

# SPIKE AND SLAB VARIATIONAL INFERENCE FOR BLIND IMAGE DECONVOLUTION

Juan G. Serra\*, Javier Mateos\*, Rafael Molina\*, and Aggelos K. Katsaggelos†

\* University of Granada, Dept. of Computer Science and Artificial Intelligence.

† Northwestern University, Dept. of Electrical Engineering and Computer Science.

## ABSTRACT

In this work, we propose a new variational blind deconvolution method for spike and slab prior models. Soft-sparse or shrinkage priors such as the Laplace and other related Gaussian Scale Mixture priors may not be ideal sparsity promoting priors. They assign zero probability mass to events we may be interested in assigning a probability greater than zero. The truly sparse nature of the spike and slab priors allows us to discard irrelevant information in the blur estimation process, resulting in improved performance. We present an efficient inference algorithm to estimate the unknown blur kernel in the filter space, from which we estimate the final deblurred image. The VB approach we propose in this paper handles the inference in a much more efficient way than MCMC, and is more accurate than the standard mean field variational approximation. We prove the efficacy of our method by means of a series of experiments on both synthetically generated and real images.

**Index Terms**— Blind deconvolution, spike-and-slab, variational Bayesian approach.

## 1. INTRODUCTION

Blind image deconvolution (BID) is an image restoration problem where both the blur and the original image are unknown and have to be estimated from the degraded observed image. This is a very challenging ill-posed problem where small variations in the estimated blur result in large variations in the restored image. Also, it is an underdetermined nonlinear inverse problem, which requires the estimation of more unknowns than available observations. To obtain good image and blur estimations, prior knowledge about the unknowns and sound estimation procedures are needed.

It is a well known fact that when high-pass filters are applied to natural images, the resulting coefficients are sparse; i.e., most of the coefficients are zero or very small while only a small number of them are large (e.g., at the edges). BID methods can be formulated in either the image or the filter space. Several authors have discussed the advantages of using one space over the other [1–3]. The formulation on the image space appears to be less sensitive to noise since noise is more noticeable in the filter space. However, filter space methods have access to more pseudo-observations to estimate the blur. Notice that images suited to blur estimation only should have many step edges and not many details. Once the blur is estimated in the filter space, a non-blind deconvolution algorithm is used to recover the sharp image from the observed image and the estimated blur. If the image space is used, both the image and blur can be estimated simultaneously [4].

This work was supported in part by the Ministerio de Economía y Competitividad under contracts TIN2013-43880-R and DPI2016-77869-C2-2-R, and the US Department of Energy grant DE-NA0002520.

Variational Bayes (VB) is a sound and well grounded probabilistic estimation approach to BID problems. VB BID methods, like the vast majority of state-of-the-art BID methods, rely on the use of sparsity promoting image models. See [3, 5] for reviews on variational BID methods and image models. Since the work [6], where a mixture of Gaussians (MoG) is used to impose sparsity and VB is utilized to perform inference, the interest on sparse priors has increased. Babacan *et al.* [7] propose a general Bayesian framework based on Super Gaussian (SG) and Scale Mixture of Gaussian (SMG) priors for BID. The framework includes, among others, [6] as a particular case. Notice that popular sparse prior models, such as TV,  $\ell_p$ , MoG, and Student-t [8] are included in the proposed framework. The work in [7] has been extended in [9] to handle Huber Super Gaussian (HSG) priors, which solve the problem of lack of differentiability around zero of most common SG priors. Power Exponential Scale Mixture (PESM) models have been recently introduced, see [10, 11]. They use a mixture of exponential distributions (being the Laplace distribution a typical one) to represent a sparse signal. It is interesting to note that while most of the BID research concentrates on image modelling, work has also been carried out to force the posterior blur distribution to be a member of a particular class of probability distributions, see, for instance, [12].

As indicated in [13], soft-sparse or shrinkage priors such as the Laplace and other related GSM priors may not be ideal sparsity promoting priors. They assign zero probability mass to events we may be interested in assigning a probability greater than zero. For instance, to obtain better blur estimates in BID and get rid of noise that compromises the estimation procedure we may want to assign non-zero probability to a zero output in the filter space. Priors that combine Bernoulli and continuous distributions are starting to be used to better approximate  $\ell_0$  penalization [14, 15]. Spike-and-slab priors [16], also named Bernoulli-Gaussian priors [15] since they are a mixture of a Bernoulli and a Gaussian distribution, are the gold standard in sparse machine learning. They have the ability of selectively shrink irrelevant variables while relevant variables are mildly regularized [17]. Applications of this prior include variable selection [18, 19], denoising [13, 20], inpainting [13], unsupervised learning sparse features [21], hyper-spectral image fusion [22] and sparse signal recovery [15]. Notice that the spike-and-slab prior can be approximated as the mixture of two Gaussians, one very peaky (the spike) and another with very high variance (the slab) [23] but this is still a mixture of two continuous distributions. According to [24] Spike-and-slab models are more effective than other sparse priors (Laplacian or Student-t priors, for instance) in enforcing sparsity and, also, the degree of sparsity can be directly adjusted by modifying the weight of the spike in the mixture .

