
Pan-Sharpening with a Bayesian Nonparametric Dictionary Learning Model

Xinghao Ding Yiyong Jiang Yue Huang
Department of Communications Engineering
Xiamen University, Fujian, China

John Paisley
Department of Electrical Engineering
Columbia University, New York, USA

Abstract

Pan-sharpening, a method for constructing high resolution images from low resolution observations, has recently been explored from the perspective of compressed sensing and sparse representation theory. We present a new pan-sharpening algorithm that uses a Bayesian nonparametric dictionary learning model to give an underlying sparse representation for image reconstruction. In contrast to existing dictionary learning methods, the proposed method infers parameters such as dictionary size, patch sparsity and noise variances. In addition, our regularization includes image constraints such as a total variation penalization term and a new gradient penalization on the reconstructed PAN image. Our method does not require high resolution multiband images for dictionary learning, which are unavailable in practice, but rather the dictionary is learned directly on the reconstructed image as part of the inversion process. We present experiments on several images to validate our method and compare with several other well-known approaches.

1 Introduction

Remote sensing technology has developed rapidly in recent decades. There are currently a variety of optical remote sensors, such as IKONOS, Pléiades and QuickBird, collecting satellite imagery with varying spatial and spectral resolutions. Due to physical constraints, many of these remote sensors do not collect images that are simultaneously of high spatial and spectral resolution [11]. As a result, data provided by most satellites comprise a high resolution panchromatic (PAN), i.e., gray-scale image, and several low resolution multispectral (LRMS) images, which

Appearing in Proceedings of the 17th International Conference on Artificial Intelligence and Statistics (AISTATS) 2014, Reykjavik, Iceland. JMLR: W&CP volume 33. Copyright 2014 by the authors.

include the color spectrum and other frequencies such as near infrared. In many remote sensing applications, such as classification, feature extraction and positioning, the corresponding high resolution multispectral (HRMS) images are desired. Pan-sharpening is a method whereby the spectral information in the LRMS images is fused with the spatial information in the PAN image to output an approximation of the underlying HRMS images [3].

Various methods have been proposed for pan-sharpening. The most common are based on a projection-substitution approach, which assumes that the PAN image is equivalent to a linear combination of the HRMS images [12]. Among the projection-substitution methods, intensity-hue-saturation (IHS) [19, 1, 16], principal component analysis (PCA) [17], and the Brovey transform [7] are the most popular because of their relatively straightforward implementation and fast computation. These three methods reconstruct HRMS images with good spatial details, but tend to have significant spectral distortions because the PAN and HRMS images do not share the same spectral range.

Recently, another approach has been to treat the fusion problem as an ill-posed inverse problem and employ a regularization scheme that encourages a reconstruction having the desired image characteristics. One example is motivated by compressed sensing (CS), which uses sparse regularization. The CS-based methods in [12, 5] are successful attempts at pan-sharpening in which LRMS image patches are assumed to have a sparse representation with respect to a patch-level dictionary. In these methods, a dictionary is randomly sampled (not learned) from either pre-obtained HRMS images or the PAN image itself. A limitation of these approaches is that important parameters such as the dictionary size and patch-specific sparsity are preset rather than learned from the data, and also that it leaves the dictionary quality to chance. Other dictionary learning algorithms use HRMS images to learn the dictionary off-line, but these are often unavailable in practice.

Motivated by the sparse representation perspective, in this paper we propose a pan-sharpening method that uses a beta-Bernoulli process as a Bayesian nonparametric prior for dictionary learning [18, 15]. Compared with the previous CS-based pan-sharpening methods, the proposed

model has the following unique properties:

1. It uses a Bayesian nonparametric dictionary learning approach [15, 21] to learn the dictionary directly on the reconstructed HRMS images of the current iteration for the problem at hand, as opposed to images obtained off-line. The model infers a sparse representation by learning the number of dictionary elements and the sparse coding of each patch.
2. We consider image properties not considered by other pan-sharpening approaches. We include the total variation (TV) as a penalty term for the pan-sharpened images and use the alternating direction method of multipliers (ADMM) to derive an efficient optimization procedure [4, 8]. We also consider a first derivative penalty for the PAN/HRMS fidelity term, which reduces the spectral distortion of the reconstructed HRMS images.

2 Pan-sharpening and dictionary learning

Images obtained by optical remote sensors contain both a high resolution panchromatic (i.e., black and white) image (PAN) and low resolution multispectral images (LRMS). The spectral bands typically consist of the colors red, green and blue, as well as a near infrared band. The goal is to fuse the color information of the LRMS images with the spatial information of the PAN image to obtain a high resolution multispectral image (HRMS).

In this paper, for concreteness we consider the case where the LRMS images are downsampled by a factor of four. We represent each of the $M \times N$ LRMS images in vectorized form as $y_b \in \mathbb{R}^{MN}$, where b indexes a spectral band (RGB and NIR). This observed data is treated as a noisy version of the corresponding $4M \times 4N$ HRMS images, which are vectorized as $x_b \in \mathbb{R}^{16MN}$. We follow the standard approach of assuming that the measured y_b follows from filtering x_b through a predefined and shared subsampling and blurring matrix $H \in \mathbb{R}^{MN \times 16MN}$ and adding noise,

$$y_b = Hx_b + \epsilon_b. \quad (1)$$

The monochromatic PAN image is usually considered to be a linear combination of the unknown HRMS images,

$$y_P = \sum_{b=1}^4 w_b x_b + \epsilon_P. \quad (2)$$

The vector $y_P \in \mathbb{R}^{16MN}$ represents the vectorized $4M \times 4N$ PAN image, ϵ_P is a Gaussian noise vector and the weights (w_1, \dots, w_4) are predefined.

