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].

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, ℓp, 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 an 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 ℓ0 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 peaked (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 dis-
In BID, the degraded observed image is modeled as a Gaussian random variable \( \tilde{x}_{\gamma i} \sim \mathcal{N}(\tilde{x}_{\gamma i}|0, \alpha_{\gamma i}^{-1}) \) and a Bernoulli random variable \( s_{\gamma i} \sim \pi_{\gamma i}^{s_{\gamma i}}(1 - \pi_{\gamma i})^{1-s_{\gamma i}} \), that is,

\[
x_{\gamma i} \sim s_{\gamma i}\tilde{x}_{\gamma i},
\]

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

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

where \( s_{\gamma i} \in \{0, 1\} \). We use the notation \( \tilde{x}_r = \{\tilde{x}_1, \ldots, \tilde{x}_L\} \) and \( s_r = \{s_1, \ldots, s_L\} \). With all the above, utilizing a flat prior on \( h \), and denoting the whole set of unknowns by \( \Theta = \{h, \tilde{x}_r, s_r\} \), we can write the joint distribution as

\[
p(\Theta, y_r) = p(h) \prod_{y_i} p(y_i|\tilde{x}_i, s_i, \beta_i) \times \prod_{s} \prod_{\gamma} p(\tilde{x}_{\gamma i}, s_{\gamma i}|\alpha_{\gamma i}, \pi_{\gamma i}).
\]

### 3. VARIATIONAL INFERENCE FOR BLUR ESTIMATION

Since \( p(\Theta|y_r) \) cannot be calculated in closed form, the standard mean field approximation [25] that factorizes \( q(\tilde{x}_i, s_i) = q(\tilde{x}_i)q(s_i) \) 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, using the factorization

\[
q(\tilde{x}_{\gamma i}, s_{\gamma i}) = \prod_{\gamma} \prod_{s} q(\tilde{x}_{\gamma i}, s_{\gamma i})
\]

and utilize the following mean field approximation

\[
q(\Theta) = q(h) \prod_{s} \prod_{\gamma} q(\tilde{x}_{\gamma i}, s_{\gamma i}).
\]

The distribution approximating the posterior \( p(\Theta|y_r) \) is found by minimizing the Kullback-Leibler divergence assuming that \( q(h) \) 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 \( q(h) \) 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

\[
q(\tilde{x}_{\gamma i}, s_{\gamma i}) = \frac{1}{Z} \exp \left[ \ln p(\gamma_i|y, h, x, s, \beta_{\gamma_i})q(\theta_{\gamma_i}, s_{\gamma_i}) \right]
\]

\[
\times \mathcal{N}(\tilde{x}_{\gamma i}|0, \alpha_{\gamma_i}^{-1})\pi_{\gamma_i}^{s_{\gamma_i}}(1 - \pi_{\gamma_i})^{1-s_{\gamma_i}},
\]

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 \( s_{\gamma i} \) and redefine the prior on the two components of \( x_{\gamma i} \), to \( \Theta \rightarrow \Theta_\theta \).

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

\[
x_{\gamma i} = s_{\gamma i}\tilde{x}_{\gamma i},
\]

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

\[
p(\tilde{x}_{\gamma i}, s_{\gamma i}|\alpha_{\gamma i}, \pi_{\gamma i}) = \mathcal{N}(\tilde{x}_{\gamma i}|0, \alpha_{\gamma_i}^{-1})\pi_{\gamma_i}^{s_{\gamma_i}}(1 - \pi_{\gamma_i})^{1-s_{\gamma_i}},
\]

The rest of the paper is organized as follows. Section 2 describes the Bayesian modeling of the blur estimation problem. Section 3.1 obtains the variational inference for blur estimation. Finally, Section 6 concludes the paper.
where $h_i$ denotes the $i$th column of $H$ and
\[ \rho_i = \beta \gamma \| h_i \|^2 + \alpha \gamma, \] (12)
with $\| h \|^2 = h^T h, \forall i$ since spatially invariant blur is assumed.

Following an analogous procedure for $s_{\gamma i} = 0$, we have
\[ q(s_{\gamma i} = 0) = \int q(\tilde{x}_{\gamma i}, 0) d\tilde{x}_{\gamma i} = \frac{1}{2} (1 - \pi \gamma)(2\pi)^{-N/2} \times \beta \gamma^{N/4} \exp \left[ -\frac{\beta \gamma}{2} \| y - \sum_{h_i \neq s_i} s_{\gamma i} \tilde{x}_{\gamma i} h_i \| ^2 \right]. \] (13)

Defining now
\[ \omega_{\gamma i} = q(s_{\gamma i} = 1) = \frac{1}{1 + e^{-u_{\gamma i}}}, \] (14)
it is easy to show that
\[ u_{\gamma i} = \ln q(s_{\gamma i} = 1) - \ln q(s_{\gamma i} = 0) = \ln \frac{\pi \gamma}{1 - \pi \gamma} + \frac{1}{2} \ln \alpha \gamma \]
\[ - \frac{1}{2} \ln(\rho) + \frac{\beta^2}{2\rho} \gamma (h^T (y - \sum_{h_i \neq s_i} s_{\gamma i} \tilde{x}_{\gamma i} h_i))^2. \] (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
\[ q(\tilde{x}_{\gamma i}|s_{\gamma i} = 0) = N(\tilde{x}_{\gamma i}|0, \alpha^{-1}), \] (16)
\[ q(\tilde{x}_{\gamma i}|s_{\gamma i} = 1) = N(\tilde{x}_{\gamma i}|\mu x_{\gamma i}, \rho^{-1}), \] (17)
where
\[ \mu x_{\gamma i} = \frac{\beta}{\rho} h_i^T (y - \sum_{h_i \neq s_i} s_{\gamma i} \tilde{x}_{\gamma i} h_i) \] (18)
is the mean value of the slabs.