Unfortunately, due to the form of the prior, Bayesian inference for spike and slab models is a very challenging task. The exact posterior can not be calculated. Since classical mean field variational

inference removes essential dependencies in the posterior distribution approximation, until recently, costly Monte-Carlo Markov Chain (MCMC) sampling was the usual way to perform inference. The work by Titsias *et al.* [13], proposes an alternative VB inference model to approximate the posterior distribution using a simple and efficient algorithm. Instead of using a unimodal variational distribution, the authors propose an alternative approximation that more accurately matches the combinatorial nature of the posterior distribution over the spike-and-slab weights, see also [24] for the use of Expectation Propagation for posterior approximation.

In this paper, we formulate the BID problem in the filter space. Then, we introduce a spike-and-slab prior to model our knowledge on the original image in that space. This will allow us to distinguish relevant observations for blur estimation from noisy ones, discarding the latter, and making the blur estimation more robust and accurate. The VB approach we propose in this paper handles inference in a much more efficient way than MCMC, and is more accurate than the standard mean field variational approximation.

The rest of the paper is organized as follows. Section 2 describes the proposed model for blur estimation. In Sec. 3, Bayesian inference is performed and, in Sec. 4, a BID algorithm is synthesized. The performance of the proposed method is assessed in Section 5. Finally, Section 6 concludes the paper.

## 2. BAYESIAN MODELING OF THE BLUR ESTIMATION PROBLEM

In BID, the degraded observed image  $\mathbf{y}$  is modeled as [3]

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n}, \quad (1)$$

where  $\mathbf{x}$  and  $\mathbf{n}$  are the original image and noise respectively, and  $\mathbf{H}$  is the unknown convolution matrix whose row elements are obtained from the blur kernel  $\mathbf{h}$ . We utilize a column vector notation for the observed and original images and the noise. We also assume that their size is  $N$ .

The BID problem is formulated here in the filter space. We create  $L$  pseudo-observations  $\mathbf{y}_\gamma$  by applying high-pass filters  $\{\mathbf{f}_\gamma\}_{\gamma=1}^L$  to the blurred and noisy image  $\mathbf{y}$  obtaining

$$\mathbf{y}_\gamma = \mathbf{F}_\gamma \mathbf{y} = \mathbf{H}\mathbf{F}_\gamma \mathbf{x} + \mathbf{F}_\gamma \mathbf{n} = \mathbf{H}\mathbf{x}_\gamma + \mathbf{n}_\gamma, \quad (2)$$

with  $\mathbf{x}_\gamma = \mathbf{F}_\gamma \mathbf{x}$  and  $\mathbf{F}_\gamma$  is the convolution matrix associated with the filter  $\mathbf{f}_\gamma$ .

Assuming that the pseudo-observations are independent, i.i.d. Gaussian noise, that is,  $\mathbf{n}_\gamma \sim \mathcal{N}(\mathbf{0}, \beta_\gamma^{-1} \mathbf{I})$ , and denoting  $\mathbf{y}_\Gamma = \{\mathbf{y}_1, \dots, \mathbf{y}_L\}$ ,  $\mathbf{x}_\Gamma = \{\mathbf{x}_1, \dots, \mathbf{x}_L\}$ , and  $\boldsymbol{\beta}_\Gamma = \{\beta_1, \dots, \beta_L\}$ , we can write

$$p(\mathbf{y}_\Gamma | \mathbf{h}, \mathbf{x}_\Gamma, \boldsymbol{\beta}_\Gamma) = \prod_\gamma p(\mathbf{y}_\gamma | \mathbf{h}, \mathbf{x}_\gamma, \beta_\gamma), \quad (3)$$

where  $p(\mathbf{y}_\gamma | \mathbf{h}, \mathbf{x}_\gamma, \beta_\gamma) = \mathcal{N}(\mathbf{y}_\gamma | \mathbf{H}\mathbf{x}_\gamma, \beta_\gamma \mathbf{I})$ .

Notice that  $\{\mathbf{x}_\gamma\}_{\gamma=1}^L$  are sparse since they represent high-pass filtered instances of the original image. Therefore, to impose sparsity on the solution, we define a spike-and-slab prior on the value of each pixel  $x_{\gamma i}$  of  $\mathbf{x}_\gamma$ ,

$$p(x_{\gamma i} | \alpha_\gamma, \pi_\gamma) = \pi_\gamma \mathcal{N}(x_{\gamma i} | 0, \alpha_\gamma^{-1}) + (1 - \pi_\gamma) \delta(x_{\gamma i}), \quad (4)$$

where  $\delta(x_{\gamma i})$  denotes the Dirac delta function centered at zero. Observe that this is a truly sparse prior, with probability  $(1 - \pi_\gamma)$ ,  $x_{\gamma i}$  is exactly zero.

