Bayesian Sparse Representation for Hyperspectral Image Super Resolution

Naveed Akhtar, Faisal Shafait and Ajmal Mian
The University of Western Australia
35 Stirling Highway, Crawley, 6009. WA
naveed.akhtar@research.uwa.edu.au, {faisal.shafait & ajmal.mian} @ uwa.edu.au

Abstract

Despite the proven efficacy of hyperspectral imaging in many computer vision tasks, its widespread use is hindered by its low spatial resolution, resulting from hardware limitations. We propose a hyperspectral image super resolution approach that fuses a high resolution image with the low resolution hyperspectral image using non-parametric Bayesian sparse representation. The proposed approach first infers probability distributions for the material spectra in the scene and their proportions. The distributions are then used to compute sparse codes of the high resolution image. To that end, we propose a generic Bayesian sparse coding strategy to be used with Bayesian dictionaries learned with the Beta process. We theoretically analyze the proposed strategy for its accurate performance. The computed codes are used with the estimated scene spectra to construct the super resolution hyperspectral image. Exhaustive experiments on two public databases of ground based hyperspectral images and a remotely sensed image show that the proposed approach outperforms the existing state of the art.

1. Introduction

Spectral characteristics of hyperspectral imaging have recently been reported to enhance performance in many computer vision tasks, including tracking [22], recognition and classification [14], [32], [28], segmentation [25] and document analysis [20]. They have also played a vital role in medical imaging [34], [18] and remote sensing [13], [4]. Hyperspectral imaging acquires a faithful spectral representation of the scene by integrating its radiance against several spectrally well-localized basis functions. However, contemporary hyperspectral systems lack in spatial resolution [2], [18], [11]. This fact is impeding their widespread use. In this regard, a simple solution of using high resolution sensors is not viable as it further reduces the density of the photons reaching the sensors, which is already limited by the high spectral resolution of the instruments.

Figure 1. Left: A 16×16 spectral image at 600nm. Center: The 512×512 super resolution spectral image constructed by the proposed approach. Right: Ground truth (CAVE database [30]).

Due to hardware limitations, software based approaches for hyperspectral image super resolution (*e.g*. see Fig. 1) are considered highly attractive [2]. At present, the spatial resolution of the systems acquiring images by a gross quantization of the scene radiance (*e.g*. RGB and RGB-NIR) is much higher than their hyperspectral counterparts. In this work, we propose to fuse the spatial information from the images acquired by these systems with the hyperspectral images of the same scenes using non-parametric Bayesian sparse representation.

The proposed approach fuses a hyperspectral image with the high resolution image in a four-stage process, as shown in Fig. 2. In the first stage, it infers probability distributions for the material reflectance spectra in the scene and a set of Bernoulli distributions, indicating their proportions in the image. Then, it estimates a dictionary and transforms it according to the spectral quantization of the high resolution image. In the third stage, the transformed dictionary and the Bernoulli distributions are used to compute the sparse codes of the high resolution image. To that end, we propose a generic Bayesian sparse coding strategy to be used with Bayesian dictionaries learned with the Beta process [23]. We theoretically analyze the proposed strategy for its accurate performance. Finally, the computed codes are used with the estimated dictionary to construct the super resolution hyperspectral image. The proposed approach not only improves the state of the art results, which is verified by exhaustive experiments on three different public data sets, it also maintains the advantages of the non-parametric Bayesian framework over the typical optimization based approaches [2], [18], [29], [31].

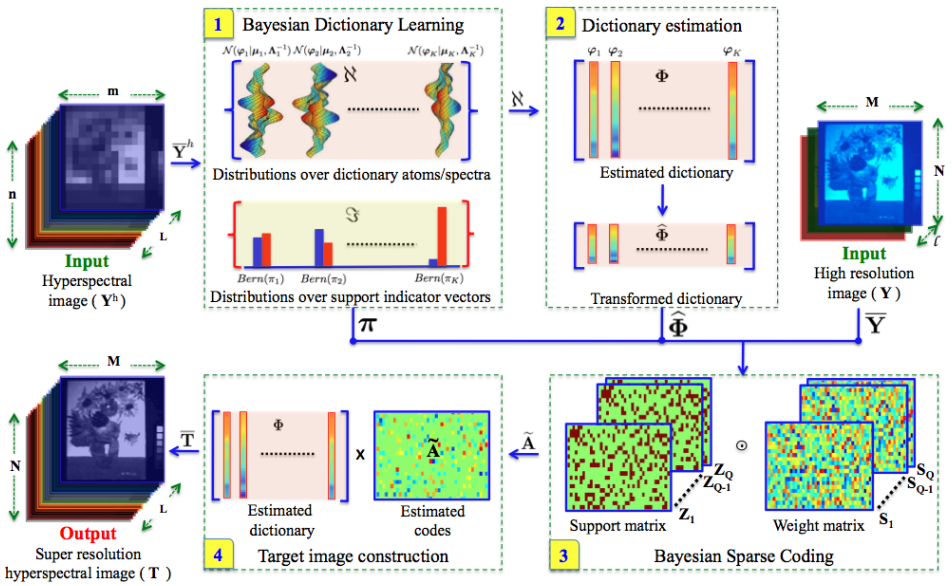

Figure 2. Schematics of the proposed approach: (1) Sets of distributions over the dictionary atoms and the support indicator vectors are inferred non-parametrically. (2) A dictionary $\mathbf{\Phi}$ is estimated and transformed according to the spectral quantization of the high resolution image $\mathbf{Y}$. (3) The transformed dictionary and the distributions over the support indicator vectors are used for sparse coding $\mathbf{Y}$. This step is performed by the proposed Bayesian sparse coding strategy. (4) The codes are used with $\mathbf{\Phi}$ to construct the target super resolution image.

The rest of the paper is organized as follows. After reviewing the related literature in Section 2, we formalize the problem in Section 3. The proposed approach is presented in Section 4 and evaluated in Section 5. Section 6 provides a discussion on the parameter settings of the proposed approach, and Section 7 concludes the paper.

2. Related Work

Hyperspectral sensors have been in use for nearly two decades in remote sensing [13]. However, it is still difficult to obtain high resolution hyperspectral images by the satellite sensors due to technical and budget constraints [17]. This fact has motivated considerable research in hyperspectral image super resolution, especially for remote sensing. To enhance the spatial resolution, hyperspectral images are usually fused with the high resolution pan-chromatic images (*i.e*. pan-sharpening) [25], [11]. In this regard, conventional approaches are generally based on projection and substitution, including the intensity hue saturation [16] and the principle component analysis [10]. In [1] and [7] , the authors have exploited the sensitivity of human vision to luminance and fused the luminance component of the high resolution images with the hyperspectral images. However, this approach can also cause spectral distortions in the resulting image [8].

Minghelli-Roman *et al*. [21] and Zhukov *et al*. [35] have used hyperspectral unmixing [19], [3] for spatial resolution enhancement of hyperspectral images. However, their methods require that the spectral resolutions of the images being fused are close to each other. Furthermore, these approaches struggle in highly mixed scenarios [17]. Zurita-Milla *et al*. [36] have enhanced their performance for such cases using the sliding window strategy.