The pan-sharpening problem is to use the measured data y_P, y_1, y_2, y_3, y_4 , along with settings for H and (w_1, \dots, w_4) to recover the underlying HRMS images x_1, x_2, x_3, x_4 via some predefined inversion model. For

Algorithm 1 Dictionary Learning with BPFA

-
1. Construct a dictionary $D = [d_1, \dots, d_k]$, with $d_k \sim \mathcal{N}(0, L^{-1}I_L)$ for each k .
 2. Draw probability π_k for each d_k , $\pi_k \sim \text{Beta}(c\gamma/K, c(1 - \gamma/K))$.
 3. For the i th patch in X :
 - (a) Draw the weight vector $s_i \sim \mathcal{N}(0, \gamma_s^{-1}I_K)$.
 - (b) Draw the binary vector $z_{ik} \sim \text{Bernoulli}(\pi_k)$.
 - (c) Define $\alpha_i = s_i \circ z_i$ by an element-wise product.
 - (d) Construct the patch $R_i X = D\alpha_i + \epsilon_i$ with noise $\epsilon_i \sim \mathcal{N}(0, \gamma_\epsilon^{-1}I_L)$.
 4. Construct the image using the average of all $R_i X$ that overlap on a given pixel.
-

this inversion, we propose using a Bayesian nonparametric (BNP) dictionary learning model based on the beta process called BPFA [15].

2.1 BNP dictionary learning for pan-sharpening

Sparse dictionary learning methods attempt to find a sparse basis for patches extracted from an image class of interest. The canonical dictionary learning approach is with the K-SVD model [2], which iterates between orthogonal matching pursuits and least squares to learn a matrix factorization of patches extracted from images. In this paper we instead focus on a BNP dictionary learning model. This model performs dictionary learning in a similar manner as K-SVD, in that it iterates between learning a weighted binary coefficient matrix and a regularized least squares dictionary. However in addition, Bayesian nonparametric methods provide the flexibility of allowing the data to determine the complexity of the model [9].

We learn a dictionary using the reconstructed HRMS images of the current iteration. Therefore, the “data” is slightly changing with each iteration. We extract $\sqrt{q} \times \sqrt{q}$ patches from these reconstructed HRMS images, one centered on each pixel, and put them in a matrix X . Let R_i be the i th patch extraction operator, which is a $q \times 16MN$ matrix containing a one in each row and zeros elsewhere. The i th column in the $4q \times 16MN$ matrix X is

$$X_i = [R_i x_1; R_i x_2; R_i x_3; R_i x_4] \in \mathbb{R}^{4q}.$$

A dictionary learning model learns a factorization of this matrix, $X = D\alpha + E$, where D is a $4q \times K$ dictionary, α is a $K \times 16MN$ coefficient matrix and E is noise.

We use the BPFA model shown in Algorithm 1 for dictionary learning from X . This model treats the columns of D as Gaussian vectors and α as being a weighted binary matrix, $S \circ Z$. The matrix S contains Gaussian random variables, and a row of Z is drawn from a Bernoulli process using a row-specific probability. This probability has a

beta distribution prior that is nonparametric through its parameterization; the parameters are set such that, for a large K , many of the probabilities will be so small that an entire row of Z will contain all zeros [18]. In this case the corresponding dictionary element is removed from the model. Therefore, unlike K-SVD which will learn a full dictionary of size K , the BPFA model will learn a dictionary that is smaller than the initial setting of K , and is in some sense “appropriate” for the data at hand because the actual size of the dictionary is determined by the data. As $K \rightarrow \infty$, this approximation converges to the beta process [10].

This model allows for learning an overcomplete dictionary with the size of the dictionary not being preset (assuming K is large enough). Additionally, conjugate gamma prior distributions can be placed on the noise precisions γ_s and γ_ϵ , which allows for these values to be inferred.

2.2 Total variation and high frequency constraints

We discuss two other image constraints that we consider for our reconstruction algorithm before presenting the final objective function and inference algorithm.

The first constraint is the total variation (TV) of the reconstructed HRMS images. TV regularization was introduced for image denoising by Rudin, Osher and Fatemi in [13], and has since evolved into a more general tool for solving a wide variety of image restoration problems, including deconvolution and inpainting. The isotropic TV norm of a vectorized image x is defined by

$$\|x\|_{TV} = \sum_i \|B_i x\|_2 = \sum_i \sqrt{|B_i^h x|^2 + |B_i^v x|^2}. \quad (3)$$

The vectors B_i^h and B_i^v are the rows of $B_i \in \mathbb{R}^{2 \times 16MN}$, which is a finite difference matrix that has two nonzero entries in each row corresponding to the partial derivatives of x at the pixel i along the horizontal and vertical directions. An efficient algorithm was explored by [8].

Another aspect of pan-sharpening algorithms is to use the PAN image for its high resolution spatial information. Motivated by Equation (2), most models do this by incorporating a squared error term for $y_P \approx \sum_b w_b x_b$ [3]. A drawback of this approach is that the weights w_b need significant tuning to minimize the spectral distortion that this tends to introduce in the learned x_b . In fact, the information we wish to exploit in the PAN image y_P is the high frequency edge details, while the low frequency (color) information is provided by the LRMS y_1, \dots, y_4 . Therefore, we propose using the squared error of the horizontal and vertical derivatives of the residual instead. This will enforce that $\sum_b w_b x_b$ agrees with y_P in the edge details only and allow for the spectral information in y_1, \dots, y_4 to inform the smooth regions of the corresponding HRMS images x_1, \dots, x_4 that we learn.