To investigate the model capability, we assume here that  $\pi_\gamma$  and  $\alpha_\gamma$  are known and postpone their estimation to future work.

Notice that these parameters determine the amount of sparsity and the variability of nonzero values, respectively.

Let us now write, following [13],  $x_{\gamma i}$  as the product of a Gaussian random variable  $\tilde{x}_{\gamma i} \sim \mathcal{N}(\tilde{x}_{\gamma i} | 0, \alpha_\gamma^{-1})$  and a Bernoulli random variable  $s_{\gamma i} \sim \pi_\gamma^{s_{\gamma i}} (1 - \pi_\gamma)^{1-s_{\gamma i}}$ , that is,

$$x_{\gamma i} = s_{\gamma i} \tilde{x}_{\gamma i}, \quad (5)$$

and redefine the prior on the two components of  $x_{\gamma i}$ , as

$$p(\tilde{x}_{\gamma i}, s_{\gamma i} | \alpha_\gamma, \pi_\gamma) = \mathcal{N}(\tilde{x}_{\gamma i} | 0, \alpha_\gamma^{-1}) \pi_\gamma^{s_{\gamma i}} (1 - \pi_\gamma)^{1-s_{\gamma i}}, \quad (6)$$

where  $s_{\gamma i} \in \{0, 1\}$ . We use the notation  $\tilde{\mathbf{x}}_\Gamma = \{\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_L\}$  and  $\mathbf{s}_\Gamma = \{\mathbf{s}_1, \dots, \mathbf{s}_L\}$ . With all the above, utilizing a flat prior on  $\mathbf{h}$ , and denoting the whole set of unknowns by  $\Theta = \{\mathbf{h}, \tilde{\mathbf{x}}_\Gamma, \mathbf{s}_\Gamma\}$ , we can write the joint distribution as

$$\begin{aligned} p(\Theta, \mathbf{y}_\Gamma) = & p(\mathbf{h}) \prod_\gamma p(\mathbf{y}_\gamma | \mathbf{h}, \tilde{\mathbf{x}}_\gamma, \mathbf{s}_\gamma, \beta_\gamma) \\ & \times \prod_\gamma \prod_i p(\tilde{x}_{\gamma i}, s_{\gamma i} | \alpha_\gamma, \pi_\gamma). \end{aligned} \quad (7)$$

## 3. VARIATIONAL INFERENCE FOR BLUR ESTIMATION

Since  $p(\Theta | \mathbf{y}_\Gamma)$  cannot be calculated in closed form, the standard mean field approximation [25] that factorizes  $q(\tilde{\mathbf{x}}_\gamma, \mathbf{s}_\gamma) = q(\tilde{\mathbf{x}}_\gamma)q(\mathbf{s}_\gamma)$  could be used. However this is a unimodal distribution [13] and, therefore, not a good approximation of the true posterior distribution. Since the pairs  $\{\tilde{x}_{\gamma i}, s_{\gamma i}\}$  are strongly correlated (remind that  $x_{\gamma i} = s_{\gamma i} \tilde{x}_{\gamma i}$ ), we treat them as a unit, use the factorization

$$q(\tilde{\mathbf{x}}_\gamma, \mathbf{s}_\gamma) = \prod_i q(\tilde{x}_{\gamma i}, s_{\gamma i}) \quad (8)$$

and utilize the following mean field approximation

$$q(\Theta) = q(\mathbf{h}) \prod_\gamma \prod_i q(\tilde{x}_{\gamma i}, s_{\gamma i}). \quad (9)$$

The distribution approximating the posterior  $p(\Theta | \mathbf{y}_\Gamma)$  is found by minimizing the Kullback-Leibler divergence assuming that  $q(\mathbf{h})$  is a degenerate distribution and that the blur is non-negative and its coefficients add up to one.

We now present how inference on  $q(\tilde{x}_{\gamma i}, s_{\gamma i})$  and  $q(\mathbf{h})$  is performed.

### 3.1. Obtaining $q(\tilde{x}_{\gamma i}, s_{\gamma i})$

Using the Kullback-Leibler criterion and the mean field approximation in Eq. (9), we have

$$\begin{aligned} q(\tilde{x}_{\gamma i}, s_{\gamma i}) = & \frac{1}{Z} \exp \left[ \langle \ln p(\mathbf{y}_\gamma | \mathbf{h}, \tilde{\mathbf{x}}_\gamma, \mathbf{s}_\gamma, \beta_\gamma) \rangle_{q(\Theta_{\tilde{x}_{\gamma i}, s_{\gamma i}})} \right] \\ & \times \mathcal{N}(\tilde{x}_{\gamma i} | 0, \alpha_\gamma^{-1}) \pi_\gamma^{s_{\gamma i}} (1 - \pi_\gamma)^{1-s_{\gamma i}}, \end{aligned} \quad (10)$$

where  $Z$  is the partition function,  $\theta \in \Theta$ , and  $\Theta_\theta$  denotes  $\Theta$  with  $\theta$  removed. To compute the explicit expression for the above posterior we need to separate the derivations for  $s_{\gamma i} = 1$  and  $s_{\gamma i} = 0$ .