More recently, matrix factorization based hyperspectral image super resolution for ground based and remote sensing imagery has been actively investigated [18], [29], [17], [31], [2]. Approaches developed under this framework fuse high resolution RGB images with hyperspectral images. Kawakami *et al*. [18] represented each image from the two modalities by two factors and constructed the desired image with the complementary factors of the two representations. Similar approach is applied in [17] to the remotely acquired images, where the authors used a down-sampled version of the RGB image in the fusion process. Wycoff *et al*. [29] developed a method based on Alternating Direction Method of Multipliers (ADMM) [6]. Their approach also requires prior knowledge about the spatial transform between the images being fused. Akhtar *et al*. [2] proposed a method based on sparse spatio-spectral representation of hyperspectral images that also incorporates the non-negativity of the spectral signals. The strength of the approach comes from exploiting the spatial structure in the scene, which requires processing the images in terms of spatial patches and solving a simultaneous sparse optimization problem [27]. Yokoya *et al*. [31] made use of coupled feature space between a hyperspectral and a multispectral image of the same scene.

Matrix factorization based approaches have been able to show state of the art results in hyperspectral image su-

per resolution using the image fusion technique. However, Akhtar *et al*. [2] showed that their performance is sensitive to the algorithm parameters, especially to the sizes of the matrices (*e.g*. dictionary) into which the images are factored. Furthermore, there is no principled way to incorporate prior domain knowledge to enhance the performance of these approaches.

3. Problem Formulation

Let $\mathbf{Y}^h \in \mathbb{R}^{m\times n\times L}$ be the acquired low resolution hyperspectral image, where L denotes the spectral dimension. We assume availability of a high resolution image $\mathbf{Y} \in \mathbb{R}^{M\times N\times l}$ (*e.g*. RGB) of the same scene, such that $M \gg m, N \gg n$ and $L \gg l$. Our objective is to estimate the super resolution hyperspectral image $\mathbf{T} \in \mathbb{R}^{M\times N\times L}$ by fusing $\mathbf{Y}$ and $\mathbf{Y}^h$. For our problem, $\mathbf{Y}^h = \mathbf{\Psi}^h(\mathbf{T})$ and $\mathbf{Y} = \mathbf{\Psi}(\mathbf{T})$, where $\mathbf{\Psi}^h : \mathbb{R}^{M\times N\times L} \rightarrow \mathbb{R}^{m\times n\times L}$ and $\mathbf{\Psi} : \mathbb{R}^{M\times N\times L} \rightarrow \mathbb{R}^{M\times N\times l}$.

Let $\mathbf{\Phi} \in \mathbb{R}^{L\times|\mathcal{K}|}$ be an unknown matrix with columns $\boldsymbol{\varphi}_k$, where $k \in \mathcal{K} = \{1, ..., K\}$ and $|.|$ denotes the cardinality of the set. Let $\overline{\mathbf{Y}}^h = \mathbf{\Phi}\mathbf{B}$, where the matrix $\overline{\mathbf{Y}}^h \in \mathbb{R}^{L\times mn}$ is created by arranging the pixels of $\mathbf{Y}^h$ as its columns and $\mathbf{B} \in \mathbb{R}^{|\mathcal{K}|\times mn}$ is a coefficient matrix. For our problem, the basis vectors $\boldsymbol{\varphi}_k$ represent the reflectance spectra of different materials in the imaged scene. Thus, we also allow for the possibility that $|\mathcal{K}| > L$. Normally, $|\mathcal{K}| \ll mn$ because a scene generally comprises only a few spectrally distinct materials [2]. Let $\widehat{\mathbf{\Phi}} \in \mathbb{R}^{l\times|\mathcal{K}|}$ be such that $\overline{\mathbf{Y}} = \widehat{\mathbf{\Phi}}\mathbf{A}$, where $\overline{\mathbf{Y}} \in \mathbb{R}^{l\times MN}$ is formed by arranging the pixels of $\mathbf{Y}$ and $\mathbf{A} \in \mathbb{R}^{|\mathcal{K}|\times MN}$ is a coefficient matrix. The columns of $\widehat{\mathbf{\Phi}}$ are also indexed in $\mathcal{K}$. Since $\overline{\mathbf{Y}}^h$ and $\overline{\mathbf{Y}}$ represent the images of the same scene, $\widehat{\mathbf{\Phi}} = \mathbf{\Upsilon}\mathbf{\Phi}$, where $\mathbf{\Upsilon} \in \mathbb{R}^{l\times L}$ is a transformation matrix, associating the spectral quantizations of the two imaging modalities. Similar to the previous works [2], [18], [29], this transform is considered to be known *a priori*.

In the above formulation, pixels of $\mathbf{Y}$ and $\mathbf{Y}^h$ are likely to admit sparse representations over $\widehat{\mathbf{\Phi}}$ and $\mathbf{\Phi}$, respectively, because a pixel generally contains very few spectra as compared to the whole image. Furthermore, the value of $|\mathcal{K}|$ can vary greatly between different scenes, depending on the number of spectrally distinct materials present in a scene. In the following, we refer to $\mathbf{\Phi}$ as the *dictionary* and $\widehat{\mathbf{\Phi}}$ as the *transformed dictionary*. The columns of the dictionaries are called their *atoms* and a complementary coefficient matrix (*e.g*. $\mathbf{A}$) is referred as the *sparse code matrix* or the *sparse codes* of the corresponding image. We adopt these conventions from the sparse representation literature [24].

4. Proposed Approach

We propose a four-stage approach for hyperspectral image super resolution that is illustrated in Fig. 2. The proposed approach first separates the scene spectra by learning a dictionary from the low resolution hyperspectral image under a Bayesian framework. The dictionary is transformed using the known spectral transform $\mathbf{\Upsilon}$ between the two input images as $\widehat{\mathbf{\Phi}} = \mathbf{\Upsilon}\mathbf{\Phi}$. The transformed dictionary is used for encoding the high-resolution image. The codes $\widetilde{\mathbf{A}} \in \mathbb{R}^{|\mathcal{K}|\times MN}$ are computed using the proposed strategy. As shown in the figure, we eventually use the dictionary and the codes to construct $\overline{\mathbf{T}} = \mathbf{\Phi}\widetilde{\mathbf{A}}$, where $\overline{\mathbf{T}} \in \mathbb{R}^{L\times MN}$ is formed by arranging the pixels of the target image $\mathbf{T}$. Hence, accurate estimation of $\widetilde{\mathbf{A}}$ and $\mathbf{\Phi}$ is crucial for our approach, where the dictionary estimation also includes finding its correct size, *i.e*. $|\mathcal{K}|$. Furthermore, we wish to incorporate the ability of using the prior domain knowledge in our approach. This naturally leads towards exploiting the non-parametric Bayesian framework. The proposed approach is explained below, following the sequence in Fig. 2.