3 Bayesian nonparametric pan-sharpening

We combine the three elements of the model from Section 2 in the following algorithm for pan-sharpening.

3.1 A BNP pan-sharpening model

Let the set $\varphi_i = \{D, s_i, z_i, \gamma_\epsilon, \gamma_s, \pi\}$ denote the parameters for the BPFA dictionary learning model and $f(\varphi_i)$ be the log joint likelihood of the BPFA algorithm for all prior terms. We define the gradient matrix G_i , which is a $2 \times 16MN$ matrix of zeros except for a $+1$ and -1 in each row for the horizontal and vertical directions as defined by pixel i . The objective we seek to maximize is:

$$\mathcal{L}(x_1, x_2, x_3, x_4, \varphi) = \quad (4)$$

$$\begin{aligned} & \frac{v_1}{2} \sum_b \|Hx_b - y_b\|_2^2 + \frac{v_2}{2} \sum_i \|G_i(\sum_b w_b x_b - y_P)\|_2^2 \\ & + \sum_i \frac{\gamma_\epsilon}{2} \|R_i X - D\alpha_i\|_2^2 + f(\varphi_i) + \frac{\lambda_g}{2} \sum_b \|x_b\|_{TV}. \end{aligned}$$

The first row contains the squared error fidelity term for the multispectral information of each band b ; this enforces that the downsampling/blurring of our reconstruction x_b is in agreement with the observed y_b . The second term in the first row enforces that the gradient of the approximate PAN image from the reconstructions is in agreement with the gradient of the observed PAN image; this term is where the high resolution edge information from the PAN image is enforced. The second row contains the beta process prior on the stacked patches from each x_b and a TV penalty term. This attempts to learn HRMS reconstructions that have a sparse dictionary representation and little noise. The goal is to find a (local) optimal solution for \mathcal{L} with respect to the HRMS images (x_1, \dots, x_4) and the dictionary learning parameters φ .

3.2 An optimization algorithm

We next summarize our algorithm for approximately optimizing the non-convex objective function \mathcal{L} . Our algorithm incorporates regularized least squares, MCMC sampling and the alternating direction method of multipliers (ADMM) algorithm for TV minimization. ADMM is a general algorithmic approach to convex optimization [4]. For our model, ADMM works by performing dual ascent on the augmented Lagrangian objective function introduced for the total variation coefficients for each spectral band. Though our overall objective is not convex because of the dictionary learning terms, we note that when holding the dictionary learning parameters fixed, the resulting TV denoising problem for which we use ADMM is convex.

The ADMM algorithm entails modifying the objective function as follows: For the b th spectral band, we re-write

the TV coefficients for the i th pixel as $\beta_{i,b} := B_i x_b$. We introduce the vector of Lagrange multipliers $\eta_{i,b}$, and then split $\beta_{i,b}$ from $B_i x_b$ by relaxing the inequality via an augmented Lagrangian. This results in the following modification to the TV portion of the objective:

$$\begin{aligned}\|x_b\|_{TV} \rightarrow & \frac{\lambda_g}{2} \sum_i \|\beta_{i,b}\|_2 + \sum_i \eta_{i,b}^T (B_i x_b - \beta_{i,b}) \\ & + \frac{\rho}{2} \sum_i \|B_i x_b - \beta_{i,b}\|_2^2.\end{aligned}$$

From the ADMM theory, this objective will have optimal values $\beta_{i,b}^*$ and x_b^* with $\beta_{i,b}^* = B_i x_b^*$, and so the equality constraints will be satisfied. Following the standard convenient re-parameterization, we let $u_{i,b} := (1/\rho)\eta_{i,b}$ to absorb $u_{i,b}$ within the squared term [4, 8].

The resulting augmented objective function can be split into three separate sub-problems that are cycled through in each iteration. One problem is for TV, one for BPFA and one for reconstructing the HRMS:

$$\begin{aligned}P1 : \beta_{i,b}' &= \arg \min_{\beta} \frac{\lambda_g}{2} \|\beta_{i,b}\|_2 + \frac{\rho}{2} \|B_i x_b - \beta_{i,b} + u_{i,b}\|_2^2. \\ P2 : \varphi' &= \arg \min_{\varphi} \sum_i \frac{\gamma_\epsilon}{2} \|R_i X - D \alpha_i\|_2^2 + f(\varphi_i). \\ P3 : x_b' &= \arg \min_{x_b} \frac{\rho}{2} \|B_i x_b - \beta_{i,b}' + u_{i,b}\|_2^2 \\ &+ \sum_i \frac{\gamma_\epsilon}{2} \|R_i x_b - D'_b \alpha_i'\|_2^2 \\ &+ \frac{v_1}{2} \|H x_b - y_b\|_2^2 + \frac{v_2}{2} \sum_i \|G_i (\sum_b w_b x_b - Y_P)\|_2^2. \\ u_{i,b}' &= u_{i,b} + B_i x_b' - \beta_{i,b}', \quad i = 1, \dots, 16MN.\end{aligned}\tag{5}$$

We let D_b denote the part of the dictionary relevant to band b . (We recall that we combine the spectral bands to learn a dictionary on vectorized $\sqrt{q} \times \sqrt{q} \times 4$ blocks.) The update of the Lagrange multiplier $u_{i,b}$ follows from the ADMM algorithm [4].

For each sub-problem, we use the most recent values of all other parameters. Solutions for $P1$ and $P3$ are globally optimal and in closed form. Since $P2$ is non-convex, we cannot perform the desired minimization, and so an approximation is required. Furthermore, this problem requires iterating through several parameters, and so a local optimal solution cannot be given in closed form either. Our approach is to update variables for $P2$ by a combination of MCMC sampling and least squares. We next present the updates for these sub-problems.

