Parameter Estimation in 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*—Most current state of the art blind image deconvolution methods model the underlying image (either in the image or filter space) using sparsity promoting priors and perform inference, that is, image, blur, and parameter estimation using variational approximation. In this paper we propose the use of the spike-and-slab prior model in the filter space and a variational posterior approximation more expressive than mean field. The spike-and-slab prior model, which is the "gold-standard" in sparse machine learning, has the ability to selectively shrink irrelevant variables while relevant variables are mildly regularized. This allows to discard irrelevant information while preserving important features for the estimation of the blur which results in more precise and less noisy blur kernel estimates. In this paper we present a variational inference algorithm for estimating the blur in the filter space, which is both more efficient than MCMC and more accurate than the standard mean field variational approximation. The parameters of the prior model are automatically estimated together with the blur. Once the blur is estimated, a non-blind image restoration algorithm is used to obtain the sharp image. We prove the efficacy of our method on both synthetically generated and real images.**

I. Introduction

In image deconvolution, it is usual to model the degradation suffered by an original image $\mathbf{x}$ as [1]

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1}$$

where, using matrix-vector notation, $\mathbf{y}$ is the observed image, $\mathbf{n}$ represents the noise and the matrix $\mathbf{H}$ is the convolution matrix whose row elements are obtained from the blur kernel $\mathbf{h}$. In blind image deconvolution (BID) both the original image $\mathbf{x}$ and the blur $\mathbf{h}$ are unknown and both have to be estimated from the available data, that is, the degraded observed image. Notice that the inverse problem in (1) is underdetermined since for estimating the original image and the blur we have more unknowns than available observations. It is also an ill-posed problem in the sense of Hadamard, i.e., small variations in the observed data can lead to large variations in the solution and, even more, small variations in the estimated blur result in large variations in the restored image. To obtain good image and blur estimates, prior knowledge about the unknowns and sound estimation procedures are needed.

BID methods can be formulated in either the image or the filter space. In the image space, the restored image and blur are estimated directly from the observed image. In the filter space, however, the observed image is first filtered with different high-pass filters to obtain a set of sparse pseudo-observations. Several authors have discussed the advantages of using one space over the other [1]–[3]. While filter space methods have the advantage of having access to more pseudo-observations (one for each used filter) to estimate the blur, they appear to be more noise-sensitive. In the filter space, a non-blind deconvolution algorithm is always used to recover the sharp image after blur estimation. Image space approaches can estimate the image and blur simultaneously [4] although they usually first estimate the blur from the observed image and then they resort to a non-blind deconvolution algorithm to recover the final sharp image since images suited to blur estimation tend to have many steep edges and not many details.

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 [1], [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 special case. Notice also that popular sparse prior models, such as TV, ℓ_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. It is interesting to note that while most of the BID research concentrates on image modeling, 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, [10].

As indicated in [11], soft-sparse or shrinkage priors such as the Laplace and other related SMG 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 the ℓ_0 penalization [12], [13]. Such is the case

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.

of spike-and-slab priors [14], also named Bernoulli-Gaussian priors [13] since they consist of a mixture of a Bernoulli and a Gaussian distribution. These priors constitute the gold standard in sparse machine learning, having the ability to selectively shrink irrelevant variables, while mildly regularizing the relevant ones [15]. Applications of this prior include variable selection [16], [17], denoising [11], [18], inpainting [11], unsupervised latent variable models [19], hyper-spectral image fusion [20] and sparse signal recovery [13]. 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) [21] but this is still a mixture of two continuous distributions. According to [22] spike-and-slab models are more effective than other sparse priors (Laplacian or Student-t priors, for instance) in enforcing sparsity. The degree of sparsity can also 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.* [11], 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 [22] for the use of Expectation Propagation for posterior approximation).

Note that spike-and-slab models depend on a series of parameters that need to be set or estimated. The estimation of the model parameters is a very important issue in BID since they dramatically affect the quality of the final result. Usually, their values are chosen by trial-and-error, using a grid search on a set of selected values, generalized cross-validation (GCV) [23], discrepancy principle [24], or measures such as the whiteness measure [25]. However, when using the VB approach, the parameter estimation can be performed within the framework [4], [7], [9].