Finally, we can write the posterior as
\[ q(\tilde{x}_{\gamma i}, s_{\gamma i}) = q(\tilde{x}_{\gamma i}|s_{\gamma i}) q(s_{\gamma i}) \] (19)
\[ = N(\tilde{x}_{\gamma i}|s_{\gamma i} \mu x_{\gamma i}, s_{\gamma i} \rho^{-1}) + (1-s_{\gamma i}) \alpha^{-1} \omega_{\gamma i} (1 - \omega_{\gamma i})^{-1} \gamma_{\gamma i}. \]
Furthermore,
\[ \langle s_{\gamma i}, \tilde{x}_{\gamma i} \rangle = \langle x_{\gamma i} \rangle = \omega_{\gamma i} \mu x_{\gamma i} \] (20)
\[ \langle x_{\gamma i}^2, \tilde{x}_{\gamma i} \rangle = \langle s_{\gamma i}, \tilde{x}_{\gamma i}^2 \rangle = \langle \tilde{x}_{\gamma i}^2 \rangle = \omega_{\gamma i} (\mu x_{\gamma i}^2 + \rho^{-1}), \] (21)
where $\omega_{\gamma i}$ has been defined in Eq. (14) and $\mu x_{\gamma i}$ and $\rho$ in Eqs. (18) and (12), respectively.

### 3.2. Obtaining $q(h)$

Notice that we are assuming a degenerate distribution on $q(h)$, that is,
\[ q(h) = \begin{cases} 1 & \text{if } h = \hat{h} \\ 0 & \text{elsewhere} \end{cases}, \] (22)
where $\hat{h} = \arg \min_{h \in D} \sum_\gamma \langle y - H x, h_j \rangle^2 > q(\alpha h)$ and $D$ denotes the set of valid blurs, i.e., non-negative and add up to one.

Notice that
\[ \sum_\gamma \langle y - H x, h_j \rangle^2 > q(\alpha h) = \sum_\gamma \langle y - \sum_j \langle x_{\gamma j} \rangle h_j \rangle^2 \]
\[ + \sum_\gamma \sum_j \langle x_{\gamma j} - \langle x_{\gamma j} \rangle \rangle h_j \|^2 \]
\[ = \sum_\gamma \| y - H x \|^2 + \sum_\gamma \sum_j \langle x_{\gamma j}^2 - \langle x_{\gamma j} \rangle^2 \rangle \| h \| ^2 \] (23)
and so we have
\[ \hat{h} = \arg \min_{h \in D} \sum_\gamma \left[ \| y - H \langle x \rangle \| ^2 + \sum_j \langle x_{\gamma j}^2 - \langle x_{\gamma j} \rangle^2 \rangle \| h \| ^2 \right], \] (24)
constrained to $h_i \geq 0, \sum h_i = 1$. 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 $\hat{h}$ and the estimation of $q(h)$ from the current estimates of $\tilde{x}_{\gamma i}$ and $s_{\gamma i}$. That is, starting from initial values $h^{(0)}, \mu^{(0)}$, and $\omega^{(0)}$, and $k = 0$, the algorithm computes $\mu^{(k+1)}$ and $\omega^{(k+1)}$ to obtain $\tilde{x}_{\gamma i}^{(k+1)}, \forall \gamma i$ in Eq. (20), and the estimation of $h^{(k+1)}$ using Eq. (24). It then sets $k = k + 1$ and iterates again until convergence.

Following [6, 7, 9], we perform kernel estimation using a multi-scale 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, $\hat{h}$, 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
\[ x = \arg \min_x \frac{1}{2} \| H x - y \| ^2 + \frac{\lambda}{p} \sum \| x \|_p, \] (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 $\beta$ were set to 5000, $\gamma_i$, its real value according to the noise in the image (notice that $\beta - \beta_0 / \| f \|^2$), $\alpha$ and $\gamma$, were selected by grid search. At each scale, $\mu^{(k+1)}$ is initialized at the current pseudo-observation $y_i$ and $\omega^{(k+1)} \forall \gamma i$ 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 $f_3 = [1, -1, 1]$, $f_2 = [1, -1, 1]$, and the same parameter values to make a fair the 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 3 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].

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

<table>
<thead>
<tr>
<th>image</th>
<th>method</th>
<th>kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Proposed</td>
<td>31.01</td>
</tr>
<tr>
<td></td>
<td>Zhou [9]</td>
<td>29.24</td>
</tr>
<tr>
<td>2</td>
<td>Proposed</td>
<td>31.03</td>
</tr>
<tr>
<td>3</td>
<td>Proposed</td>
<td>29.55</td>
</tr>
<tr>
<td></td>
<td>Zhou [9]</td>
<td>30.77</td>
</tr>
<tr>
<td>4</td>
<td>Proposed</td>
<td>31.04</td>
</tr>
<tr>
<td></td>
<td>Zhou [9]</td>
<td>30.08</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>image</th>
<th>method</th>
<th>kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Proposed</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>Zhou [9]</td>
<td>0.90</td>
</tr>
<tr>
<td>2</td>
<td>Proposed</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>Zhou [9]</td>
<td>0.86</td>
</tr>
<tr>
<td>3</td>
<td>Proposed</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>Zhou [9]</td>
<td>0.90</td>
</tr>
<tr>
<td>4</td>
<td>Proposed</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>Zhou [9]</td>
<td>0.81</td>
</tr>
</tbody>
</table>

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