3.2.1 P1 sub-problem: Total variation

We can solve for $\beta_{i,b}$ exactly for each pixel $i = 1, \dots, 16MN$ by using a generalized shrinkage operation,

$$\beta_{i,b}' = \max\{\|B_i x_b + u_{i,b}\|_2 - \frac{\lambda_g}{\rho}, 0\} \cdot \frac{B_i x_b + u_{i,b}}{\|B_i x_b + u_{i,b}\|_2}.$$

We recall that after updating x_b , we update the Lagrange multiplier for this part as $u_{i,b}' = u_{i,b} + B_i x_b' - \beta_{i,b}'$.

3.2.2 P2 sub-problem: BNP dictionary learning

We give the variable updates for BPFA in the appendix. We make one update for each variable in this step.

3.2.3 P3 sub-problem: HRMS image reconstructions

In this sub-problem we reconstruct each HRMS spectral band x_b for $b = 1, \dots, 4$. This is a least squares problem and has a closed form solution, but requires a special approach to make it computationally tractable.

For notational convenience, we define the spatial differences matrices $B = [B_1^T, \dots, B_{16MN}^T]^T$ and $G = [G_1^T, \dots, G_{16MN}^T]^T$. We also defined the stacked vectors $\beta_b = [\beta_{1,b}^T, \dots, \beta_{16MN,b}^T]^T$ and $u_b = [u_{1,b}^T, \dots, u_{16MN,b}^T]^T$. For band b , the least squares problem for x_b has a solution that satisfies

$$\begin{aligned}(\rho B^T B + \sum_i R_i^T R_i + v_1 H^T H + v_2 w_b^2 G^T G) x_b \\ = \rho B^T (\beta_b - u_b) + q x_b^{\text{BPFA}} + v_1 H^T y_b \\ + v_2 w_b G^T G (y_P - \sum_{m \neq b} w_m x_m).\end{aligned}\tag{6}$$

The vector $x_b^{\text{BPFA}} = (1/q) \sum_i R_i^T D_b \alpha_i$ is the reconstructed HRMS image for band b according to the BPFA dictionary learning model. We need to solve for x_b to get the reconstruction.

The left matrix is too large to invert. However, we can solve this problem exactly by working within the Fourier domain. Let $\theta^b = \mathcal{F} x_b$ be the Fourier transform of x_b . We replace x_b with $\mathcal{F}^T \theta^b$ and take the Fourier transform of each side of Equation (6). This diagonalizes the matrix on the left hand side, which can be seen as follows: The product of the finite difference operator matrix $B^T B$ yields a circulant matrix, which has the rows of the Fourier matrix \mathcal{F} as its eigenvectors. Therefore $\mathcal{F} B^T B \mathcal{F}^H$ diagonalizes $B^T B$ to give a matrix of eigenvalues Λ_1 . The matrix $G^T G$ is the same as $B^T B$ since it is also a spatial finite difference matrix, and so has the same eigenvalues. $H^T H$ is a different circulant matrix and so has unique eigenvalues Λ_2 , though the same Fourier basis as eigenvectors. The matrix $\sum_i R_i^T R_i = qI$ and so the product of Fourier matrices cancel.

Table 1: Quality metrics of different methods on IKONOS images from Figure 1.

	RMSE	CC	WB	ERGAS	Q4
Brovey	0.120	0.851	0.599	10.07	0.776
Wavelet	0.079	0.898	0.618	6.76	0.760
PCA	0.140	0.601	0.350	12.08	0.600
FIHS	0.156	0.577	0.429	13.99	0.537
Adaptive IHS	0.077	0.916	0.666	6.39	0.761
BPFA	0.074	0.921	0.697	6.13	0.794
BPFA+TV	0.071	0.923	0.686	5.95	0.814
BPFA w/o G_i	0.079	0.917	0.676	6.22	0.778

This reduces the problem to solving $16MN$ one dimensional problems in the Fourier domain for each band:

$$\begin{aligned} \theta_i^b = & \mathcal{F}_i(\rho B^T(\beta_b - u_b) + qx_b^{\text{BPFA}} + v_1 H^T y_b \\ & + v_2 w_b G^T G(y_P - \sum_{m \neq b} w_m x_m)) / \\ & (q + (\rho + v_2 w_b^2) \Lambda_{1,i} + v_1 \Lambda_{2,i}). \quad (7) \end{aligned}$$

We then invert θ^b using the inverse Fourier transform to obtain the reconstruction x_b for spectral band b .

4 Experiments

In this section, we evaluate our proposed models, denoted BPFA+TV and BPFA (without TV), on three remote sensing images. The data sets are as follows:

1. IKONOS. This dataset consists of a 512×512 PAN image and low resolution multispectral images of dimension 128×128 pixels each. The MS images have four bands, RGB and near infrared (NIR).
2. Pléiades. The data set from the Pléiades satellite contains a 256×256 PAN image and four LRMS images of 64×64 pixels each.
3. QuickBird. In this experiment we have a 512×512 PAN image and four 128×128 LRMS images.

For all experiments we extract $4 \times 4 \times 4$ patches from the reconstructed HRMS images of the current iteration, for a dictionary of 64 dimensions. We initialize the dictionary size to 256. We also set $w_1 = 0.25$, $w_2 = 0.25$, $w_3 = 0.25$, $w_4 = 0.25$ and $\lambda_g = 0.1$. The BPFA hyperparameters $c, \gamma, e_0, f_0, g_0, h_0$ are all set to one.