In this paper we formulate the BID problem in the filter space and introduce a spike-and-slab prior to model our knowledge on the original image in that space. This allows us to distinguish relevant observations for blur estimation from noisy ones, discarding the latter, and make the blur estimation more robust and accurate. We also derive a method to estimate the prior model parameters automatically from the observed data within the VB framework. 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 II describes the proposed model for blur estimation. In Sec. III, Bayesian inference is performed and, in Sec. IV, a BID algorithm is synthesized. The performance of the proposed method is assessed in Section V. Finally, Section VI concludes the paper.

II. Bayesian Modeling of the Blur Estimation Problem

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

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

with $\mathbf{x}_\gamma = \mathbf{F}_\gamma \mathbf{x}$ and $\mathbf{F}_\gamma$ is the convolution matrix associated with the filter $\mathbf{f}_\gamma$. In this work a linear convolution is assumed. The size of each pseudo-observation is N.

Assuming that the pseudo-observations are independent, and denoting $\mathbf{y}_\mathbf{\Gamma} = \{\mathbf{y}_\gamma\}_{\gamma=1}^L$, $\mathbf{x}_\mathbf{\Gamma} = \{\mathbf{x}_\gamma\}_{\gamma=1}^L$, and $\boldsymbol{\beta}_\mathbf{\Gamma} = \{\beta_\gamma\}_{\gamma=1}^L$ where β_γ, is the inverse of the noise variance, we can write

$$\mathrm{p}(\mathbf{y}_\mathbf{\Gamma}|\mathbf{h}, \mathbf{x}_\mathbf{\Gamma}, \boldsymbol{\beta}_\mathbf{\Gamma}) = \prod\nolimits_\gamma \mathrm{p}(\mathbf{y}_\gamma|\mathbf{h}, \mathbf{x}_\gamma, \beta_\gamma), \tag{3}$$

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

Notice that $\{\mathbf{x}_\gamma\}_{\gamma=1}^L$ should be 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 $\mathbf{x}_\gamma$,

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

where $\delta(\cdot)$ denotes the Dirac delta function and $\mathcal{N}(\tilde{x}_{\gamma i}|0, \alpha_\gamma^{-1})$ denotes a zero mean Gaussian distribution on $\tilde{x}_{\gamma i}$ with variance α_γ^{-1}. Observe that this is a truly sparse prior: $x_{\gamma i}$ is exactly zero with probability $1-\pi_\gamma$.

Let us now write, following [11], $x_{\gamma i}$ as the product of a Gaussian zero-mean 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,

$$x_{\gamma i} = s_{\gamma i}\tilde{x}_{\gamma i}, \tag{5}$$

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

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

where $s_{\gamma i} \in \{0, 1\}$. We use the notation $\tilde{\mathbf{x}}_\mathbf{\Gamma} = \{\tilde{\mathbf{x}}_\gamma\}_{\gamma=1}^L$, $\mathbf{s}_\mathbf{\Gamma} = \{\mathbf{s}_\gamma\}_{\gamma=1}^L$, $\boldsymbol{\alpha}_\mathbf{\Gamma} = \{\alpha_\gamma\}_{\gamma=1}^L$ and $\boldsymbol{\pi}_\mathbf{\Gamma} = \{\pi_\gamma\}_{\gamma=1}^L$ and utilize flat priors on the hyperparameters $\boldsymbol{\alpha}_\mathbf{\Gamma}$ and $\boldsymbol{\pi}_\mathbf{\Gamma}$ as well as on the blur $\mathbf{h}$. Other distributions can be used on the hyperparameters, such us gamma distributions [4], if any information about their value is available. Notice that, since the size of the blur is much smaller than the size of the image, enough data is available to obtain a precise estimation, even if no information on the prior form of the blur is available, which supports the use of an improper non-informative priors.