4.1. Bayesian Dictionary Learning

We denote the i^{th} pixel of $\mathbf{Y}^h$ by $\mathbf{y}_i^h \in \mathbb{R}^L$, that admits to a sparse representation $\boldsymbol{\beta}_i^h \in \mathbb{R}^{|\mathcal{K}|}$ over the dictionary $\mathbf{\Phi}$ with a small error $\boldsymbol{\epsilon}_i^h \in \mathbb{R}^L$. Mathematically, $\mathbf{y}_i^h = \mathbf{\Phi}\boldsymbol{\beta}_i^h + \boldsymbol{\epsilon}_i^h$. To learn the dictionary in these settings[1], Zhou *et al*. [33] proposed a beta process [23] based non-parametric Bayesian model, that is shown below in its general form. In the given equations and the following text, we have dropped the superscript 'h' for brevity, as it can be easily deduced from the context.

$$
\begin{aligned}
&\mathbf{y}_i = \mathbf{\Phi}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i && \forall i \in \{1, ..., mn\}\\
&\boldsymbol{\beta}_i = \mathbf{z}_i \odot \mathbf{s}_i\\
&\boldsymbol{\varphi}_k \sim \mathcal{N}(\boldsymbol{\varphi}_k|\boldsymbol{\mu}_{k_o}, \mathbf{\Lambda}_{k_o}^{-1}) && \forall k \in \mathcal{K}\\
&z_{ik} \sim \text{Bern}(z_{ik}|\pi_{k_o})\\
&\pi_k \sim \text{Beta}(\pi_k|a_o/K, b_o(K-1)/K)\\
&s_{ik} \sim \mathcal{N}(s_{ik}|\mu_{s_o}, \lambda_{s_o}^{-1})\\
&\boldsymbol{\epsilon}_i \sim \mathcal{N}(\boldsymbol{\epsilon}_i|\mathbf{0}, \mathbf{\Lambda}_{\epsilon_o}^{-1})
\end{aligned}
$$

In the above model, $\odot$ denotes the Hadamard/element-wise product; $\sim$ denotes a draw (*i.i.d.*) from a distribution; $\mathcal{N}$ refers to a Normal distribution; Bern and Beta represent Bernoulli and Beta distributions, respectively. Furthermore, $\mathbf{z}_i \in \mathbb{R}^{|\mathcal{K}|}$ is a binary vector whose k^{th} component z_{ik} is drawn from a Bernoulli distribution with parameter π_{k_o}. Conjugate Beta prior is placed over π_k, with hyper-parameters a_o and b_o. We have used the subscript 'o' to distinguish the parameters of the prior distributions. We refer to $\mathbf{z}_i$ as the *support indicator vector*, as the value $z_{ik} = 1$ indicates that the k^{th} dictionary atom participates in the expansion of $\mathbf{y}_i$. Also, each component s_{ik} of $\mathbf{s}_i \in \mathbb{R}^{|\mathcal{K}|}$ (the *weight vector*) is drawn from a Normal distribution.

[1]The sparse code matrix $\mathbf{B}$ (with $\boldsymbol{\beta}_{i\in\{1,...,mn\}}^h$ as its columns) is also learned. However, it is not required by our approach.

For tractability, we restrict the precision matrix $\mathbf{\Lambda}_{k_o}$ of the prior distribution over a dictionary atom to $\lambda_{k_o}\mathbf{I}_L$, where $\mathbf{I}_L$ denotes the identity in $\mathbb{R}^{L\times L}$ and $\lambda_{k_o} \in \mathbb{R}$ is a predetermined constant. A zero vector is used for the mean parameter $\boldsymbol{\mu}_{k_o} \in \mathbb{R}^L$, since the distribution is defined over a basis vector. Similarly, we let $\mathbf{\Lambda}_{\epsilon_o} = \lambda_{\epsilon_o}\mathbf{I}_L$ and $\mu_{s_o} = 0$, where $\lambda_{\epsilon_o} \in \mathbb{R}$. These simplifications allow for fast inferencing in our application without any noticeable degradation of the results. We further place non-informative gamma hyper-priors over λ_{s_o} and λ_{ϵ_o}, so that $\lambda_s \sim \Gamma(\lambda_s|c_o, d_o)$ and $\lambda_\epsilon \sim \Gamma(\lambda_\epsilon|e_o, f_o)$, where Γ denotes the Gamma distribution and c_o, d_o, e_o and f_o are the hyper-parameters. The model thus formed is completely conjugate, therefore Bayesian inferencing can be performed over it with Gibbs sampling using analytical expressions. We derive these expressions for the proposed approach and state the final sampling equations below. Detailed derivations of the Gibbs sampling equations can be found in the provided supplementary material.

We denote the contribution of the k^{th} dictionary atom $\boldsymbol{\varphi}_k$ to $\mathbf{y}_i$ as, $\mathbf{y}_{i_{\varphi_k}} = \mathbf{y}_i - \mathbf{\Phi}(\mathbf{z}_i \odot \mathbf{s}_i) + \boldsymbol{\varphi}_k(z_{ik}s_{ik})$, and the ℓ_2 norm of a vector by $\|.\|_2$. Using these notations, we obtain the following analytical expressions for the Gibbs sampling process used in our approach:

Sample $\boldsymbol{\varphi}_k$: from $\mathcal{N}(\boldsymbol{\varphi}_k|\boldsymbol{\mu}_k, \lambda_k^{-1}\mathbf{I}_L)$, where

$$\lambda_k = \lambda_{k_o} + \lambda_{\epsilon_o}\sum_{i=1}^{mn}(z_{ik}s_{ik})^2; \boldsymbol{\mu}_k = \frac{\lambda_{\epsilon_o}}{\lambda_k}\sum_{i=1}^{mn}(z_{ik}s_{ik})\mathbf{y}_{i_{\varphi_k}}$$

Sample z_{ik}: from $\text{Bern}\Big(z_{ik}|\frac{\xi\pi_{k_o}}{1-\pi_{k_o}+\xi\pi_{k_o}}\Big)$, where

$$\xi = \exp\Big(-\frac{\lambda_{\epsilon_o}}{2}(\boldsymbol{\varphi}_k^{\mathrm{T}}\boldsymbol{\varphi}_k s_{ik}^2 - 2s_{ik}\mathbf{y}_{i_{\varphi_k}}^{\mathrm{T}}\boldsymbol{\varphi}_k)\Big)$$

Sample s_{ik}: from $\mathcal{N}(s_{ik}|\mu_s, \lambda_s^{-1})$, where

$$\lambda_s = \lambda_{s_o} + \lambda_{\epsilon_o}z_{ik}^2\boldsymbol{\varphi}_k^{\mathrm{T}}\boldsymbol{\varphi}_k \;;\; \mu_s = \frac{\lambda_{\epsilon_o}}{\lambda_s}z_{ik}\boldsymbol{\varphi}_k^{\mathrm{T}}\mathbf{y}_{i_{\varphi_k}}$$

Sample π_k: from $\text{Beta}(\pi_k|a, b)$, where

$$a = \frac{a_o}{K} + \sum_{i=1}^{mn}z_{ik} \;;\; b = \frac{b_o(K-1)}{K} + (mn) - \sum_{i=1}^{mn}z_{ik}$$

Sample λ_s: from $\Gamma(\lambda_s|c, d)$, where

$$c = \frac{Kmn}{2} + c_o \;;\; d = \frac{1}{2}\sum_{i=1}^{mn}||\mathbf{s}_i||_2^2 + d_o$$

Sample λ_ϵ: from $\Gamma(\lambda_\epsilon|e, f)$, where

$$e = \frac{Lmn}{2} + e_o \;;\; f = \frac{1}{2}\sum_{i=1}^{mn}||\mathbf{y}_i - \mathbf{\Phi}(\mathbf{z}_i \odot \mathbf{s}_i)||_2^2 + f_o$$

As a result of Bayesian inferencing, we obtain sets of posterior distributions over the model parameters. We are interested in two of them. (a) The set of distributions over the atoms of the dictionary, $\aleph \stackrel{\text{def}}{=} \{\mathcal{N}(\boldsymbol{\varphi}_k|\boldsymbol{\mu}_k, \mathbf{\Lambda}_k^{-1}) : k \in \mathcal{K}\} \subset \mathbb{R}^L$ and (b) the set of distributions over the components of the support indicator vectors $\beth \stackrel{\text{def}}{=} \{\text{Bern}(\pi_k) : k \in \mathcal{K}\} \subset \mathbb{R}$. Here, $\text{Bern}(\pi_k)$ is followed by the k^{th} components of all the support indicator vectors simultaneously, *i.e.* $\forall i \in \{1, ..., mn\}$, $z_{ik} \sim \text{Bern}(\pi_k)$. These sets are used in the later stages of the proposed approach.