We compare the performance with other existing fusion methods, namely, model-based fusion [1], Brovey transform [7], the fast IHS (FIHS) [19], a PCA-based method [17], wavelet-based image fusion [14] and the adaptive IHS method [16]. Since there are no ground truth HRMS images for IKONOS and Pléiades, to compare these algorithms we re-project the fused images and compare with

Table 2: Quality metrics of different methods on Pléiades images from Figure 2.

	RMSE	CC	WB	ERGAS	Q4
Brovey	0.274	0.791	0.235	65.522	0.252
Wavelet	0.110	0.877	0.754	8.304	0.794
PCA	0.144	0.830	0.537	11.056	0.635
FIHS	0.133	0.790	0.642	10.503	0.666
Adaptive IHS	0.109	0.901	0.754	8.223	0.771
BPFA	0.102	0.898	0.812	7.78	0.835
BPFA+TV	0.096	0.906	0.814	7.23	0.836
BPFA w/o G_i	0.103	0.897	0.807	7.758	0.836

Table 3: Quality metrics of different methods on QuickBird images from Figure 3.

	RMSE	CC	WB	ERGAS	Q4
Brovey	0.133	0.924	0.695	6.388	0.761
Wavelet	0.069	0.939	0.646	4.625	0.757
PCA	0.247	0.608	0.422	17.295	0.586
FIHS	0.161	0.655	0.507	11.410	0.564
Adaptive IHS	0.058	0.958	0.720	3.843	0.811
BPFA	0.059	0.965	0.733	3.609	0.821
BPFA+TV	0.053	0.967	0.745	3.453	0.824
BPFA w/o G_i	0.057	0.967	0.731	3.523	0.749

the ground truth LRMS images. For the QuickBird experiment we have the ground truth HRMS images. After some parameter tuning, we set $v_1 = 50$, $v_2 = 70$ and the ADMM parameter $\rho = 3$ for the IKONOS image and set $v_1 = 10$, $v_2 = 30$ and the ADMM parameter $\rho = 5$ for the Pléiades image for the BPFA+TV pan-sharpening model. For the QuickBird image we set $v_1 = 10$, $v_2 = 30$ and the ADMM parameter $\rho = 20$ for BPFA+TV pan-sharpening model.

4.1 Reconstruction results

The pan-sharpening results for the different algorithms are shown in Figure 1 for IKONOS, in Figure 2 for Pléiades and Figure 3 for QuickBird. By visually comparing the fusion results, we can see that while several algorithms have equally clear edges, they exhibit significant spectral distortion. For example, the red blocks in the IKONOS image are much brighter than the original LRMS images for the PCA-based model and several algorithms have spectral distortion in the vegetation areas for the Pléiades image. The model-based method is more blurry than BPFA+TV, and the wavelet-based results exhibit a stair-casing effect. The BPFA results do not appear to be as sharp as some other algorithms, such as adaptive IHS. This may be because the dictionary learning algorithm, which averages 16 values for each pixel, is slightly “denoising” these edges.

In pan-sharpening, the spectral quality is considered as important as the spatial quality, but is more difficult to judge visually – especially considering that there are bands in the non-visible spectrum. Therefore we quantitatively evaluate the reconstructions using the following standard pan-

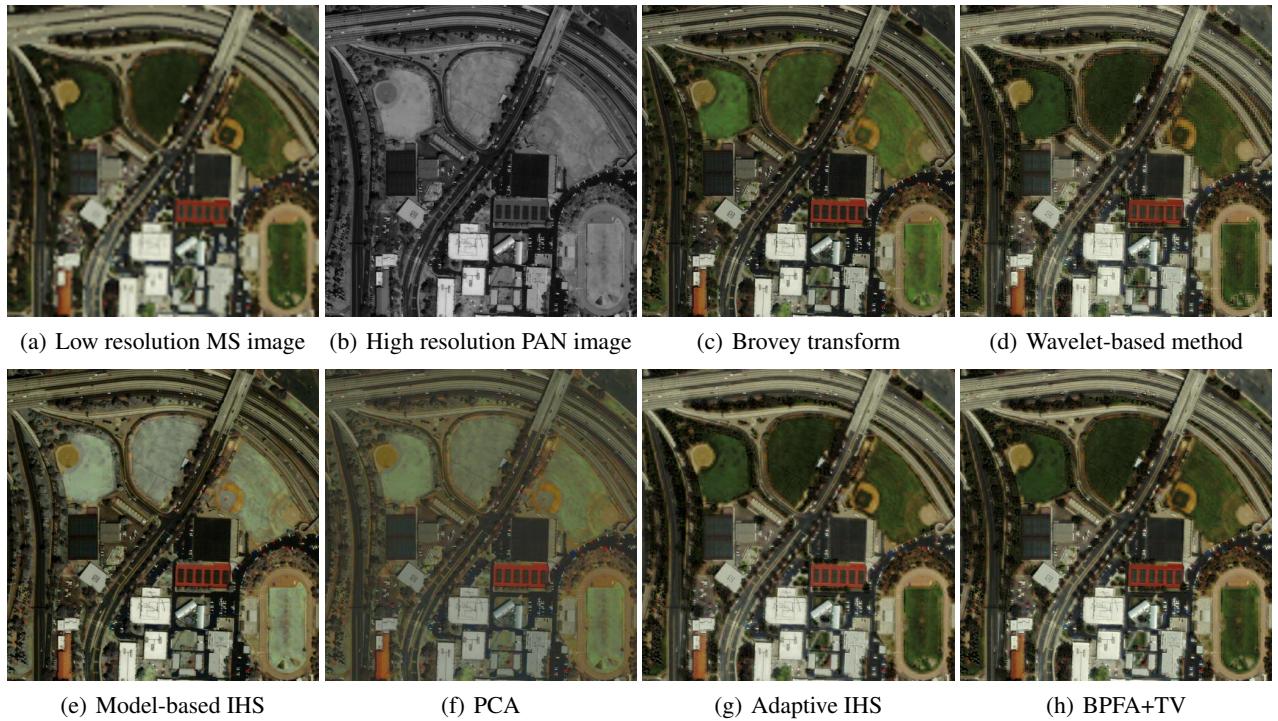


Figure 1: IKONOS images and pan-sharpening results for different methods.