First, we find the marginal  $q(s_{\gamma i} = 1)$ . Particularizing Eq. (10) to  $s_{\gamma i} = 1$  and integrating on  $\tilde{x}_{\gamma i}$ , we obtain

$$\begin{aligned} q(s_{\gamma i} = 1) = & \int q(\tilde{x}_{\gamma i}, 1) d\tilde{x}_{\gamma i} = \frac{1}{Z} (2\pi)^{-N/2} \beta_\gamma^{N/2} \alpha_\gamma^{1/2} \pi_\gamma \rho_\gamma^{-1/2} \\ & \times \exp \left[ -\frac{\beta_\gamma}{2} \| \mathbf{y}_\gamma - \sum_{k \neq i} \langle s_{\gamma k} \tilde{x}_{\gamma k} \rangle \mathbf{h}_k \|^2 \right] \\ & \times \exp \left[ \frac{\beta_\gamma^2}{2\rho_\gamma} (\mathbf{h}_i^\top (\mathbf{y}_\gamma - \sum_{k \neq i} \langle s_{\gamma k} \tilde{x}_{\gamma k} \rangle \mathbf{h}_k))^2 \right], \end{aligned} \quad (11)$$

where  $\mathbf{h}_i$  denotes the  $i$ th column of  $\mathbf{H}$  and

$$\rho_\gamma = \beta_\gamma \|\mathbf{h}\|^2 + \alpha_\gamma, \quad (12)$$

with  $\|\mathbf{h}\|^2 = \mathbf{h}_i^T \mathbf{h}_i, \forall i$  since spatially invariant blur is assumed.

Following an analogous procedure for  $s_{\gamma i} = 0$ , we have

$$\begin{aligned} q(s_{\gamma i} = 0) &= \int q(\tilde{x}_{\gamma i}, 0) d\tilde{x}_{\gamma i} = \frac{1}{Z} (1 - \pi_\gamma) (2\pi)^{-N/2} \\ &\times \beta_\gamma^{N/2} \exp \left[ -\frac{\beta_\gamma}{2} \|\mathbf{y}_\gamma - \sum_{k \neq i} \langle s_{\gamma k} \tilde{x}_{\gamma k} \rangle \mathbf{h}_k\|^2 \right]. \end{aligned} \quad (13)$$

Defining now

$$\omega_{\gamma i} = q(s_{\gamma i} = 1) = \frac{1}{1 + e^{-u_{\gamma i}}}, \quad (14)$$

it is easy to show that

$$\begin{aligned} u_{\gamma i} &= \ln q(s_{\gamma i} = 1) - \ln q(s_{\gamma i} = 0) = \ln \frac{\pi_\gamma}{1 - \pi_\gamma} + \frac{1}{2} \ln \alpha_\gamma \\ &- \frac{1}{2} \ln(\rho_\gamma) + \frac{\beta_\gamma^2}{2\rho_\gamma} (\mathbf{h}_i^T (\mathbf{y}_\gamma - \sum_{k \neq i} \langle s_{\gamma k} \tilde{x}_{\gamma k} \rangle \mathbf{h}_k))^2. \end{aligned} \quad (15)$$

We now calculate the conditional distributions  $q(\tilde{x}_{\gamma i} | s_{\gamma i})$  from Eq. (10). It can be shown that they are both Gaussians of the form

$$q(\tilde{x}_{\gamma i} | s_{\gamma i} = 0) = \mathcal{N}(\tilde{x}_{\gamma i} | 0, \alpha_\gamma^{-1}), \quad (16)$$

$$q(\tilde{x}_{\gamma i} | s_{\gamma i} = 1) = \mathcal{N}(\tilde{x}_{\gamma i} | \mu_{x_{\gamma i}}, \rho_\gamma^{-1}), \quad (17)$$

where

$$\mu_{x_{\gamma i}} = \frac{\beta_\gamma}{\rho_\gamma} \mathbf{h}_i^T (\mathbf{y}_\gamma - \sum_{k \neq i} \langle s_{\gamma k} \tilde{x}_{\gamma k} \rangle \mathbf{h}_k) \quad (18)$$

is the mean value of the slabs.