In the above model, we have placed Gaussian priors over the dictionary atoms, enforcing our prior belief of relative smoothness of the material spectra. Note that, the correct value of $|\mathcal{K}|$ is also inferred at this stage. We refer to the pioneering work by Paisley and Carin [23] for the theoretical details in this regard. In our inferencing process, the desired value of $|\mathcal{K}|$ manifests itself as the total number of dictionary atoms for which $\pi_k \neq 0$ after convergence. To implement this, we start with $K \to \infty$ and later drop the dictionary atoms corresponding to $\pi_k = 0$ during the sampling process.

With the computed $\aleph$, we estimate $\mathbf{\Phi}$ (stage 2 in Fig. 2) by drawing multiple samples from the distributions in the set and computing their respective means. It is also possible to directly use the mean parameters of the inferred distributions as the estimates of the dictionary atoms, but the former is preferred for robustness. Henceforth, we will consider the dictionary, instead of the distributions over its atoms, as the final outcome of the Bayesian dictionary learning process. The transformed dictionary is simply computed as $\widehat{\mathbf{\Phi}} = \mathbf{\Upsilon}\mathbf{\Phi}$. Recall that, the matrix $\mathbf{\Upsilon}$ relates the spectral quantizations of the two imaging modalities under consideration and it is known *a priori*.

4.2. Bayesian Sparse Coding

Once $\widehat{\mathbf{\Phi}}$ is known, we use it to compute the sparse codes of $\mathbf{Y}$. The intention is to obtain the codes of the high resolution image and use them with $\mathbf{\Phi}$ to estimate $\mathbf{T}$. Although some popular strategies for sparse coding already exist, *e.g.* Orthogonal Matching Pursuit [26] and Basis Pursuit [12], but their performance is inferior when used with the Bayesian dictionaries learned using the Beta process. There are two main reasons for that. (a) Atoms of the Bayesian dictionaries are not constrained to ℓ_2 unit norm. (b) With these atoms, there is an associated set of Bernoulli distributions which must not be contradicted by the underlying support of the sparse codes. In some cases, it may be easy to modify an existing strategy to cater for (a), but it is not straightforward to take care of (b) in these approaches.

We propose a simple, yet effective method for Bayesian sparse coding that can be generically used with the dictionaries learned using the Beta process. The proposal is to follow a procedure similar to the Bayesian dictionary learning,

with three major differences. For a clear understanding, we explain these differences as modifications to the inferencing process of the Bayesian dictionary learning, following the same notational conventions as above.

1) Use $\mathcal{N}(\widehat{\boldsymbol{\varphi}}_k|\boldsymbol{\mu}_{k_o}, \lambda_{k_o}^{-1}\mathbf{I}_l)$ as the prior distribution over the k^{th} dictionary atom, where $\lambda_{k_o} \to \infty$ and $\boldsymbol{\mu}_{k_o} = \widehat{\boldsymbol{\varphi}}_k$. Considering that $\widehat{\boldsymbol{\Phi}}$ is already a good estimate of the dictionary[2], this is an intuitive prior. It entails, $\widehat{\boldsymbol{\varphi}}_k$ is sampled from the following posterior distribution while inferencing: **Sample** $\widehat{\boldsymbol{\varphi}}_k$**:** from $\mathcal{N}(\widehat{\boldsymbol{\varphi}}_k|\boldsymbol{\mu}_k, \lambda_k^{-1}\mathbf{I}_l)$, where

$$\lambda_k = \lambda_{k_o} + \lambda_{\epsilon_o}\sum_{i=1}^{MN}(z_{ik}s_{ik})^2;$$

$$\boldsymbol{\mu}_k = \frac{\lambda_{\epsilon_o}}{\lambda_k}\sum_{i=1}^{MN}(z_{ik}s_{ik})\mathbf{y}_{i_{\widehat{\varphi}_k}} + \frac{\lambda_{k_o}}{\lambda_k}\boldsymbol{\mu}_{k_o}$$

In the above equations, $\lambda_{k_o} \to \infty$ signifies $\lambda_k \approx \lambda_{k_o}$ and $\boldsymbol{\mu}_k \approx \boldsymbol{\mu}_{k_o}$. It further implies that we are likely to get similar samples against multiple draws from the distribution. In other words, we can not only ignore to update the posterior distributions over the dictionary atoms during the inferencing process, but also approximate them with a fixed matrix. A sample from the k^{th} posterior distribution is then the k^{th} column of this matrix. Hence, from the implementation perspective, Bayesian sparse coding directly uses the atoms of $\widehat{\boldsymbol{\Phi}}$ as the samples from the posterior distributions.

2) Sample the support indicator vectors in accordance with the Bernoulli distributions associated with the fixed dictionary atoms. To implement this, while inferencing, we fix the distributions over the support indicator vectors according to $\mathfrak{F}$. As shown in Fig. 2, we use the vector $\boldsymbol{\pi} \in \mathbb{R}^{|\mathcal{K}|}$ for this purpose, which stores the parameters of the distributions in the set $\mathfrak{F}$. While sampling, we directly use the k^{th} component of $\boldsymbol{\pi}$ as π_k. It is noteworthy that using $\boldsymbol{\pi}$ in coding $\mathbf{Y}$ also imposes the self-consistency of the scene between the high resolution image $\mathbf{Y}$ and the hyperspectral image $\mathbf{Y}^h$.

Incorporating the above proposals in the Gibbs sampling process and performing the inferencing can already result in a reasonably accurate sparse representation of $\mathbf{y}$ over $\widehat{\boldsymbol{\Phi}}$. However, a closer examination of the underlying probabilistic settings reveals that a more accurate estimate of the sparse codes is readily obtainable.