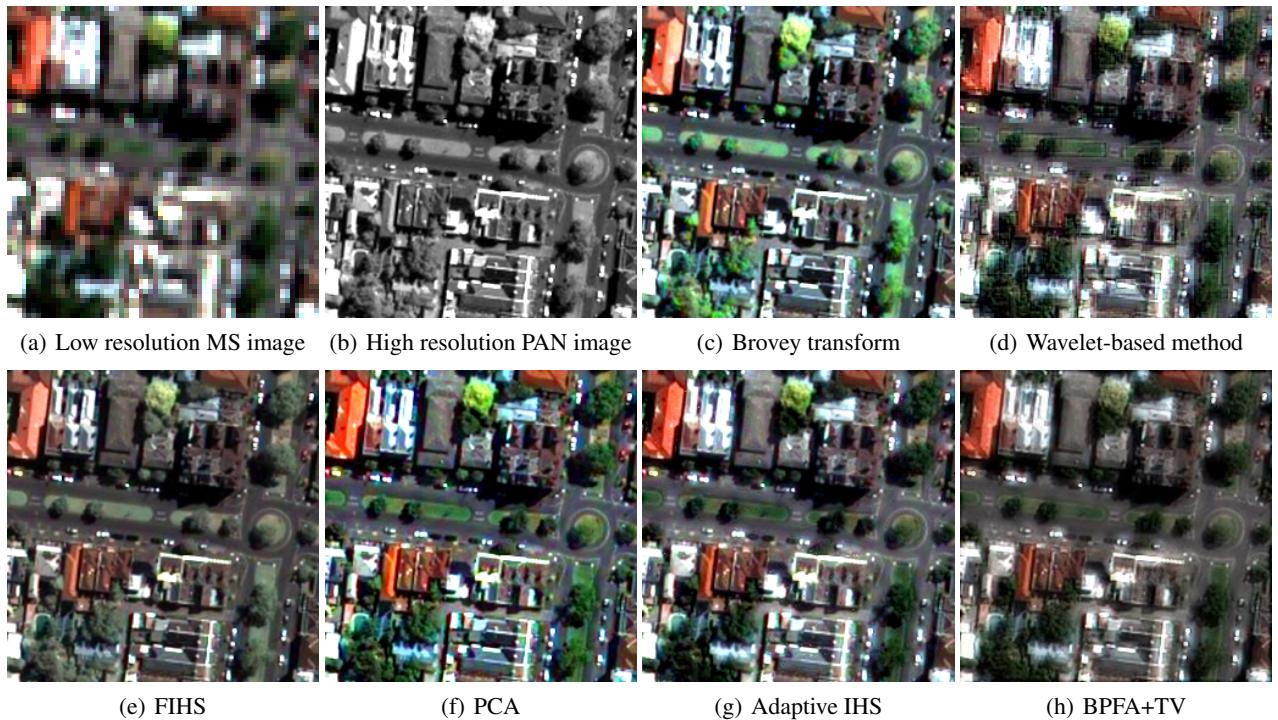


Figure 2: Pléiades images and pan-sharpening results for different methods.