Finally, we can write the posterior as

$$\begin{aligned} q(\tilde{x}_{\gamma i}, s_{\gamma i}) &= q(\tilde{x}_{\gamma i} | s_{\gamma i}) q(s_{\gamma i}) \\ &= \mathcal{N}(\tilde{x}_{\gamma i} | s_{\gamma i} \mu_{x_{\gamma i}}, s_{\gamma i} \rho_\gamma^{-1} + (1 - s_{\gamma i}) \alpha_\gamma^{-1}) \omega_{\gamma i}^{s_{\gamma i}} (1 - \omega_{\gamma i})^{1-s_{\gamma i}}. \end{aligned} \quad (19)$$

Furthermore,

$$\langle s_{\gamma i} \tilde{x}_{\gamma i} \rangle = \langle x_{\gamma i} \rangle = \omega_{\gamma i} \mu_{x_{\gamma i}}, \quad (20)$$

$$\langle s_{\gamma i}^2 \tilde{x}_{\gamma i}^2 \rangle = \langle s_{\gamma i} \tilde{x}_{\gamma i}^2 \rangle = \langle x_{\gamma i}^2 \rangle = \omega_{\gamma i} (\mu_{x_{\gamma i}}^2 + \rho_\gamma^{-1}), \quad (21)$$

where  $\omega_{\gamma i}$  has been defined in Eq. (14) and  $\mu_{x_{\gamma i}}$  and  $\rho_\gamma$  in Eqs. (18) and (12), respectively.

### 3.2. Obtaining $q(\mathbf{h})$

Notice that we are assuming a degenerate distribution on  $q(\mathbf{h})$ , that is,

$$q(\mathbf{h}) = \begin{cases} 1 & \text{if } \mathbf{h} = \hat{\mathbf{h}} \\ 0 & \text{elsewhere} \end{cases}, \quad (22)$$

where  $\hat{\mathbf{h}} = \arg \min_{\mathbf{h} \in \mathcal{D}} \sum_\gamma \langle \|\mathbf{y}_\gamma - \mathbf{H}\mathbf{x}_\gamma\|^2 \rangle_{q(\Omega_h)}$  and  $\mathcal{D}$  denotes the set of valid blurs, i.e., non negative and add up to one.

Notice that

$$\begin{aligned} \sum_\gamma \langle \|\mathbf{y}_\gamma - \mathbf{H}\mathbf{x}_\gamma\|^2 \rangle_{q(\Omega_h)} &= \sum_\gamma \|\mathbf{y}_\gamma - \sum_j \langle x_{\gamma j} \rangle \mathbf{h}_j\|^2 \\ &+ \sum_\gamma \|\sum_j (x_{\gamma j} - \langle x_{\gamma j} \rangle) \mathbf{h}_j\|^2 \\ &= \sum_\gamma \|\mathbf{y}_\gamma - \mathbf{H}\langle \mathbf{x}_\gamma \rangle\|^2 + \sum_\gamma \sum_j (\langle x_{\gamma j}^2 \rangle - \langle x_{\gamma j} \rangle^2) \|\mathbf{h}\|^2 \end{aligned} \quad (23)$$

and so we have

$$\begin{aligned} \hat{\mathbf{h}} &= \arg \min_{\mathbf{h} \in \mathcal{D}} \sum_\gamma \left[ \|\mathbf{y}_\gamma - \mathbf{H}\langle \mathbf{x}_\gamma \rangle\|^2 \right. \\ &\quad \left. + \sum_j (\langle x_{\gamma j}^2 \rangle - \langle x_{\gamma j} \rangle^2) \|\mathbf{h}\|^2 \right], \end{aligned} \quad (24)$$

constrained to  $h_i \geq 0, \sum_i h_i = 1$ . This estimation problem can be efficiently solved using the ADMM method in [9].

## 4. BLIND DECONVOLUTION ALGORITHM

The blur estimation algorithm iterates on the estimation of the distribution  $q(\tilde{x}_{\gamma i}, s_{\gamma i})$  given the current estimate of  $\mathbf{h}$  and the estimation of  $q(\mathbf{h})$  from the current estimates of  $\tilde{x}_{\gamma i}$  and  $s_{\gamma i}$ . That is, starting from initial values  $\mathbf{h}^{(0)}$ ,  $\boldsymbol{\mu}_\Gamma^{(0)}$  and  $\boldsymbol{\omega}_\Gamma^{(0)}$ , and  $k = 0$ , the algorithm computes  $\boldsymbol{\mu}_\Gamma^{(k+1)}$  and  $\boldsymbol{\omega}_\Gamma^{(k+1)}$  to obtain  $\langle x_{\gamma i} \rangle^{(k+1)}, \forall \gamma \forall i$  in Eq. (20), and the estimation of  $\mathbf{h}^{(k+1)}$  using Eq. (24). It then sets  $k = k + 1$  and iterates again until convergence.

Following [6, 7, 9], we perform kernel estimation using a multiscale approach. This allows us to obtain a good kernel approximation at coarse scales, where it is easier to estimate, and provide a good starting point to finer levels by upsampling the kernel estimated at the previous scale.

Notice that whereas this algorithm does not provide an estimate of the image since it works on the filtered images, not on the image itself, it provides an estimate of the blur. Once the estimate of the blur,  $\hat{\mathbf{h}}$ , has been obtained, a non-blind deconvolution algorithm is used to recover an estimation of the original sharp image. In this paper we obtain an estimate of the original image by solving the problem

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\hat{\mathbf{H}}\mathbf{x} - \mathbf{y}\|_2^2 + \frac{\lambda}{p} \sum_\gamma \|\mathbf{x}_\gamma\|_p, \quad (25)$$

using the fast iterative method in [26], [9].