Lemma 4.1 *With* $\mathbf{y} \in \mathcal{R}(\widehat{\boldsymbol{\Phi}})$ (i.e. $\exists\boldsymbol{\alpha}\ s.t.\ \mathbf{y} = \widehat{\boldsymbol{\Phi}}\boldsymbol{\alpha}$) *and* $|\mathcal{K}| > l$*, the best estimate of the representation of* $\mathbf{y}$*, in the mean squared error sense*[3]*, is given by* $\tilde{\boldsymbol{\alpha}}_{\text{opt}} = \mathbb{E}\big[\mathbb{E}[\boldsymbol{\alpha}|\mathbf{z}]\big]$*, where* $\mathcal{R}(.)$ *is the range operator,* $\mathbb{E}[.]$ and $\mathbb{E}[.|.]$ *are the expectation and the conditional expectation operators, respectively.*

Proof: Let $\tilde{\boldsymbol{\alpha}} \in \mathbb{R}^{|\mathcal{K}|}$ be an estimate of the representation $\boldsymbol{\alpha}$ of $\mathbf{y}$, over $\widehat{\boldsymbol{\Phi}}$. We can define the mean square error (MSE) as the following:

$$\text{MSE} = \mathbb{E}\big[||\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}||_2^2\big] \tag{1}$$

In our settings, the components of a support indicator vector $\mathbf{z}$ are independent draws from Bernoulli distributions. Let $\mathcal{Z}$ be the set of all possible support indicator vectors in $\mathbb{R}^{|\mathcal{K}|}$, *i.e.* $|\mathcal{Z}| = 2^{|\mathcal{K}|}$. Thus, there is a non-negative probability of selection $P(\mathbf{z})$ associated with each $\mathbf{z} \in \mathcal{Z}$ such that $\sum_{\mathbf{z}\in\mathcal{Z}} P(\mathbf{z}) = 1$. Indeed, the probability mass function $p(\mathbf{z})$ depends on the vector $\boldsymbol{\pi}$ that assigns higher probabilities to the vectors indexing more important dictionary atoms.

We can model the generation of $\boldsymbol{\alpha}$ as a two step sequential process: 1) Random selection of $\mathbf{z}$ with probability $P(\mathbf{z})$. 2) Random selection of $\boldsymbol{\alpha}$ according to a conditional probability density function $p(\boldsymbol{\alpha}|\mathbf{z})$. Here, the selection of $\boldsymbol{\alpha}$ implies the selection of the corresponding weight vector $\mathbf{s}$ and then computing $\boldsymbol{\alpha} = \mathbf{z} \odot \mathbf{s}$. Under this perspective, MSE can be re-written as:

$$\text{MSE} = \sum_{\mathbf{z}\in\mathcal{Z}} P(\mathbf{z})\mathbb{E}\big[||\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}||_2^2 \mid \mathbf{z}\big] \tag{2}$$

The conditional expectation in (2) can be written as:

$$\mathbb{E}\big[||\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}||_2^2|\mathbf{z}\big] = ||\tilde{\boldsymbol{\alpha}}||_2^2 - 2\tilde{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbb{E}[\boldsymbol{\alpha}|\mathbf{z}] + \mathbb{E}\big[||\boldsymbol{\alpha}||_2^2|\mathbf{z}\big] \tag{3}$$

We can write the last term in (3) as the following:

$$\mathbb{E}\big[||\boldsymbol{\alpha}||_2^2|\mathbf{z}\big] = \big\|\mathbb{E}[\boldsymbol{\alpha}|\mathbf{z}]\big\|_2^2 + \mathbb{E}\Big[\big\|\boldsymbol{\alpha} - \mathbb{E}[\boldsymbol{\alpha}|\mathbf{z}]\big\|_2^2|\mathbf{z}\Big] \tag{4}$$

For brevity, let us denote the second term in (4) as $\mathbb{V}_{\mathbf{z}}$. By combining (2)-(4) we get:

$$\text{MSE} = \sum_{\mathbf{z}\in\mathcal{Z}} P(\mathbf{z})\big\|\tilde{\boldsymbol{\alpha}} - \mathbb{E}[\boldsymbol{\alpha}|\mathbf{z}]\big\|_2^2 + \sum_{\mathbf{z}\in\mathcal{Z}} P(\mathbf{z})\mathbb{V}_{\mathbf{z}} \tag{5}$$

$$= \mathbb{E}\big[\big\|\tilde{\boldsymbol{\alpha}} - \mathbb{E}[\boldsymbol{\alpha}|\mathbf{z}]\big\|_2^2\big] + \mathbb{E}\big[\mathbb{V}_{\mathbf{z}}\big] \tag{6}$$

Differentiating R.H.S. of (6) with respect to $\tilde{\boldsymbol{\alpha}}$ and equating it to zero, we get $\tilde{\boldsymbol{\alpha}}_{\text{opt}} = \mathbb{E}\big[\mathbb{E}[\boldsymbol{\alpha}|\mathbf{z}]\big]$[4], that minimizes the mean squared error.

Notice that, with the aforementioned proposals incorporated in the sampling process, it is possible to independently perform the inferencing multiple, say Q, times. This would result in Q support indicator vectors $\mathbf{z}_q$ and weight vectors $\mathbf{s}_q$ for $\mathbf{y}$, where $q \in \{1, ...Q\}$.

Lemma 4.2 *For* $Q \to \infty$, $\frac{1}{Q}\sum_{q=1}^{Q}\mathbf{z}_q \odot \mathbf{s}_q = \mathbb{E}\big[\mathbb{E}[\boldsymbol{\alpha}|\mathbf{z}]\big]$.

[2]This is true because $\widehat{\boldsymbol{\Phi}}$ is an exact transform of $\boldsymbol{\Phi}$, which in turn, is computed with high confidence.

[3]The metric is chosen based on the existing literature in hyperspectral image super resolution [18],[17],[2].

[4]Detailed mathematical derivation of each step used in the proof is also provided in the supplementary material.

Proof: We only discuss an informal proof of Lemma 4.2. The following statements are valid in our settings:
(a) $\exists\, \boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j \; s.t. \; (\boldsymbol{\alpha}_i \neq \boldsymbol{\alpha}_j) \wedge (\boldsymbol{\alpha}_i = \mathbf{z} \odot \mathbf{s}_i) \wedge (\boldsymbol{\alpha}_j = \mathbf{z} \odot \mathbf{s}_j)$
(b) $\exists\, \mathbf{z}_i, \mathbf{z}_j \; s.t. \; (\mathbf{z}_i \neq \mathbf{z}_j) \wedge (\boldsymbol{\alpha} = \mathbf{z}_i \odot \mathbf{s}_i) \wedge (\boldsymbol{\alpha} = \mathbf{z}_j \odot \mathbf{s}_j)$
where $\wedge$ denotes the logical *and*; $\boldsymbol{\alpha}_i$ and $\boldsymbol{\alpha}_j$ are instances of two distinct solutions of the underdetermined system $\mathbf{y} = \widehat{\boldsymbol{\Phi}}\boldsymbol{\alpha}$. In the above statements, (a) refers to the possibility of distinct representations with the same support and (b) refers to the existence of distinct support indicator vectors for a single representation. Validity of these conditions can be easily verified by noticing that $\mathbf{z}$ and $\mathbf{s}$ are allowed to have zero components. For a given inferencing process, the final computed vectors $\mathbf{z}$ and $\mathbf{s}$ are drawn according to valid probability distributions. Thus, (a) and (b) entail that the mean of Q independently computed representations, is equivalent to $\mathbb{E}\big[\mathbb{E}[\boldsymbol{\alpha}|\mathbf{z}]\big]$ when $Q \rightarrow \infty$.