With all the above and denoting the whole set of unknowns by $\mathbf{\Theta} = \{\boldsymbol{\alpha}_\mathbf{\Gamma}, \boldsymbol{\pi}_\mathbf{\Gamma}, \mathbf{h}, \tilde{\mathbf{x}}_\mathbf{\Gamma}, \mathbf{s}_\mathbf{\Gamma}\}$, we can write the joint distribution as

$$\begin{aligned}\mathrm{p}(\mathbf{\Theta}, \mathbf{y}_\mathbf{\Gamma}) =& \mathrm{p}(\boldsymbol{\alpha}_\mathbf{\Gamma})\mathrm{p}(\boldsymbol{\pi}_\mathbf{\Gamma})\mathrm{p}(\mathbf{h}) \prod\nolimits_\gamma \mathrm{p}(\mathbf{y}_\gamma|\mathbf{h}, \tilde{\mathbf{x}}_\gamma, \mathbf{s}_\gamma, \beta_\gamma) \\ &\times \prod\nolimits_\gamma \prod\nolimits_i \mathrm{p}(\tilde{x}_{\gamma i}, s_{\gamma i}|\alpha_\gamma, \pi_\gamma).\end{aligned} \tag{7}$$

III. Variational Inference for Blur estimation

Since $\mathrm{p}(\boldsymbol{\Theta}|\mathbf{y}_{\boldsymbol{\Gamma}})$ cannot be calculated in closed form, the standard mean field approximation [26] that factorizes $\mathrm{q}(\tilde{x}_{\boldsymbol{\gamma}}, \mathbf{s}_{\boldsymbol{\gamma}}) = \mathrm{q}(\tilde{x}_{\boldsymbol{\gamma}})\mathrm{q}(\mathbf{s}_{\boldsymbol{\gamma}})$ could be used. However this is a unimodal distribution [11] and, therefore, not a good approximation of the true posterior distribution. Since the pairs $\{\tilde{x}_{\gamma i}, s_{\gamma i}\}$ are strongly correlated (recall that $x_{\gamma i} = s_{\gamma i}\tilde{x}_{\gamma i}$), we treat them as a unit, hence we use the factorization

$$\mathrm{q}(\tilde{x}_{\boldsymbol{\gamma}}, \mathbf{s}_{\boldsymbol{\gamma}}) = \prod\nolimits_{\gamma} \prod\nolimits_{i} \mathrm{q}(\tilde{x}_{\gamma i}, s_{\gamma i}) \tag{8}$$

and utilize the following mean field approximation

$$\mathrm{q}(\boldsymbol{\Theta}) = \mathrm{q}(\boldsymbol{\alpha}_{\boldsymbol{\Gamma}})\mathrm{q}(\boldsymbol{\pi}_{\boldsymbol{\Gamma}})\mathrm{q}(\mathbf{h}) \prod\nolimits_{\gamma} \prod\nolimits_{i} \mathrm{q}(\tilde{x}_{\gamma i}, s_{\gamma i}). \tag{9}$$

The distribution approximating the posterior $\mathrm{p}(\boldsymbol{\Theta}|\mathbf{y}_{\boldsymbol{\Gamma}})$ is found by minimizing the Kullback-Leibler divergence. In absence of any other knowledge, we assume that $\mathrm{q}(\boldsymbol{\alpha}_{\boldsymbol{\Gamma}})$, $\mathrm{q}(\boldsymbol{\pi}_{\boldsymbol{\Gamma}})$ and $\mathrm{q}(\mathbf{h})$ are degenerate distributions and that the blur is non-negative and its coefficients add up to one.

