$. Finally, we estimate the expected value of these $Q+1$ reconstructions by computing their mean. Note that, the reconstructions are performed pixel-wise in our approach, where pixels represent natural spectra. In light of Lemma 1 and the above discussion, the mean reconstructed pixels are smoothened spectra with minimum squared error with respect to the ground-truth pixels.

5 Experiments

We have evaluated our approach on both remote sensing and ground-based hyperspectral images. For the evaluation metrics, we used the Root Mean Squared

Error (RMSE) [16] and the Spectral Angle Mapper (Sam) [50]. We mainly compare our approach to the Matrix Factorization based approach (MF) [4], the Spatial-Spectral Fusion Method (SSFM) [33], the Generalized Simultaneous OMP based method (GSOMP) [17] and the Bayesian Sparse Representation approach (BSR) [16]. These are the state-of-the-art approaches in the first category of the hyperspectral super-resolution techniques (see Sect. 2) to which our approach also belongs. We note that few approaches from the second and the third category have recently reported impressive results [29,37,40]. However, those results are obtained by exploiting additional prior knowledge, which is not always available (and hence, not assumed in this work).

In our experiments, we used the author-provided implementations for GSOMP and BSR, with the parameter values reported for the same data sets in [16,17]. Due to the unavailability of public codes for MF and SSFM, we implemented these approaches using the SPAMS library [51], that is well-know for its accuracy. The parameter values of these approaches were carefully optimized such that the achieved results were the same or better than the previously reported best results for these approaches on common images. We defer the discussion on the parameter settings of our approach to Sect. 6. We follow a common evaluation protocol [4,16,17] that considers an available hyperspectral image as the ground truth and constructs a low resolution hyperspectral image by averaging 32×32 disjoint blocks of the ground truth. A high resolution multi-spectral image is constructed by spectral transformation of the ground truth, with a known transformation operator $\boldsymbol{\Delta}$.

Table 1. Benchmarking on remote sensing images: The results are in the range of 8 bit images. The best results are given in bold in green cells. The second best values are in blue cells. Image names are according to the source data set.

	AVIRIS data set							
Image	SC01		SC02		SC03		SC04	
Method	RMSE	Sam	RMSE	Sam	RMSE	Sam	RMSE	Sam
MF [4]	1.32	1.85	1.55	1.60	1.62	1.51	2.73	2.49
SSFM [33]	1.35	1.68	1.56	1.59	1.77	1.59	2.68	2.31
GSOMP [17]	1.30	1.39	1.52	1.63	1.79	1.80	1.54	2.05
BSR [16]	1.21	1.33	1.54	1.61	1.46	1.58	1.62	1.77
Proposed	**0.92**	**1.26**	**1.32**	**1.47**	**1.11**	**1.35**	**1.36**	**1.54**

For the remote sensing images, we used a data set provided by NASA[1], that contains four hyperspectral images collected by the airborne sensor AVIRIS [52]. These $512 \times 512 \times 224$ images are acquired in the wavelength range 370–2500 nm, over Cuprite mines in Nevada, US. Due to water absorptions and low signal-to-noise ratio, we removed 36 channels from the images, corresponding to the

[1] Download link: ftp://popo.jpl.nasa.gov/pub/free_data/f970619t01p02r02c_rfl.tar.

wavelengths 370, 380, 1330 to 1430, 1780 to 1970, 2490 and 2500 nm. The resulting images are considered as the ground truth. We constructed a high resolution multi-spectral image by selecting six bands of the ground truth, corresponding to the wavelengths 480, 560, 660, 830, 1650 and 2220 nm. These bands roughly correspond to the visible and mid-infrared wavelength channels of NASA-Landsat 7 satellite. Thus, the spectral transform $\boldsymbol{\Delta} \in \mathbb{R}^{6\times 188}$ is a binary matrix in this experiment, similar to [16,17,33].