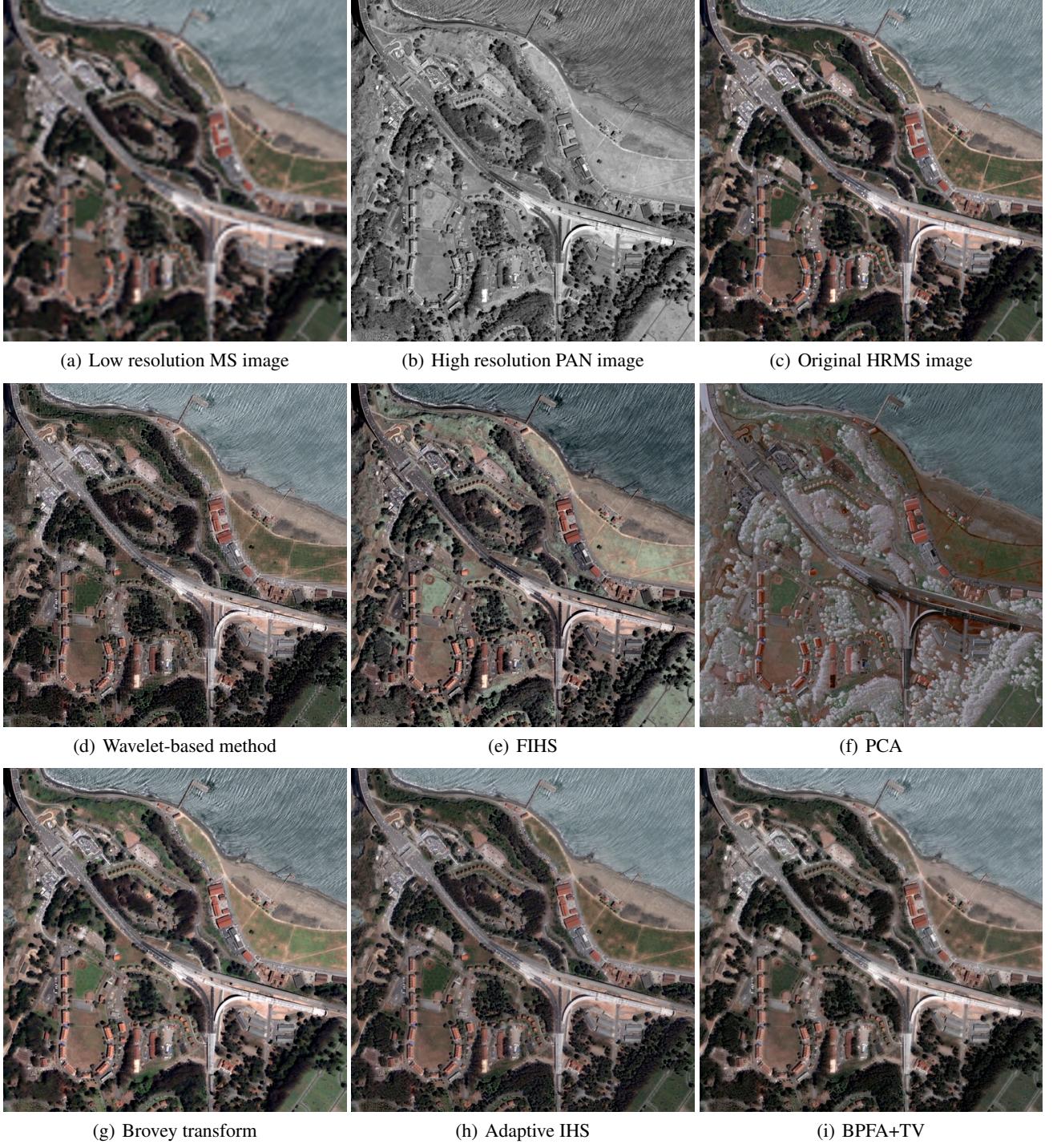


Figure 3: QuickBird images and pan-sharpening results for different methods.

sharpening quality metrics [20]: Root mean square error (RMSE), Correlation coefficient (CC), Wang-Bovik measure (WB), ERGAS and Q4. We show quantitative results in Tables 1-3 for each image, where we highlight the best result in bold. From these tables, we see that BPFA and BPFA+TV perform very competitively compared with the other methods in its ability to accurately learn the spectral

content of the HRMS images. We also show results without using the first derivative penalty, that is, replacing G_i with the identity matrix when penalizing the PAN reconstruction. We can see an advantage to our method of only focusing on the high frequency edge information within the PAN image and allowing the spectral information in the smooth regions to only be determined by the LRMS images. We

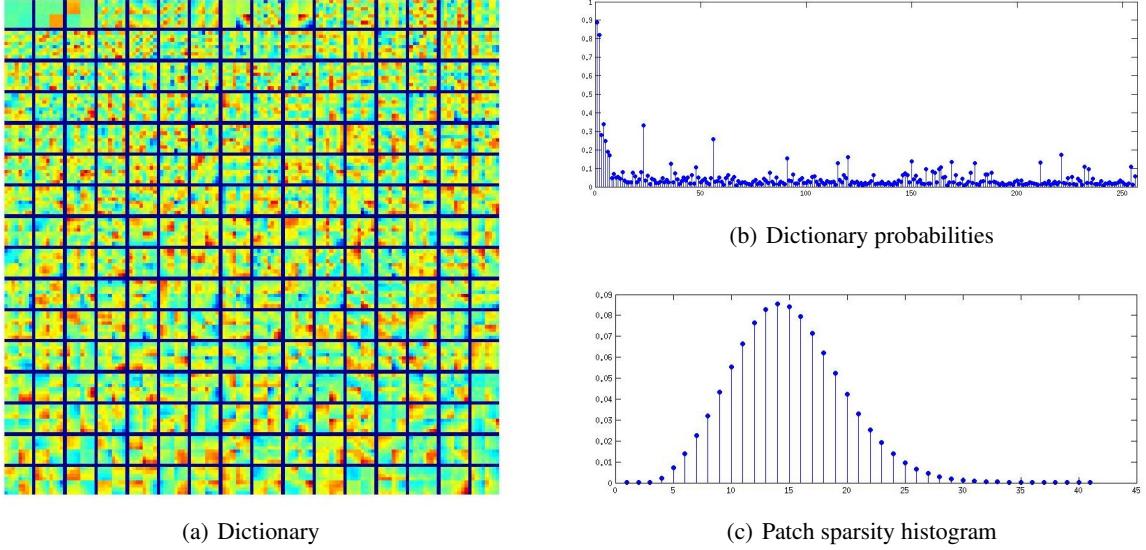


Figure 4: Dictionary learning results for QuickBird.

also notice that in general penalizing the total variation has a positive impact.

4.2 BNP dictionary learning results

We show example dictionary learning results for the QuickBird image in Figure 4. In Figure 4(a), we show the $4 \times 4 \times 4$ dictionary as 8×8 patches by putting the dictionary elements along the third dimension into quadrants within a cell. In Figure 4(b) we show a sample of the dictionary probabilities, where we can see that most of the dictionary elements have very small probability. Roughly speaking, about 50 of the 256 dictionary elements have significant probability. Figure 4(c) shows a histogram of the patch sparsity for one sample. Here we see that the patches use a variable number of dictionary elements, which is determined probabilistically through the sampler. Also, each patch has a low rank representation using the learned dictionary since each patch has 64 dimensions and on average 15 basis functions are needed per patch.