3) In the light of Lemma 4.1 and 4.2, we propose to independently repeat the inferencing process Q times, where Q is a large number (*e.g.* 100), and finally compute the code matrix $\widetilde{\mathbf{A}}$ (in Fig. 2) as $\widetilde{\mathbf{A}} = \frac{1}{Q}\sum_{q=1}^{Q} \mathbf{Z}_q \odot \mathbf{S}_q$, where $\widetilde{\mathbf{A}}$ has $\tilde{\boldsymbol{\alpha}}_{i \in \{1,\ldots,MN\}}$ as its columns. The matrices $\mathbf{Z}_q, \mathbf{S}_q \in \mathbb{R}^{|\mathcal{K}| \times MN}$ are the support matrix and the weight matrix, respectively, formed by arranging the support indicator vectors and the weight vectors as their columns. Note that, the finally computed codes $\widetilde{\mathbf{A}}$ may by dense as compared to individual $\mathbf{Z}_q$.

With the estimated $\widetilde{\mathbf{A}}$ and the dictionary $\boldsymbol{\Phi}$, we compute the target super resolution image $\mathbf{T}$ by re-arranging the columns of $\overline{\mathbf{T}} = \boldsymbol{\Phi}\widetilde{\mathbf{A}}$ (stage 4 in Fig. 2) into the pixels of hyperspectral image.

5. Experimental Evaluation

The proposed approach has been thoroughly evaluated using ground based imagery as well as remotely sensed data. For the former, we performed exhaustive experiments on two public databases, namely, the CAVE database [30] and the Harvard database [9]. CAVE comprises 32 hyperspectral images of everyday objects with dimensions $512 \times 512 \times 31$, where 31 represents the spectral dimension. The spectral images are in the wavelength range 400 - 700nm, sampled at a regular interval of 10nm. The Harvard database consists of 50 images of indoor and outdoor scenes with dimensions $1392 \times 1040 \times 31$. The spectral samples are taken at every 10nm in the range 420 - 720nm. For the remote sensing data, we chose a $512 \times 512 \times 224$ hyperspectral image[5] acquired by the NASA's Airborne Visible Infrared Imaging Spectrometer (AVIRIS) [15]. This image has been acquired over the Cuprite mines in Nevada, in the wavelength range 400 - 2500nm with 10nm sampling interval. We followed the experimental protocol of [2] and [18]. For benchmarking, we compared the results with the existing best reported results in the literature under the same protocol, unless the code was made public by the authors. In the latter case, we performed experiments using the provided code and the optimized parameter values. The reported results are in the range of 8 bit images.

In our experiments, we consider the images from the databases as the ground truth. A low resolution hyperspectral image $\mathbf{Y}^h$ is created by averaging the ground truth over 32×32 spatially disjoint blocks. For the Harvard database, $1024 \times 1024 \times 31$ image patches were cropped from the top left corner of the images, to make the spatial dimensions of the ground truth multiples of 32. For the ground based imagery, we assume the high resolution image $\mathbf{Y}$ to be an RGB image of the same scene. We simulate this image by integrating the ground truth over its spectral dimension using the spectral response of Nikon D700[6]. For the remote sensing data, we consider $\mathbf{Y}$ to be a multispectral image. Following [2], we create this image by directly selecting six spectral images from the ground truth against the wavelengths 480, 560, 660, 830, 1650 and 2220 nm. Thus, in this case, $\boldsymbol{\Upsilon}$ is a 6×224 binary matrix that selects the corresponding rows of $\boldsymbol{\Phi}$. The mentioned wavelengths correspond to the visible and mid-infrared channels of USGS/NASA Landsat 7 satellite.

We compare our results with the recently proposed approaches, namely, the Matrix Factorization based method (MF) [18], the Spatial Spectral Fusion Model (SSFM) [17], the ADMM based approach [29], the Coupled Matrix Factorization method (CMF) [31] and the spatio-spectral sparse representation approach, GSOMP [2]. These matrix factorization based approaches constitute the state of the art in this area [2]. In order to show the performance difference between these methods and the other approaches mentioned in Section 2, we also report some results of the Component Substitution Method (CSM) [1], taken directly from [18].

The top half of Table 1 shows results on seven different images from the CAVE database. We chose these images because they are commonly used for benchmarking in the existing literature [2],[29],[18]. The table shows the root mean squared error (RMSE) of the reconstructed super resolution images. The approaches highlighted in red additionally require the knowledge of the down-sampling matrix that converts the ground truth to the acquired hyperspectral image. Hence, they are of less practical value [2]. As can be seen, our approach outperforms most of the existing methods by a considerable margin on all the images. Only the results of GSOMP are comparable to our method. However, GSOMP operates under the assumption that nearby pixels in the target image are spectrally similar. The assumption is enforced with the help of two extra algorithm parameters. Fine tuning these parameters is often non-trivial, as many

[5] *http://aviris.jpl.nasa.gov/data/free_data.html*.

[6] The response and integration limits can be found at *http://www.maxmax.com/spectral_response.htm*

Table 1. Benchmarking of the proposed approach: The RMSE values are in the range of 8 bit images. The best results are shown in bold. The approaches highlighted in red additionally require the knowledge of the spatial transform between the input images.

Method	CAVE database [30]						
	Beads	Spools	Painting	Balloons	Photos	CD	Cloth
CSM [1]	28.5	-	12.2	13.9	13.1	13.3	-
MF [18]	8.2	8.4	4.4	3.0	3.3	8.2	6.1
SSFM [17]	9.2	6.1	4.3	-	3.7	-	10.2
ADMM [29]	6.1	5.3	6.7	2.1	3.4	6.5	9.5
CMF [31]	6.6	15.0	26.0	5.5	11.0	11.0	20.0
GSOMP [2]	6.1	5.0	4.0	2.3	2.2	7.5	**4.0**
Proposed	**5.4**	**4.6**	**1.9**	**2.1**	**1.6**	**5.3**	**4.0**
	Harvard database [9]						
	Img 1	Img b5	Img b8	Img d4	Img d7	Img h2	Img h3
MF [18]	3.9	2.8	6.9	3.6	3.9	3.7	2.1
SSFM [17]	4.3	2.6	7.6	4.0	4.0	4.1	2.3
GSOMP [2]	1.2	**0.9**	5.9	2.4	2.1	1.0	**0.5**
Proposed	**1.1**	**0.9**	**4.3**	**0.5**	**0.8**	**0.7**	**0.5**

Table 2. Exhaustive experiment results: The means and the standard deviations of the RMSE values are computed over the complete databases.

Method	CAVE database [30]	Harvard database [9]
	Mean $\pm$ Std. Dev	Mean $\pm$ Std. Dev
GSOMP [2]	3.66 ± 1.51	2.84 ± 2.24
Proposed	$\mathbf{3.06 \pm 1.12}$	$\mathbf{1.74 \pm 1.49}$

of the nearby pixels in an image can also have dissimilar spectra. There is no provision for automatic adjustment of the parameter values for such cases. Therefore, an image reconstructed by GSOMP can often suffer from spatial artifacts. For instance, even though the parameters of GSOMP are optimized specifically for the sample image in Fig. 3, the spatial artifacts are still visible. The figure also compares the RMSE of our approach with that of GSOMP, as a function of the spectral bands of the image. The RMSE curve is lower and smoother for the proposed approach.