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

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

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

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

where $\mathcal{Z}$ is the partition function, $\theta \in \boldsymbol{\Theta}$, and $\boldsymbol{\Theta}_{\theta}$ denotes $\boldsymbol{\Theta}$ with θ removed. To compute the explicit expression for the above posterior we separate the derivations for $\mathrm{q}(s_{\gamma i})$ and $\mathrm{q}(\tilde{x}_{\gamma i}|s_{\gamma i})$ obtaining

$$\omega_{\gamma i} = q(s_{\gamma i} = 1) = \frac{1}{1 + e^{-u_{\gamma i}}}, \tag{11}$$

where

$$\begin{aligned}u_{\gamma i} = \ln \mathrm{q}(s_{\gamma i} = 1) - \ln \mathrm{q}(s_{\gamma i} = 0) = \ln \frac{\pi_{\gamma}}{1-\pi_{\gamma}} + \frac{1}{2}\ln \alpha_{\gamma} \\ - \frac{1}{2}\ln(\rho_{\gamma}) + \frac{\beta_{\gamma}^2}{2\rho_{\gamma}}\big(\mathbf{h}_i^{\mathrm{T}}(\mathbf{y}_{\gamma} - \sum\nolimits_{k \neq i} \langle s_{\gamma k}\tilde{x}_{\gamma k}\rangle \mathbf{h}_k)\big)^2\end{aligned} \tag{12}$$

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

$$\rho_{\gamma} = \beta_{\gamma} \parallel \mathbf{h} \parallel^2 + \alpha_{\gamma}, \tag{13}$$

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

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

$$\mathrm{q}(\tilde{x}_{\gamma i}|s_{\gamma i} = 0) = \mathcal{N}(\tilde{x}_{\gamma i}|0, \alpha_{\gamma}^{-1}), \tag{14}$$

$$\mathrm{q}(\tilde{x}_{\gamma i}|s_{\gamma i} = 1) = \mathcal{N}(\tilde{x}_{\gamma i}|\mu_{x_{\gamma i}}, \rho_{\gamma}^{-1}), \tag{15}$$

where

$$\mu_{x_{\gamma i}} = \frac{\beta_{\gamma}}{\rho_{\gamma}}\mathbf{h}_i^{\mathrm{T}}(\mathbf{y}_{\gamma} - \sum\nolimits_{k \neq i} \langle s_{\gamma k}\tilde{x}_{\gamma k}\rangle \mathbf{h}_k) \tag{16}$$

is the mean value of the slabs.

Finally, we can write the posterior as

$$\begin{aligned}\mathrm{q}(\tilde{x}_{\gamma i}, s_{\gamma i}) = \mathrm{q}(\tilde{x}_{\gamma i}|s_{\gamma i})\mathrm{q}(s_{\gamma i}) = \omega_{\gamma i}^{s_{\gamma i}}(1-\omega_{\gamma i})^{1-s_{\gamma i}} \\ \times \mathcal{N}(\tilde{x}_{\gamma i}|s_{\gamma i}\mu_{x_{\gamma i}}, s_{\gamma i}\rho_{\gamma}^{-1} + (1-s_{\gamma i})\alpha_{\gamma}^{-1}).\end{aligned} \tag{17}$$

Furthermore,

$$\langle s_{\gamma i}\tilde{x}_{\gamma i}\rangle = \langle x_{\gamma i}\rangle = \omega_{\gamma i}\mu_{x_{\gamma i}}, \tag{18}$$

$$\langle s_{\gamma i}^2\tilde{x}_{\gamma i}^2\rangle = \langle s_{\gamma i}\tilde{x}_{\gamma i}^2\rangle = \langle x_{\gamma i}^2\rangle = \omega_{\gamma i}(\mu_{x_{\gamma i}}^2 + \rho_{\gamma}^{-1}), \tag{19}$$

where $\omega_{\gamma i}$, $\mu_{x_{\gamma i}}$ and ρ_{γ} have been defined in (11), (16) and (13), respectively.