5 Conclusion

We have proposed a new dictionary learning approach to pan-sharpening that uses a Bayesian nonparametric model to learn a sparse patch-level basis. We also considered a total variation penalty for reconstruction, and a gradient penalty on the PAN image to enforce high frequency fidelity while minimizing spectral distortion. Experimental results on three images showed that the proposed model improves the spectral content of the reconstructions over other commonly used methods. This suggests that our algorithm has potential uses for other problems where spectral content is important, such as hyperspectral imaging [21].

Appendix: An algorithm for BPFA

a) Update D : The least squares update for the dictionary is

$$D = X\alpha^T(\alpha\alpha^T + (L/\gamma_\epsilon)I_L)^{-1}. \quad (8)$$

The matrix X is constructed from the patches of the HRMS images of the current iteration.

b) Update α_i : From Algorithm 1, $\alpha_i := z_{ik}s_{ik}$. We update z_{ik} first stochastically by sampling from a Bernoulli distribution with probability

$$p_{ik} \propto \pi_k \sqrt{1 + \frac{\gamma_\epsilon}{\gamma_s} d_k^T d_k} \exp \left\{ \frac{\gamma_\epsilon}{2} \frac{(d_k^T r_{i,-k})^2}{\gamma_s/\gamma_\epsilon + d_k^T d_k} \right\},$$

$$1 - p_{ik} \propto 1 - \pi_k, \quad (9)$$

where $r_{i,-k}$ is the residual error in approximating the i th patch ignoring the k th dictionary element. The corresponding least squares weight s_{ik} is

$$s_{ik} = z_{ik} d_k^T r_{i,-k} / (\gamma_s/\gamma_\epsilon + d_k^T d_k), \quad (10)$$

and we then calculate the weight vector, $\alpha_i = s_i \circ z_i$.

c) Update γ_ϵ : The least squares solution is

$$\gamma_\epsilon = (2g_0 + LN_1) / (2h_0 + \sum_i \|R_i X_0 - D\alpha_i\|_2^2). \quad (11)$$

We observe that this is approximately the inverse of the empirical squared error.

d) Update π_k : We update π_k stochastically via Gibbs sampling from a beta distribution as follows,

$$\pi_k \sim Beta(a_0 + \sum_i z_{ik}, b_0 + \sum_i (1 - z_{ik})). \quad (12)$$

We observe that π_k will be close to the empirical fraction of times that dictionary element d_k is used by the patches for the previous iteration.

Acknowledgements

The project is supported by the National Natural Science Foundation of China (No. 30900328, 61172179, 61103121, 81301278).

References

- [1] H. Aanaes and J. Sveinsson, "Model-based satellite image fusion," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1336-1346, 2008.
- [2] M. Aharon, M. Elad, A. Bruckstein, and Y. Katz, "K-SVD: An Algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Processing*, vol. 54, pp. 4311-4322, 2006.
- [3] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pan-sharpening algorithms: Outcome of the 2006 GRS-S data fusion contest," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3012-3021, 2007.
- [4] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1-122, 2010.
- [5] J. Cheng, H. Zhang, H. Shen, and L. Zhang, "A practical compressed sensing-based pan-sharpening method," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 4, pp. 629-633, 2012.
- [6] X. Ding, L. He, and L. Carin, "Bayesian robust principal component analysis," *IEEE Trans. on Image Processing*, vol. 20, no. 12, pp. 3419-3430, 2011.
- [7] D. Fasbender, J. Radoux, and P. Bogaert, "Bayesian data fusion for adaptable image pan-sharpening," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 46, no. 6, pp. 1847-1857, 2008.
- [8] T. Goldstein and S. Osher, "The Split Bregman Method for L1 Regularized Problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323-343, 2009.
- [9] N.L. Hjort, C. Holmes, P. Müller, S.G. Walker (editors), "Bayesian Nonparametrics," *Cambridge Series in Statistical and Probabilistic Mathematics*, 2010.
- [10] N.L. Hjort, "Nonparametric Bayes estimators based on beta processes in models for life history data," *Annals of Statistics*, vol. 18, pp. 1259-1294, 1990.
- [11] M. V. Joshi, L. Bruzzone, and S. Chaudhuri, "A model-based approach to multiresolution fusion in remotely sensed images," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2549-2562, 2006.
- [12] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 49, no. 2, pp. 738-746, 2011.
- [13] S. Osher, L. Rudin and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D.*, vol. 60, pp. 259-268, 1992.
- [14] X. Otazu and M. Gonzalez-Ausicana, "Introduction of sensor spectral response into image fusion methods: Application to wavelet-based methods," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no. 10, pp. 2376-2385, 2005.
- [15] J. Paisley and L. Carin, "Nonparametric factor analysis with beta process priors," in *International Conference on Machine Learning*, 2009.
- [16] S. Rahmani, M. Strait, D. Merkurjev, M. Moeller, and T. Wittman, "An adaptive IHS pan-sharpening method," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 4, pp. 746-750, 2010.
- [17] V. P. Shah and N. H. Younan, "An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1323-1335, 2008.
- [18] R. Thibaux and M. Jordan, "Hierarchical beta processes and the Indian buffet process," in *International Conference on Artificial Intelligence and Statistics*, 2007.
- [19] T. M. Tu, P. S. Huang, C. L. Hung, and C. P. Chang, "A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 4, pp. 309-312, 2004.
- [20] Z. Wang and A.C. Bovik, "A universal image quality index," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 9, no. 3, pp. 81-84, 2002.
- [21] M. Zhou, H. Chen, J. Paisley, L. Ren, L. Li, Z. Xing, D. Dunson, G. Sapiro and L. Carin, "Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images," *IEEE Trans. on Image Processing*, vol. 21, no. 1, pp. 130-144, 2012.