Comparison of performance on the remote sensing images is summarized in Table 1. The results are computed using the 8-bit intensity range. As visible in the table, a considerable improvement in the results is achieved by our approach. On average, the RMSE values for our approach are ∼18 % better than the previous lowest values. Similarly, the average gain in the spectral angle mapper values is ∼9 %. This gain can be attributed to the spectral smoothness and the spatial consistency of the images reconstructed by our approach, which are common attributes of remote sensing hyperspectral images. In Fig. 2, we compare the reconstructions of two randomly selected contiguous pixels of image SC01 by BSR [16] and our approach. BSR showed the second best results on this image. Our approach not only reconstructs each pixel better due to the spectral smoothness, but also due to the similarities between the adjacent pixels. The angles (in $\mathbb{R}^{188}$) between the shown ground truth pixels is 1.26°. For our approach, this angle is 1.28°, whereas for BSR, its value is 3.23°. We also show examples of the reconstructed spectral images for SC01 by our approach and BSR in Fig. 3.

For the ground-based images, we evaluated our approach on hyperspectral images of everyday objects from the CAVE database [53] and the images of indoor and outdoor scenes from the Harvard database [54]. The 512 × 512 × 31 images of the CAVE database are acquired using tunable filters over the

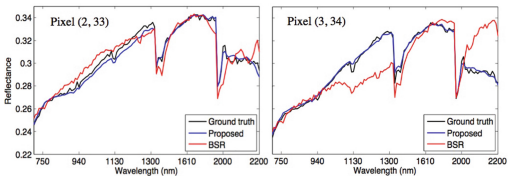

Fig. 2. Effect of spectral smoothness and spatial consistency: Two contiguous ground truth pixels for SC01 are shown along their estimates with BSR [16] and our approach.

wavelength range 400–700 nm. The fixed focal length of the sensor has resulted in a blur for the first two spectral bands of the images. We removed these bands in our experiments to avoid any bias in the results. The remaining images are considered as the ground truth. The use of tunable filters in the Harvard database has resulted in spatial distortions in some images with moving objects (e.g. grass, trees). We also avoid these images in our experiments for a fair evaluation. Following [16,17] we used the top-left 1024 × 1024 spatial patches as the ground truth for this database.

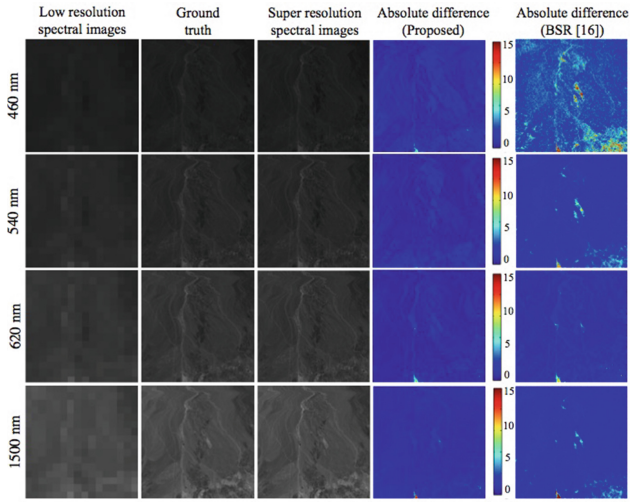

Fig. 3. Spectral images for SC01 at 460, 540, 620 and 1500 nm: The 512×512 reconstructions of the ground truth are shown along the used 16×16 low resolution images. Absolute differences between the reconstructions and the ground truth are also shown. The absolute difference images for BSR [16] are included for comparison.

Table 2. Benchmarking on ground-based images: CMF [40] and TDGU [29] (in red) are included for *reference only* as these approaches belong to different categories.

	CAVE database [53]									
Image	Balloons		Beads		Cloth		Pompos		CD	
Method	RMSE	Sam	RMSE	Sam	RMSE	Sam	RMSE	Sam	RMSE	Sam
MF [4] (Cat. 1)	2.3	8.0	8.2	14.9	6.0	7.9	4.3	10.7	7.9	14.9
SSFM [33] (Cat. 1)	2.4	8.4	8.9	15.3	7.6	8.2	4.3	11.9	8.1	16.4
GSOMP [17] (Cat. 1)	2.3	8.1	6.3	14.1	4.2	5.2	4.4	**10.0**	7.5	18.7
BSR [16] (Cat. 1)	2.1	7.9	5.9	14.2	4.0	5.9	4.1	11.1	5.4	12.9
Proposed (Cat. 1)	**1.9**	**7.6**	**5.8**	**13.7**	**3.7**	**5.0**	**3.9**	10.1	**5.3**	**10.6**
CMF [40] (Cat. 2)	2.9	4.3	7.2	7.5	5.2	4.5	3.5	3.6	6.1	7.0
TDGU [29] (Cat. 3)	1.6	-	6.9	-	-	-	-	-	3.5	-