## 5. EXPERIMENTAL RESULTS

To assess the performance of the proposed method we have run the proposed algorithm on a set of 4 test images with 6 different blur kernels. The original images, displayed in the upper row of Fig. 1, were degraded by convolving them with each one of the blur kernels (see lower row of Fig. 1) and adding noise of standard deviation 0.01, thus obtaining a set of 24 degraded images.

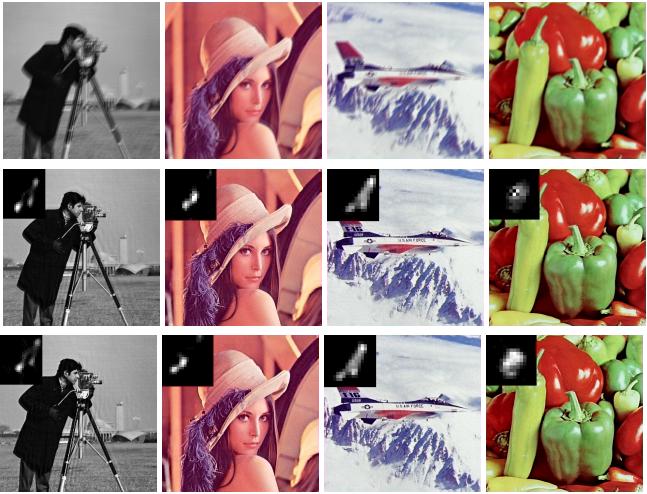
The proposed BID algorithm is initialized as follows. Initial blur at the lowest scale is initialized with a  $3 \times 3$  Gaussian kernel with very small variance. For the rest of the scales, the kernel estimated at the previous lower scale is upsampled by a factor  $\sqrt{2}$  in each direction. The precision parameters  $\beta_\Gamma$  were set to 5000,  $\forall \gamma$ , its real value according to the noise in the image (notice that  $\beta_\gamma = \beta / \|\mathbf{f}_\gamma\|^2$ ),  $\alpha_\gamma = \alpha$  and  $\pi_\gamma = \pi, \forall \gamma$ , were selected by grid search. At each scale,  $\boldsymbol{\mu}_\Gamma$  is initialized at the current pseudo-observation  $\mathbf{y}_\Gamma$  and  $\omega_{\gamma i} \forall \gamma \forall i$  is drawn from a Gaussian distribution with mean 0.5 and standard deviation 0.01.

We compared the proposed method with the BID method in [9] and, following the authors, we use the filters  $f_1 = [1, -1]$  and  $f_2 = [1, -1]^T$  for the blur estimation problem. For the final image reconstruction, we also use the second order derivative filters  $f_3 = [-1, 2, -1]$ ,  $f_4 = [-1, 2, -1]^T$ ,  $f_5 = [1 - 1; -1, 1]$  and the same parameter values to make a fair comparison.

Numerical results for PSNR and SSIM are presented in Tables 1 and 2, respectively. For the color images, PSNR and SSIM of the



**Fig. 1.** Original images and blur kernels used in the experiment. Images are  $512 \times 512$ . Blur kernel sizes range from  $13 \times 13$  to  $23 \times 23$ .



**Fig. 2.** Selected images for visual comparison. First row, degraded images; second row, restored with the proposed method; third row, restored with the method in [9]. The images are best viewed on the screen and zooming in to appreciate the details.

Y band is reported. The data suggest that the proposed method performs better than the method in [9] for most of the images and blur kernels tested, providing higher PSNR and similar SSIM values. We think that one of the proposed method strengths is its ability to identify good pixels for blur estimation. However, as in many other BID methods, good results depend on a good blur estimation at each scale and small variations or a poor estimation at a lower scale may ruin the final result. The proposed method which needs about 320 seconds to run on an i7-5500U CPU @ 2.40GHz with 16GB RAM, is slower than the method in [9] which needs about 50 seconds. The reason is that the image update in Eq. (20) is performed pixelwise, preventing the use of the FFT to speed up the computations. The modification of the image update to allow FFT processing is under study. Figure 2 depicts selected images for visual comparison. Color images were obtained by composing the deconvolved Y band with the degraded Cb and Cr bands. The images show that the proposed method provides sharper restorations with less noticeable artifacts.