The results on the images from the Harvard database are shown in the bottom half of Table 1. These results also favor our approach. Results of ADMM and CMF have never been reported for the Harvard database. In Table 2, we report the means and the standard deviations of the RMSE values of the proposed approach over the complete databases. The results are compared with GSOMP using the public code provided by the authors and the optimal parameter settings for each database, as mentioned in [2]. Fair comparison with other approaches is not possible because of the unavailability of the public codes and results on the full databases. However, based on Table 1 and the mean RMSE values of 4.24 ± 2.08 and 4.98 ± 1.97 for ADMM and MF, respectively, reported by Wycoff *et al.* [29], on 20 images

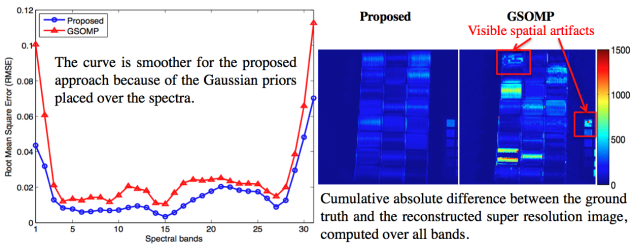

Figure 3. Comparison of the proposed approach with GSOMP [2] on image 'Spools' (CAVE database) [30].

from the CAVE database, we can safely conjecture that the other methods are unlikely to outperform our approach on the full databases. Table 2 clearly indicates the consistent performance of the proposed approach. For our approach, results on the individual images of the complete databases can be found in the supplementary material of the paper, where we also provide the Matlab code/demo for the proposed approach (that will eventually be made public).

For qualitative analysis, Fig. 4 shows spectral samples from two reconstructed super resolution hyperspectral images, against wavelengths 460, 540 and 620nm. The spectral images are shown along the ground truth and their absolute difference with the ground truth. Spectral samples of the input 16×16 hyperspectral images are also shown. Successful hyperspectral image super resolution is clearly evident from the images. Further qualitative results on the images mentioned in Table 1 are given in the supplementary material of the paper. For the remote sensing image acquired by AVIRIS, the RMSE value of the proposed approach is 1.63. This is also lower than the previously reported values of 2.14, 3.06 and 3.11 for GSOMP, MF and SSFM respectively, in [2]. For the AVIRIS image, the spectral samples of the reconstructed super resolution image at 460, 540, 620 and 1300 nm are shown in Fig. 5.

6. Discussion

In the above experiments, we initialized the *Bayesian dictionary learning stage* as follows. The parameters a_o, b_o, c_o, d_o, e_o and f_o were set to 10^{-6}. From the sampling equations in Section 4.1, it is easy to see that these values do not influence the posterior distributions much, and other such small values would yield similar results. We initialized $\pi_{k_o} = 0.5, \forall\ k$, to give the initial Bernoulli distributions the largest variance [5]. We initialized the Gibbs sampling process with $K = 50$ for all the images. This value is based on our prior belief that the total number of the materials in a given scene is generally less than 50. The final value of $|\mathcal{K}|$ was inferred by the learning process itself, which ranged over 10 to 33 for different images. We initialized λ_{ϵ_o} to the precision of the pixels in $\mathbf{Y}^h$ and randomly chose $\lambda_{s_o} = 1$. Following [33], λ_{k_o} was set to L. The parameter setting was kept unchanged for all the datasets

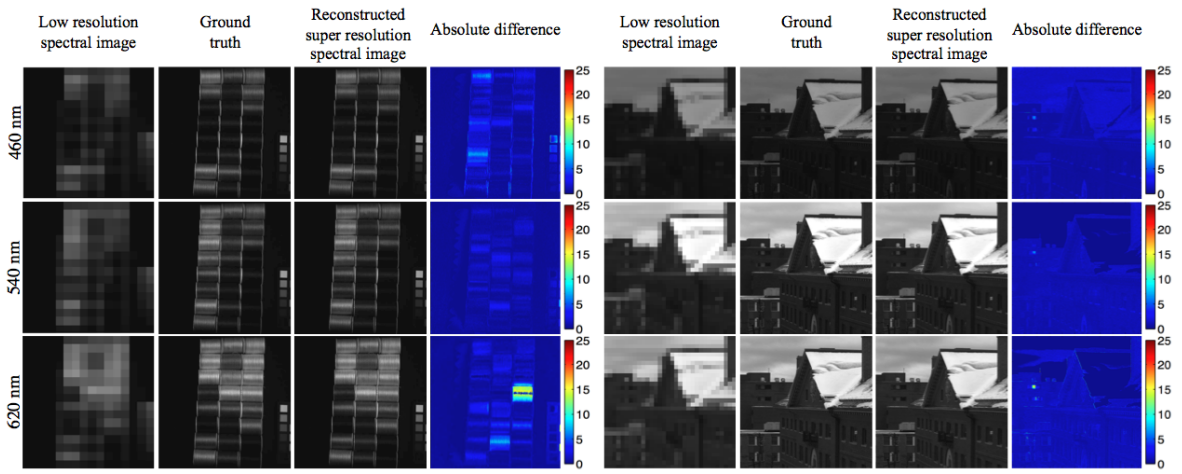

Figure 4. Super resolution image reconstruction at 460, 540 and 620 nm: The images include the low resolution spectral image, ground truth, reconstructed spectral image and the absolute difference between the ground truth and the reconstructed spectral image. (Left) 'Spools' from the CAVE database [30]. (Right) 'Img 1' form the Harvard database [9].

without further fine tuning. We ran fifty thousand Gibbs sampling iterations from which the last 100 were used to sample the distributions to compute $\mathbf{\Phi}$. On average, this process took around 3 minutes for the CAVE images and around 8 minutes for the Harvard images. For the AVIRIS data, this time was 12.53 minutes. The timing is for Matlab implementation on an Intel Core i7 CPU at 3.6 GHz with 8 GB RAM. For the *Bayesian sparse coding stage*, we again used 10^{-6} as the initial value for the parameters a_o to f_o. We respectively initialized λ_{s_o} and λ_{ϵ_o} to the final values of λ_s and λ_ϵ of the dictionary learning stage. We ran the inferencing process $Q = 128$ times with 100 iterations in each run. It is worth noticing that, in the proposed sparse coding strategy, it is possible to run the inferencing processes independent of each other. This makes the sparse coding stage naturally suitable for multi-core processing. On average, a single sampling process required around 1.75 minutes for a CAVE image and approximately 7 minutes for a Harvard image. For the AVIRIS image, this time was 11.23 minutes.

The proposed approach outperforms the existing methods on ground based imagery as well as remotely sensed data, without requiring explicit parameter tuning. This distinctive characteristic of the proposed approach comes from exploiting the non-parametric Bayesian framework.

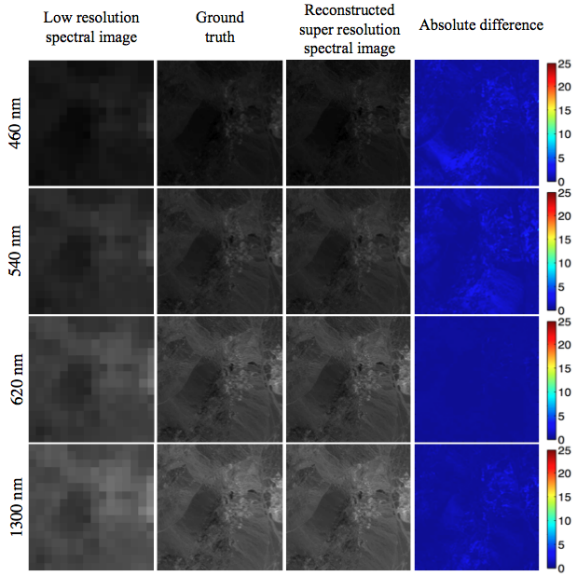

Figure 5. Spectral images at 460, 540, 620 and 1300 nm for the AVIRIS [15] data. Reconstructed spectral images (512×512) are shown along their absolute difference with the ground truth (512×512). The low resolution images (16×16) are also shown.