	Harvard database [54]									
Image	Img h0		Img c2		Img d3		Img b5		Img b2	
Method	RMSE	Sam	RMSE	Sam	RMSE	Sam	RMSE	Sam	RMSE	Sam
MF [4] (Cat. 1)	2.6	2.7	2.9	2.6	1.8	3.3	2.4	2.5	2.1	3.0
SSFM [33] (Cat. 1)	3.1	2.8	3.2	2.8	2.1	3.6	2.3	2.9	2.3	3.1
GSOMP [17] (Cat. 1)	3.3	2.9	2.8	**1.9**	1.7	3.2	0.9	2.2	1.6	2.7
BSR [16] (Cat. 1)	2.4	2.9	2.6	2.2	**1.3**	3.2	0.9	2.2	**1.1**	2.5
Proposed (Cat. 1)	**2.2**	**2.5**	**2.4**	**1.9**	1.4	**3.0**	**0.8**	**2.1**	**1.1**	**2.3**
CMF [40] (Cat. 2)	2.3	2.4	2.4	2.0	1.4	3.0	1.6	2.1	1.7	2.1
TDGU [29] (Cat. 3)	-	-	-	-	-	-	0.7	-	-	-

Following [4,16,17] we constructed the high resolution multi-spectral images by transforming the ground truth with the spectral response of Nikon D700 camera (http://www.maxmax.com/spectral_response.htm). In Table 2, we show the results on five commonly used benchmarking images of each database. For reference, we also include results of representative methods from the remaining two categories of the matrix factorization based hyperspectral super-resolution techniques. The results of Coupled Matrix Factorization (CMF) [40] based approach are directly taken from [37], whereas the performance of the Training Data Guided Up-sampling (TDGU) method are taken from [29]. In Fig. 4, we show examples of the super resolution spectral images. Although the performance of the proposed approach is generally better than the existing approaches in the same category, the improvements in the results are not as significant as for the remote sensing images. In our opinion, the lower spectral resolution and larger variations in the spatial patterns in the ground-based images are the reasons behind this phenomenon. Nevertheless, our approach is generally able to perform better than the existing approaches on the ground-based images as well.

6 Discussion on Parameters

In all the experiments we used the value 10^{-6} for $a_o, b_o, c_o, d_o, k_o, l_o$ and 1 for e_o and f_o. The values of g_o and h_o were adjusted to give a parabolic probability density function of the Beta distributions, for which $g_o/K \approx h_o(K-1)/K$. We used $\eta_o = 1/L$ and the spatial kernel width was set to 2 for the Harvard database and 1 for the remaining data sets. This resulted in $P = 5$ and $P = 3$ respectively. Except for P, our model is fairly insensitive to small perturbations in the parameter values, which is a common observation for Bayesian models [16]. The reported results are sensitive to the value of P because the considered low

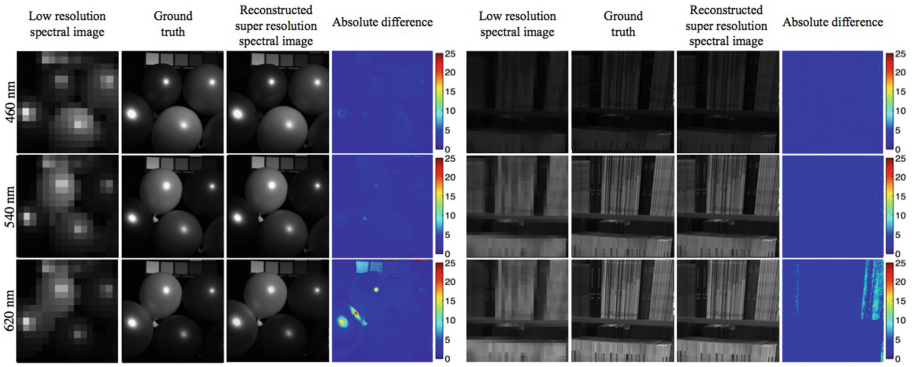

Fig. 4. Spectral reconstruction at 460, 540 and 620 nm: The used 16×16 low resolution images are shown along the reconstructions and the 512×512 ground truth. Absolute differences between the reconstructions and the ground truth are also given. (Left) 'Balloons' from the CAVE database [53]. (Right) 'Img h0' form the Harvard database [54].

resolution hyperspectral images have very small spatial dimensions. Keeping in view the spatial dimensions of the images, we used $K = 100$ for the Harvard database and 10 for the remaining data sets.