B. Obtaining $\mathrm{q}(\mathbf{h})$

Notice that we assume a degenerate distribution on $\mathrm{q}(\mathbf{h})$, which leads to obtaining the point estimate of $\mathbf{h}$, $\hat{\mathbf{h}}$, as

$$\begin{aligned}\hat{\mathbf{h}} &= \arg\min_{\mathbf{h}} \sum\nolimits_{\gamma} \langle \parallel \mathbf{y}_{\gamma} - \mathbf{H}\mathbf{x}_{\gamma} \parallel^2 \rangle_{\mathrm{q}(\Omega_{\mathbf{h}})} \\ &= \arg\min_{\mathbf{h}} \sum\nolimits_{\gamma} \Big[\parallel \mathbf{y}_{\gamma} - \mathbf{H}\langle \mathbf{x}_{\gamma}\rangle \parallel^2 \\ &\quad + \sum\nolimits_{i} (\langle x_{\gamma i}^2\rangle - \langle x_{\gamma i}\rangle^2) \parallel \mathbf{h} \parallel^2 \Big],\end{aligned} \tag{20}$$

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

C. Obtaining $\mathrm{q}(\boldsymbol{\pi}_{\boldsymbol{\Gamma}})$ and $\mathrm{q}(\boldsymbol{\alpha}_{\boldsymbol{\Gamma}})$

Finally, we proceed to estimate the parameters of the prior model, $\boldsymbol{\pi}_{\boldsymbol{\Gamma}}$ and $\boldsymbol{\alpha}_{\boldsymbol{\Gamma}}$. We factorize $\mathrm{q}(\boldsymbol{\pi}_{\boldsymbol{\Gamma}}) = \prod_{\gamma} \mathrm{q}(\pi_{\gamma})$ where

$$\mathrm{q}(\pi_{\gamma}) = \begin{cases} 1 & \text{if } \pi_{\gamma} = \hat{\pi}_{\gamma} \\ 0 & \text{elsewhere} \end{cases} \tag{21}$$

and obtain $\hat{\pi}_{\gamma} = \arg\max_{0<\pi_{\gamma}<1} \pi_{\gamma}^{\sum_i \langle s_{\gamma i}\rangle}(1-\pi_{\gamma})^{\sum_i \langle 1-s_{\gamma i}\rangle}$ which produces

$$\hat{\pi}_{\gamma} = \frac{1}{N}\sum\nolimits_{i} \langle s_{\gamma i}\rangle = \frac{\sum_i \omega_{\gamma i}}{N}. \tag{22}$$

We also factorize $\mathrm{q}(\boldsymbol{\alpha}_{\boldsymbol{\gamma}}) = \prod_{\gamma} \mathrm{q}(\alpha_{\gamma})$ where

$$\mathrm{q}(\alpha_{\gamma}) = \begin{cases} 1 & \text{if } \alpha_{\gamma} = \hat{\alpha}_{\gamma} \\ 0 & \text{elsewhere} \end{cases} \tag{23}$$

and obtain $\hat{\alpha}_{\gamma} = \arg\max_{\alpha_{\gamma}} \prod_i \mathcal{N}(\tilde{x}_{\gamma i}|0, \alpha_{\gamma}^{-1})$, which yields

$$\hat{\alpha}_{\gamma} = \frac{\sum_i \langle s_{\gamma i}\rangle}{\sum_i \langle x_{\gamma i}^2\rangle} = \frac{\sum_i \omega_{\gamma i}}{\sum_i \omega_{\gamma i}(\mu_{x_{\gamma i}}^2 + \rho_{\gamma}^{-1})}. \tag{24}$$

IV. Blind deconvolution algorithm

The blur estimation algorithm iterates on the estimation of each one of the unknowns from their respective distributions given the current estimate of the rest of the unknowns which leads to the estimation procedure summarized in Algorithm 1.

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 it, and provide a good starting point to finer levels by upsampling the kernel estimated at the previous scale.