Figure 3a shows a real blurred and noisy image borrowed from [27]. Noticeable artifacts appear on the deconvolved image with the method in [9] (Fig. 3b) while the result from the proposed method, depicted in Fig. 3c, is sharper, less noisy and present less artifacts. The estimated PSF (see the inset) is more accurate and not as noisy as the obtained with the method in [9].

| image | method   | kernel       |              |              |              |              |              |
|-------|----------|--------------|--------------|--------------|--------------|--------------|--------------|
|       |          | 1            | 2            | 3            | 4            | 5            | 6            |
| 1     | Proposed | <b>31.04</b> | 30.56        | 31.18        | <b>33.37</b> | <b>31.35</b> | <b>30.60</b> |
|       | Zhou [9] | 29.24        | <b>32.92</b> | <b>32.27</b> | 30.99        | 30.64        | 27.20        |
| 2     | Proposed | <b>31.03</b> | 30.99        | <b>31.41</b> | <b>32.63</b> | <b>30.81</b> | <b>32.04</b> |
|       | Zhou [9] | 30.11        | <b>31.00</b> | 30.29        | 29.25        | 30.40        | 31.62        |
| 3     | Proposed | 29.55        | <b>31.07</b> | 30.64        | <b>31.28</b> | 29.20        | 24.47        |
|       | Zhou [9] | <b>30.77</b> | 29.70        | <b>31.03</b> | 30.18        | <b>29.52</b> | <b>30.18</b> |
| 4     | Proposed | <b>31.04</b> | <b>30.47</b> | <b>31.75</b> | <b>31.66</b> | <b>30.31</b> | <b>31.21</b> |
|       | Zhou [9] | 30.08        | 29.96        | 30.60        | 30.15        | 29.48        | 30.44        |

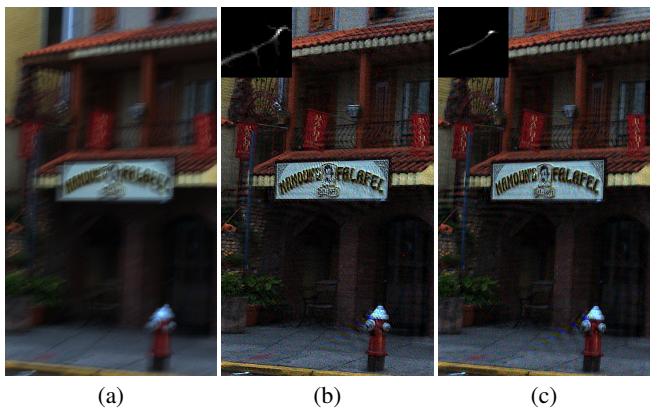
**Table 1.** PSNR value for the images and kernels in Fig. 1 for the proposed method and the method in [9].

| image | method   | kernel      |             |             |             |             |             |
|-------|----------|-------------|-------------|-------------|-------------|-------------|-------------|
|       |          | 1           | 2           | 3           | 4           | 5           | 6           |
| 1     | Proposed | 0.90        | 0.90        | 0.91        | 0.93        | 0.92        | <b>0.89</b> |
|       | Zhou [9] | 0.90        | <b>0.92</b> | <b>0.93</b> | 0.93        | 0.92        | 0.86        |
| 2     | Proposed | 0.85        | 0.86        | 0.86        | <b>0.89</b> | 0.87        | 0.87        |
|       | Zhou [9] | <b>0.86</b> | <b>0.87</b> | 0.86        | 0.86        | 0.87        | 0.87        |
| 3     | Proposed | 0.87        | 0.89        | 0.89        | 0.91        | 0.88        | 0.81        |
|       | Zhou [9] | <b>0.90</b> | 0.89        | <b>0.90</b> | 0.91        | <b>0.90</b> | <b>0.89</b> |
| 4     | Proposed | 0.81        | 0.80        | <b>0.82</b> | <b>0.83</b> | 0.80        | 0.81        |
|       | Zhou [9] | 0.81        | 0.80        | 0.81        | 0.82        | <b>0.81</b> | 0.81        |

**Table 2.** SSIM value for the images and kernels in Fig. 1 for the proposed method and the method in [9].

## 6. CONCLUSION

We have presented a new BID method formulated in the filter space. The novelty of the proposed model lies in the introduction of the spike-and-slab prior on the high-pass filtered image. A convenient reparametrization of the spike-and-slab prior makes VB inference possible and a sensible factorization provides a better approximation of the posterior. This leads to an efficient algorithm that accurately estimate the blur kernel due to the ability of such priors to shrink irrelevant information. Empirical experimentation provide sufficient proof of the competitiveness of the proposed method. Current research work is being devoted to the estimation of the model parameters, that is, noise variances and spike and slab parameters.



**Fig. 3.** A real example. (a) Blurred and noisy image. (b) Result with the method in [9]. (c) Result with the proposed method.