We initialized λ_ϵ to 10^6, λ_w to 10^3, η_k to $10^{-3}, \forall k$ and π_{ik} to $10^{-3}, \forall i, k$. These initial values were selected considering the physical significance of the parameters in our model. Nevertheless, the approach is generally insensitive to these initial values. To learn the Gaussian Processes, we initialized the dictionary with random samples of Multi-variate Gaussians and initialized the spare codes by allowing half of them to have value 1. For sparse coding, we used the LASSO solver of the SPAMS library [51] to initialize the sparse codes. In our experiments, we processed the images as 2×2 overlapping patches.

We used 500 sampling iterations for dictionary learning and 300 and 100 iterations respectively for learning the Bernoulli distributions in $\mathcal{B}$ and sparse coding. Fewer iterations were enough in the later stages because fewer probability distributions were required to be sampled in those stages. We computed the codes 25 time, i.e. $Q = 25$ in our experiments.

7 Conclusion

We proposed a Bayesian approach for hyperspectral super resolution that fuses a high resolution multi-spectral image with a hyperspectral image. It utilizes a Bayesian sparse representation model that places Gaussian Process priors on the dictionary and uses a kernel to promote spatial consistency in the representation. We also derived inference equations for the proposed model. The model is used for inferring Gaussian Processes for the dictionary atoms, estimating their popularity in the representation of the multi-spectral image and computing multiple sparse codes of that image. The sparse codes of the multi-spectral image are used with the samples of the Gaussian Processes to finally estimate the super resolution image. We tested our approach on remote sensing and ground-based images. Our results show that the approach is useful for both types of images.

Acknowledgment. Supported by ARC Discovery Grant DP110102399.

References

1. Edelman, G., Gaston, E., Van Leeuwen, T., Cullen, P., Aalders, M.: Hyperspectral imaging for non-contact analysis of forensic traces. Forensic Sci. Intl. **223**(1), 28–39 (2012)
2. Lu, G., Fei, B.: Medical hyperspectral imaging: a review. J. Biomed. Opt. **19**(1), 010901 (2014)
3. Zhou, Y., Chang, H., Barner, K., Spellman, P., Parvin, B.: Classification of histology sections via multispectral convolutional sparse coding. In: CVPR, June 2014
4. Kawakami, R., Wright, J., Tai, Y.W., Matsushita, Y., Ben-Ezra, M., Ikeuchi, K.: High-resolution hyperspectral imaging via matrix factorization. In: CVPR, pp. 2329–2336, June 2011