7. Conclusion

We proposed a Bayesian sparse representation based approach for hyperspectral image super resolution. Using the non-parametric Bayesian dictionary learning, the proposed approach learns distributions for the scene spectra and their proportions in the image. Later, this information is used to sparse code a high resolution image (*e.g.* RGB) of the same scene. For that purpose, we proposed a Bayesian sparse coding method that can be generically used with the dictionaries learned using the Beta process. Theoretical analysis is provided to show the effectiveness of the method. We used the learned sparse codes with the image spectra to construct the super resolution hyperspectral image. Exhaustive experiments on three public data sets show that the proposed approach outperforms the existing state of the art.

Acknowledgements

This work is supported by ARC Grant DP110102399.

References

[1] B. Aiazzi, S. Baronti, and M. Selva. Improving component substitution pansharpening through multivariate regression of MS+Pan data. *IEEE Trans. Geosci. Remote Sens.*, 45(10):3230–3239, 2007.

[2] N. Akhtar, F. Shafait, and A. Mian. Sparse spatio-spectral representation for hyperspectral image super-resolution. In *ECCV*, pages 63 – 78, Sept. 2014.

[3] N. Akhtar, F. Shafait, and A. Mian. Futuristic greedy approach to sparse unmixing of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.*, 53(4):2157–2174, April 2015.

[4] N. Akhtar, F. Shafait, and A. S. Mian. SUnGP: A greedy sparse approximation algorithm for hyperspectral unmixing. In *22nd International Conference on Pattern Recognition*, pages 3726–3731, 2014.

[5] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, 2011.

[7] W. J. Carper, T. M. Lilles, and R. W. Kiefer. The use of intensity-hue-saturation transformations for merging SOPT panchromatic and multispectral image data. *Photogram. Eng. Remote Sens.*, 56(4), 1990.

[8] M. Cetin and N. Musaoglu. Merging hyperspectral and panchromatic image data: Qualitative and quantitative analysis. *Int. J. Remote Sens.*, 30(7):1779–1804, 2009.

[9] A. Chakrabarti and T. Zickler. Statistics of real-world hyperspectral images. In *CVPR*, pages 193–200, June 2011.

[10] P. S. Chavez, S. C. Sides, and J. A. Anderson. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. *Photogramm. Eng. Rem. S.*, 30(7):1779–1804, 1991.

[11] C. Chen, Y. Li, W. Liu, and J. Huang. Image fusion with local spectral consistency and dynamic gradient sparsity. In *CVPR*, June 2014.

[12] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Rev.*, 43(1):129–159, 2001.

[13] J. B. Dias, A. Plaza, G. C. Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote Sens. Mag.*, 1(2):6–36, 2013.

[14] M. Fauvel, Y. Tarabalka, J. Benediktsson, J. Chanussot, and J. Tilton. Advances in spectral-spatial classification of hyperspectral images. *Proc. IEEE*, 101(3):652–675, 2013.

[15] 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, M. R. Olah, and O. Williams. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sens. Environ.*, 65(3):227 – 248, 1998.

[16] R. Haydn, G. W. Dalke, J. Henkel, and J. E. Bare. Application of the IHS color transform to the processing of multisensor data and image enhancement. In *Proc. Int. Symp. on Remote Sens. of Env.*, 1982.

[17] B. Huang, H. Song, H. Cui, J. Peng, and Z. Xu. Spatial and spectral image fusion using sparse matrix factorization. *IEEE Trans. Geosci. Remote Sens.*, 52(3):1693–1704, 2014.

[18] R. Kawakami, J. Wright, Y.-W. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi. High-resolution hyperspectral imaging via matrix factorization. In *CVPR*, pages 2329–2336, June 2011.

[19] N. Keshava and J. Mustard. Spectral unmixing. *IEEE Signal Process. Mag.*, 19(1):44–57, Jan 2002.

[20] Z. Khan, F. Shafait, and A. Mian. Hyperspectral imaging for ink mismatch detection. In *ICDAR*, pages 877 – 881, Aug 2013.

[21] A. Minghelli-Roman, L. Polidori, S. Mathieu-Blanc, L. Loubersac, and F. Cauneau. Spatial resolution improvement by merging MERIS-ETM images for coastal water monitoring. *IEEE Geosci. Remote Sens. Lett.*, 3(2):227–231, 2006.

[22] H. V. Nguyen, A. Banerjee, and R. Chellappa. Tracking via object reflectance using a hyperspectral video camera. In *CVPRW*, pages 44–51, June 2010.

[23] J. Paisley and L. Carin. Nonparametric factor analysis with beta process priors. In *ICML*, pages 777–784, June 2009.

[24] R. Rubinstein, A. Bruckstein, and M. Elad. Dictionaries for sparse representation modeling. *Proc. IEEE*, 98(6):1045–1057, 2010.

[25] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson. 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.

[26] J. Tropp and A. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory*, 53(12):4655–4666, 2007.

[27] J. A. Tropp, A. C. Gilbert, and M. J. Strauss. Algorithms for simultaneous sparse approximation. part i: Greedy pursuit. *Signal Processing*, 86(3):572–588, 2006.

[28] M. Uzair, A. Mahmood, and A. Mian. Hyperspectral face recognition using 3D-DCT and partial least squares. In *British Machine Vision Conf. (BMVC)*, pages 57.1–57.10, 2013.

[29] E. Wycoff, T.-H. Chan, K. Jia, W.-K. Ma, and Y. Ma. A non-negative sparse promoting algorithm for high resolution hyperspectral imaging. In *ICASSP*, pages 1409 – 1413, May 2013.

[30] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar. Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum. Technical report, Dept. of Comp. Sci., Columbia University CUCS-061-08, Nov 2008.

[31] N. Yokoya, T. Yairi, and A. Iwasaki. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Trans. Geosci. Remote Sens.*, 50(2):528–537, 2012.

[32] D. Zhang, W. Zuo, and F. Yue. A comparative study of palmprint recognition algorithms. *ACM Comput. Surv.*, 44(1):2:1–2:37, 2012.

[33] M. Zhou, H. Chen, J. Paisley, L. Ren, G. Sapiro, and L. Carin. Non-Parametric Bayesian Dictionary Learning for Sparse Image Representations. In *NIPS*, pages 2295–2303, Dec 2009.

[34] Y. Zhou, H. Chang, K. Barner, P. Spellman, and B. Parvin. Classification of histology sections via multispectral convolutional sparse coding. In *CVPR*, June 2014.

[35] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel. Unmixing-based multisensor multiresolution image fusion. *IEEE Trans. Geosci. Remote Sens.*, 37(3):1212–1226, 1999.

[36] R. Zurita-Milla, J. G. Clevers, and M. E. Schaepman. Unmixing-based Landsat TM and MERIS FR data fusion. *IEEE Trans. Geosci. Remote Sens.*, 5(3):453–457, 2008.