## 7. REFERENCES

- [1] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, “Efficient marginal likelihood optimization in blind deconvolution,” in *The Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.
- [2] L. Xu, S. C. Zheng, and J. Y. Jia, “Unnatural 10 sparse representation for natural image deblurring,” in *The Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.
- [3] P. Ruiz, X. Zhou, J. Mateos, R. Molina, and A.K. Katsaggelos, “Variational Bayesian blind image deconvolution: A review,” *Digital Signal Processing*, vol. 47, pp. 116–127, 2015.
- [4] R. Molina, J. Mateos, and A.K. Katsaggelos, “Blind deconvolution using a variational approach to parameter, image, and blur estimation,” *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3715–3727, Dec 2006.
- [5] T. E. Bishop, S. D. Babacan, B. Amizic, A. K. Katsaggelos, T. Chan, and R. Molina, *Blind image deconvolution: problem formulation and existing approaches*, chapter 1, CRC press, 2007.
- [6] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, “Removing camera shake from a single photograph,” *ACM Trans. Graph*, vol. 25, no. 3, 2006.
- [7] S. Babacan, R. Molina, M. Do, and A. K. Katsaggelos, “Bayesian blind deconvolution with general sparse image priors,” in *European Conference on Computer Vision (ECCV)*, 2012.
- [8] A. Mohammad-Djafari, “Bayesian blind deconvolution of images comparing JMAP, EM and BVA with a Student-t a priori model,” in *International Workshops on Electrical Computer Engineering Subfields*, 2014, pp. 98–103.
- [9] X. Zhou, M. Vega, F. Zhou, R. Molina, and A. K. Katsaggelos, “Fast Bayesian blind deconvolution with Huber super Gaussian priors,” *Digital Signal Processing*, vol. 60, pp. 122–133, 2017.
- [10] R. Giri and B. Rao, “Type I and Type II Bayesian Methods for Sparse Signal Recovery Using Scale Mixtures,” *IEEE Transactions on Signal Processing*, vol. 64, pp. 3418–3428, 2016.
- [11] R. Giri and B. Rao, “Learning Distributional Parameters for Adaptive Bayesian Sparse Signal Recovery,” *IEEE Computational Intelligence Magazine*, vol. 11, pp. 14–23, 2016.
- [12] X. Zhou, J. Mateos, F. Zhou, R. Molina, and A.K. Katsaggelos, “Variational Dirichlet blur kernel estimation,” *IEEE Transactions on Image Processing*, vol. 24, pp. 5127–5139, 2015.
- [13] M. K. Titsias and M. Lázaro-Gredilla, “Spike and slab variational inference for multi-task and multiple kernel learning,” in *Advances in Neural Information Processing Systems*, 2011, pp. 2339–2347.
- [14] L. Chaari, J.-Y. Toumeret, and H. Batatia, “Sparse Bayesian regularization using Bernoulli-Laplacian priors,” in *European Signal Processing Conference (EUSIPCO)*, 2013, pp. 1–5.
- [15] L. Chaari, J. Y. Toumeret, and C. Chaux, “Sparse signal recovery using a Bernoulli generalized Gaussian prior,” in *European Signal Processing Conference (EUSIPCO)*, 2015, pp. 1711–1715.
- [16] H. Ishwaran and J. S. Rao, “Spike and slab variable selection: Frequentist and Bayesian strategies,” *The Annals of Statistics*, vol. 33, no. 2, pp. 730–773, 2005.
- [17] S. A. Zilqurnain Naqvi, *Efficient sparse Bayesian learning using spike-and-slab priors*, Ph.D. thesis, Purdue University, 2016.
- [18] G. Malsiner-Walli and H Wagner, “Comparing spike and slab priors for Bayesian variable selection,” *Austrian Journal of Statistics*, vol. 40, pp. 241–264, 2011.
- [19] X. Huang, J. Wang, and F. Liang, “A variational Bayesian algorithm for variable selection,” in *Joint Statistical Meetings*, 2016.
- [20] X. Lu, Y. Yuan, and P. Yan, “Sparse coding for image denoising using spike and slab prior,” *Neurocomput.*, vol. 106, pp. 12–20, 2013.
- [21] S. Mohamed, K. A. Hellerand, and Z. Ghahramani, “Bayesian and  $l_1$  approaches for sparse unsupervised learning,” in *29th International Conference on Machine Learning*, 2012.
- [22] B. Lin, X. Tao, S. Li, L. Dong, and J. Lu, “Variational Bayesian image fusion based on combined sparse representations,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 1432–1436.
- [23] V. Ročková and E. I. George, “Negotiating multicollinearity with spike-and-slab priors,” *METRON*, vol. 72, pp. 217–229, 2014.
- [24] J. M. Hernández-Lobato, D. Hernández-Lobato, and A. Suárez, “Expectation propagation in linear regression models with spike-and-slab priors,” *Machine Learning*, vol. 99, pp. 437–487, 2015.
- [25] M. Zhou, H. Chen, J. Paisley, L. Ren, G. Sapiro, and L. Carin, “Non-parametric Bayesian dictionary learning for sparse image representations,” *Advances in Neural Information Processing Systems*, 2009.
- [26] X. Zhou, R. Molina, F. Zhou, and A. K. Katsaggelos, “Fast iteratively reweighted least squares for  $l_p$  regularized image deconvolution and reconstruction,” in *IEEE International Conference on Image Processing (ICIP)*, 2014.
- [27] L. Zhong, S. Cho, D. Metaxas, S. Paris, and J. Wang, “Handling noise in single image deblurring using directional filters,” in *The Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.