5. Dias, J.B., Plaza, A., Valls, G.C., Scheunders, P., Nasrabadi, N., Chanussot, J.: Hyperspectral remote sensing data analysis and future challenges. IEEE Geosci. Remote Sens. Mag. **1**(2), 6–36 (2013)
6. Charles, A., Olshausen, B., Rozell, C.: Learning sparse codes for hyperspectral imagery. IEEE J. Sel. Top. Signal Process. **5**(5), 963–978 (2011)
7. Zhang, D., Zuo, W., Yue, F.: A comparative study of palmprint recognition algorithms. ACM Comput. Surv. **44**(1), 2:1–2:37 (2012)
8. Uzair, M., Mahmood, A., Mian, A.: Hyperspectral face recognition using 3D-DCT and partial least squares. In: Proceedings British Machine Vision Conference 2013 (BMVC), pp. 57.1–57.10 (2013)
9. Di, W., Zhang, L., Zhang, D., Pan, Q.: Studies on hyperspectral face recognition in visible spectrum with feature band selection. IEEE Trans. Syst. Man Cybern. Part A: Syst. Hum. **40**(6), 1354–1361 (2010)
10. Khan, Z., Shafait, F., Mian, A.: Automatic ink mismatch detection for forensic document analysis. Pattern Recogn. **48**(11), 3615–3626 (2015)
11. Kim, S.J., Deng, F., Brown, M.S.: Visual enhancement of old documents with hyperspectral imaging. Pattern Recogn. **44**(7), 1461–1469 (2011)
12. Nguyen, H.V., Banerjee, A., Chellappa, R.: Tracking via object reflectance using a hyperspectral video camera. In: CVPRW, pp. 44–51, June 2010
13. Hwang, S., Park, J., Kim, N., Choi, Y., So Kweon, I.: Multispectral pedestrian detection: benchmark dataset and baseline. In: IEEE Conference on Computer Vision and Pattern Recognition, June 2015
14. Tarabalka, Y., Chanussot, J., Benediktsson, J.A.: Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers. IEEE Trans. Syst., Man, Cybern., Syst. **40**(5), 1267–1279 (2010)
15. Chen, C., Li, Y., Liu, W., Huang, J.: Image fusion with local spectral consistency and dynamic gradient sparsity. In: CVPR, June 2014
16. Akhtar, N., Shafait, F., Mian, A.: Bayesian sparse representation for hyperspectral image super resolution. In: IEEE Conference on Computer Vision and Pattern Recognition, June 2015
17. Akhtar, N., Shafait, F., Mian, A.: Sparse Spatio-Spectral Representation for Hyperspectral Image Super-Resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8695, pp. 63–78. Springer, Heidelberg (2014). doi:10.1007/978-3-319-10584-0_5
18. Rasmussen, C.E.: Gaussian processes for machine learning. Citeseer (2006)
19. Dobigeon, N., Tourneret, J.Y., Richard, C., Bermudez, J., Mclaughlin, S., Hero, A.O.: Nonlinear unmixing of hyperspectral images: models and algorithms. IEEE Signal Process. Mag. **31**(1), 82–94 (2014)
20. Thibaux, R., Jordan, M.I.: Hierarchical beta processes and the Indian buffet process. In: International Conference on Artificial Intelligence and Statistics, pp. 564–571 (2007)
21. Bishop, C.M.: Pattern Recognition and Machine Learning (Information Science and Statistics). Springer, Secaucus (2006)
22. Solomon, J., Rock, B.: Imaging spectrometry for earth remote sensing. Science **228**(4704), 1147–1152 (1985)
23. Alparone, L., Wald, L., Chanussot, J., Thomas, C., Gamba, P., Bruce, L.: Comparison of pansharpening algorithms: outcome of the 2006 GRS-S data-fusion contest. IEEE Trans. Geosci. Remote Sens. **45**(10), 3012–3021 (2007)
24. Nunez, J., Otazu, X., Fors, O., Prades, A., Pala, V., Arbiol, R.: Multiresolution-based image fusion with additive wavelet decomposition. IEEE Trans. Geosci. Remote Sens. **37**(3), 1204–1211 (1999)