Algorithm 1 Blur estimation algorithm

Require: Filtered degraded image set $\mathbf{y}_{\mathbf{\Gamma}}$, noise precision $\boldsymbol{\beta}_{\mathbf{\Gamma}}$ and initial values for $\mathbf{h}^0$, $\boldsymbol{\pi}^0_{\boldsymbol{\gamma}}$, $\boldsymbol{\alpha}^0_{\boldsymbol{\gamma}}$ and $\omega^0_{\gamma i}$ and $\mu^0_{x_{\gamma i}}$, for $\gamma = 1 \ldots, L$, $i = 1, \ldots, N$.
Set $k = 0$.
repeat
For each filter γ and each image pixel i, update $\omega^k_{\gamma i}$ and $\mu^k_{x_{\gamma i}}$ from (11) and (16), respectively, and update $\langle x_{\gamma i} \rangle^k$ and $\langle x^2_{\gamma i} \rangle^k$ from (18) and (19), respectively.
Find $\boldsymbol{\pi}^{k+1}_{\mathbf{\Gamma}}$ and $\boldsymbol{\alpha}^{k+1}_{\mathbf{\Gamma}}$ from (22) and (24), respectively.
Find $\mathbf{h}^{k+1}$ using (20).
Set $k = k + 1$.
until convergence
return $\hat{\mathbf{h}} = \mathbf{h}^k$.

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 obtain an estimate of the original sharp image. In this paper we obtain an estimate of the original image by solving the problem

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2} \parallel \hat{\mathbf{H}}\mathbf{x} - \mathbf{y} \parallel_2^2 + \frac{\lambda}{p} \sum\nolimits_{\gamma} \parallel \mathbf{x}_{\gamma} \parallel_p, \qquad (25)$$

using the fast iterative method in [27], [9].

V. Experimental Results

We have tested the proposed method on a set of 4 test images with 6 different blur kernels. Each original image in the upper row of Fig. 1 was convolved with each one of the blur kernels in the lower row of Fig. 1 and Gaussian noise of variance 10^{-4} was added to obtain a set of 24 degraded images.

The proposed method was compared with the BID method in [9]. Following [9], 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 $f_3 = [-1, 2, -1]$, $f_4 = [-1, 2, -1]^T$, $f_5 = [1 - 1; -1, 1]$. For a fair comparison, we also use the same values for the parameters.

The proposed algorithm is initialized as follows. Initial blur at the lowest scale is a 3×3 Gaussian kernel with variance 0.5. 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 β_γ were set to 5000, $\forall\gamma$, its real value according to the noise in the image (notice that $\beta_\gamma = \beta / \|\mathbf{f}_\gamma\|^2$). At each scale, $\boldsymbol{\mu}_{\mathbf{\Gamma}}$ is initialized at the current pseudo-observation $\mathbf{y}_{\mathbf{\Gamma}}$ and $\omega_{\gamma i}$, $\forall\gamma\forall i$, is drawn from a Gaussian distribution with mean 0.5 and standard deviation 0.01, $\pi_\gamma = 0.5$, $\forall\gamma$, and $\boldsymbol{\alpha}_{\mathbf{\Gamma}}$ is obtained from the previous scale.

Some restored images and their corresponding blur estimates are shown in Fig. 2. The proposed method obtains restorations with fewer noticeable artifacts. Color images were obtained by composing the deconvolved Y band with the degraded Cb and Cr bands. Numerical PSNR and SSIM values for the (Y band) of the whole set of images are presented in Table I. 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. 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 400 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. (18) is performed pixelwise, preventing the use of the FFT to speed up the computations and that the method in [9] does not automatically estimate the parameters so several runs may be needed to adjust the method's parameters.

Figure 3a shows a real blurred and noisy image borrowed from [28]. Noticeable artifacts appear on the deconvolved image with the method in [9] (Fig. 3b), while the result by the proposed method, depicted in Fig. 3c, is sharper, less noisy and presents fewer artifacts. The estimated PSF (see the inset) is more accurate and not as noisy as the one obtained with the method in [9].

Figure 1. Original images and blur kernels. Images are 512×512. Blur kernel sizes range from 13×13 to 23×23.