25. Haydn, R., Dalke, G.W., Henkel, J., Bare, J.E.: Application of the IHS color transform to the processing of multisensor data and image enhancement. In: Proceedings of the International Symposium on Remote Sensing of Environment (1982)
26. Carper, W.J., Lilles, T.M., Kiefer, R.W.: The use of intensity-hue-saturation transformations for merging SOPT panchromatic and multispectral image data. Photogram. Eng. Remote Sens. **56**(4), 457–467 (1990)
27. Shah, V.P., Younan, N.H., King, R.L.: An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets. IEEE Trans. Geosci. Remote Sens. **46**(5), 1323–1335 (2008)
28. Cetin, M., Musaoglu, N.: Merging hyperspectral and panchromatic image data: qualitative and quantitative analysis. Int. J. Remote Sens. **30**(7), 1779–1804 (2009)
29. Kwon, H., Tai, Y.W.: RGB-guided hyperspectral image upsampling. In: International Conference on Computer Vision (2015)
30. Zhukov, B., Oertel, D., Lanzl, F., Reinhackel, G.: Unmixing-based multisensor multiresolution image fusion. IEEE Trans. Geosci. Remote Sens. **37**(3), 1212–1226 (1999)
31. Minghelli-Roman, A., Polidori, L., Mathieu-Blanc, S., Loubersac, L., Cauneau, F.: Spatial resolution improvement by merging MERIS-ETM images for coastal water monitoring. IEEE Geosci. Remote Sens. Lett. **3**(2), 227–231 (2006)
32. Keshava, N., Mustard, J.: Spectral unmixing. IEEE Signal Process. Mag. **19**(1), 44–57 (2002)
33. Huang, B., Song, H., Cui, H., Peng, J., Xu, Z.: Spatial and spectral image fusion using sparse matrix factorization. IEEE Trans. Geosci. Remote Sens. **52**(3), 1693–1704 (2014)
34. Zurita-Milla, R., Clevers, J.G., Schaepman, M.E.: Unmixing-based Landsat TM and MERIS FR data fusion. IEEE Trans. Geosci. Remote Sens. **5**(3), 453–457 (2008)
35. He, L., Qi, H., Zaretzki, R.: Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 345–352, June 2013
36. Lanaras, C., Baltsavias, E., Schindler, K.: Advances in hyperspectral and multispectral image fusion and spectral unmixing. ISPRS – Intl. Arch. Photogrammetry, Remote Sens. Spat. Inf. Sci. **3**, W3 (2015)
37. Lanaras, C.A., Baltsavias, E., Schindler, K.: Hyperspectral super-resolution by coupled spectral unmixing. In: International Conference on Computer Vision (2015)
38. Wycoff, E., Chan, T.H., Jia, K., Ma, W.K., Ma, Y.: A non-negative sparse promoting algorithm for high resolution hyperspectral imaging. In: ICASSP, pp. 1409–1413, May 2013
39. Wei, Q., Bioucas-Dias, J., Dobigeon, N., Tourneret, J.Y.: Hyperspectral and multispectral image fusion based on a sparse representation. IEEE Trans. Geosci. Remote Sens. **53**(7), 3658–3668 (2015)
40. Yokoya, N., Yairi, T., Iwasaki, A.: Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. IEEE Trans. Geosci. Remote Sens. **50**(2), 528–537 (2012)
41. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**(1), 1–122 (2011)
42. Rubinstein, R., Bruckstein, A., Elad, M.: Dictionaries for sparse representation modeling. Proc. IEEE **98**(6), 1045–1057 (2010)

43. Zhou, M., Chen, H., Paisley, J., Ren, L., Li, L., Xing, Z., Dunson, D., Sapiro, G., Carin, L.: Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images. IEEE Trans. Image Process. **21**(1), 130–144 (2012)
44. Roweis, S., Ghahramani, Z.: A unifying review of linear Gaussian models. Neural Comput. **11**(2), 305–345 (1999)
45. Hastings, W.K.: Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**(1), 97–109 (1970)
46. Zhou, M., Yang, H., Sapiro, G., Dunson, D.B., Carin, L.: Dependent hierarchical beta process for image interpolation and denoising. In: International Conference on Artificial Intelligence and Statistics, pp. 883–891 (2011)
47. Damien, P., Wakefield, J., Walker, S.: Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. J. Roy. Stat. Soc. Ser. B Stat. Methodol. **61**, 331–344 (1999)
48. Havil, J.: Gamma: Exploring Euler's Constant. The Mathematical Intelligence **27**(1), 86–88 (2005)
49. Ross, S.M.: Introduction to Probability Models. Academic Press, San Diego (2014)
50. Yuhas, R.H., Goetz, A.F., Boardman, J.W.: Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In: Summaries of the Third Annual JPL Airborne Geoscience Workshop, vol. 1, pp. 147–149. JPL Publication, Pasadena (1992)
51. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. J. Mach. Learning Res. **11**, 19–60 (2010)
52. Green, R.O., Eastwood, M.L., Sarture, C.M., Chrien, T.G., Aronsson, M., Chippendale, B.J., Faust, J.A., Pavri, B.E., Chovit, C.J., Solis, M., Olah, M.R., Williams, O.: Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). Remote Sens. Environ. **65**(3), 227–248 (1998)
53. Yasuma, F., Mitsunaga, T., Iso, D., Nayar, S.: Generalized assorted pixel camera: post-capture control of resolution, dynamic range and spectrum. Technical report, Dept. of Comp. Sci., Columbia University, CUCS-061-08, November 2008
54. Chakrabarti, A., Zickler, T.: Statistics of real-world hyperspectral images. In: CVPR, pp. 193–200, June 2011