Table I
PSNR and SSIM value for the proposed method and the method in [9].

image	measure	method	kernel 1	2	3	4	5	6
1	PSNR	Proposed	**30.01**	30.44	30.16	**32.69**	28.70	**31.33**
		Zhou [9]	29.24	**32.92**	**32.27**	30.99	**30.64**	27.20
	SSIM	Proposed	0.885	0.890	0.897	**0.928**	0.871	**0.904**
		Zhou [9]	**0.899**	**0.920**	**0.927**	0.927	**0.917**	0.861
2	PSNR	Proposed	**30.49**	29.56	**31.72**	**30.62**	**30.88**	28.87
		Zhou [9]	30.11	**31.00**	30.29	29.25	30.40	**31.62**
	SSIM	Proposed	0.830	0.836	**0.863**	**0.870**	0.858	0.844
		Zhou [9]	**0.863**	**0.867**	0.862	0.865	**0.870**	**0.874**
3	PSNR	Proposed	29.63	26.52	**32.53**	30.04	**30.38**	28.90
		Zhou [9]	**30.77**	**29.70**	31.03	**30.18**	29.52	**30.18**
	SSIM	Proposed	0.862	0.769	**0.909**	0.893	0.892	0.884
		Zhou [9]	**0.899**	**0.891**	0.905	**0.908**	**0.896**	**0.895**
4	PSNR	Proposed	**31.38**	**31.05**	**32.46**	**31.35**	**30.46**	29.04
		Zhou [9]	30.08	29.96	30.60	30.15	29.48	**30.44**
	SSIM	Proposed	**0.807**	0.800	**0.821**	0.810	0.805	0.789
		Zhou [9]	0.806	**0.803**	0.812	**0.817**	**0.807**	**0.807**

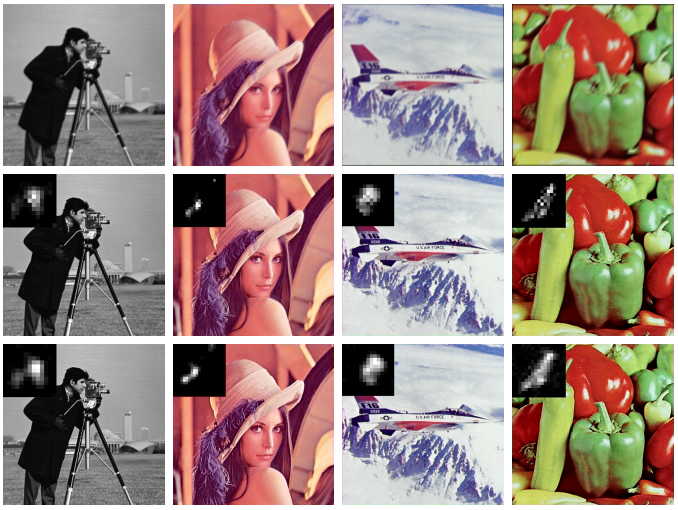

Figure 2. Selected images for visual comparison. First row, degraded images; second row, restored with the proposed method; third row, restored with the method in [9]. The images are best viewed on the screen and zooming in.

(a) (b) (c)

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

VI. 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 estimates the blur kernel due to the ability of such priors to shrink irrelevant information. All the prior model parameters are estimated along with the blur within the VB framework. Automatic estimation of the noise parameter, as well as the use of other hyperprior distributions on the parameters, is under study. Empirical experimentation provides sufficient proof of the competitiveness of the proposed method.

References

[1] P. Ruiz, X. Zhou, J. Mateos, R. Molina, and A.K. Katsaggelos, "Variational Bayesian blind image deconvolution: A review," *Dig. Sig. Proc.*, vol. 47, pp. 116–127, 2015.

[2] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," in *CVPR*, 2011.

[3] L. Xu, S. C. Zheng, and J. Y. Jia, "Unnatural l0 sparse representation for natural image deblurring," in *CVPR*, 2013.

[4] R. Molina, J. Mateos, and A.K. Katsaggelos, "Blind deconvolution using a variational approach to parameter, image, and blur estimation," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3715–3727, 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 *ECCV*, 2012.

[8] A. Mohammad-Djafari, "Bayesian blind deconvolution of images comparing JMAP, EM and BVA with a Student-t a priori model," in *Int. Work. on Elec. Comp. Eng. 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," *Dig. Sig. Proc.*, vol. 60, pp. 122–133, 2017.

[10] X. Zhou, J. Mateos, F. Zhou, R. Molina, and A.K. Katsaggelos, "Variational Dirichlet blur kernel estimation," *IEEE Trans. Image Process.*, vol. 24, pp. 5127–5139, 2015.

[11] M. K. Titsias and M. Lázaro-Gredilla, "Spike and slab variational inference for multi-task and multiple kernel learning," in *NIPS*, 2011, pp. 2339–2347.

[12] L. Chaari, J.-Y. Tourneret, and H. Batatia, "Sparse Bayesian regularization using Bernoulli-Laplacian priors," in *EUSIPCO*, 2013, pp. 1–5.

[13] L. Chaari, J. Y. Tourneret, and C. Chaux, "Sparse signal recovery using a Bernoulli generalized Gaussian prior," in *EUSIPCO*, 2015, pp. 1711–1715.

[14] H. Ishwaran and J. S. Rao, "Spike and slab variable selection: Frequentist and Bayesian strategies," *The Ann. of Stat.*, vol. 33, no. 2, pp. 730–773, 2005.

[15] S. A. Zilqurnain Naqvi, *Efficient sparse Bayesian learning using spike-and-slab priors*, Ph.D. thesis, Purdue University, 2016.

[16] G. Malsiner-Walli and H Wagner, "Comparing spike and slab priors for Bayesian variable selection," *A. J. of Stat.*, vol. 40, pp. 241–264, 2011.

[17] X. Huang, J. Wang, and F. Liang, "A variational Bayesian algorithm for variable selection," in *Joint Stat. Meetings*, 2016.

[18] X. Lu, Y. Yuan, and P. Yan, "Sparse coding for image denoising using spike and slab prior," *Neurocomput.*, vol. 106, pp. 12–20, 2013.

[19] S. Mohamed, K. A. Hellerand, and Z. Ghahramani, "Bayesian and l_1 approaches for sparse unsupervised learning," in *ICML*, 2012.

[20] B. Lin, X. Tao, S. Li, L. Dong, and J. Lu, "Variational Bayesian image fusion based on combined sparse representations," in *ICASSP*, 2016, pp. 1432–1436.

[21] V. Ročková and E. I. George, "Negotiating multicollinearity with spike-and-slab priors," *METRON*, vol. 72, pp. 217–229, 2014.

[22] J. M. Hernández-Lobato, D. Hernández-Lobato, and A. Suárez, "Expectation propagation in linear regression models with spike-and-slab priors," *Mach. Lear.*, vol. 99, pp. 437–487, 2015.

[23] H. Liao and M. K. Ng, "Blind deconvolution using generalized cross-validation approach to regularization parameter estimation," *IEEE Trans. Image Process.*, vol. 20, pp. 670–680, 2011.

[24] Y. W. Wen and R. H. Chan, "Parameter selection for total-variation-based image restoration using discrepancy principle," *IEEE Trans. Image Process.*, vol. 21, pp. 1770–1781, 2012.

[25] M. S. C. Almeida and M. A. T. Figueiredo, "Parameter estimation for blind and non-blind deblurring using residual whiteness measures," *IEEE Trans. Image Process.*, vol. 22, pp. 2751–2763, 2013.

[26] M. Zhou, H. Chen, J. Paisley, L. Ren, G. Sapiro, and L. Carin, "Non-parametric Bayesian dictionary learning for sparse image representations," *NIPS*, 2009.

[27] X. Zhou, R. Molina, F. Zhou, and A. K. Katsaggelos, "Fast iteratively reweighted least squares for l_p regularized image deconvolution and reconstruction," in *ICIP*, 2014.

[28] L. Zhong, S. Cho, D. Metaxas, S. Paris, and J. Wang, "Handling noise in single image deblurring using directional filters," in *CVPR*, 2013.