Variational EM method for blur estimation using the spike-and-slab image prior ☆

Juan G. Serra [a],[*], Javier Mateos [a],[*], Rafael Molina [a], Aggelos K. Katsaggelos [b]

[a] *Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain*
[b] *Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208-3118, USA*

ARTICLE INFO

ABSTRACT

Most state-of-the-art blind image deconvolution methods rely on the Bayesian paradigm to model the deblurring problem and estimate both the blur kernel and latent image. It is customary to model the image in the filter space, where it is supposed to be sparse, and utilize convenient priors to account for this sparsity. In this paper, we propose the use of the spike-and-slab prior together with an efficient variational Expectation Maximization (EM) inference scheme to estimate the blur in the image. The spike-and-slab prior, which constitutes the gold standard in sparse machine learning, selectively shrinks irrelevant variables while mildly regularizing the relevant ones. The proposed variational Expectation Maximization algorithm is more efficient than usual Markov Chain Monte Carlo (MCMC) inference and, also, proves to be more accurate than the standard mean-field variational approximation. Additionally, all the prior model parameters are estimated by the proposed scheme. After blur estimation, a non-blind restoration method is used to obtain the actual estimation of the sharp image. We investigate the behavior of the prior in the experimental section together with a series of experiments with synthetically generated and real blurred images that validate the method's performance in comparison with state-of-the-art blind deconvolution techniques.

1. Introduction

Image deconvolution or deblurring aims at retrieving the underlying sharp image from a blurred and noisy observation. Both blur and noise are usually undesired degradations inherent in image acquisition which negatively affect images in a variety of fields including commercial photography, astronomy, medical imaging or remote sensing, to name a few. We focus our work on the general and challenging case where both blur and image are unknown, henceforth referred to as Blind Image Deconvolution (BID).

Formally, the degradation suffered by an image $\mathbf{x}$ is typically modeled as [1]

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

where $\mathbf{y}$ is the degraded observed image, $\mathbf{n}$ represents the noise, and $\mathbf{H}$ is the circulant convolution matrix whose row elements are obtained from the blur kernel $\mathbf{h}$, also known as the point spread function (PSF). Note that the elements of $\mathbf{h}$ are non-negative and sum up to one in order to guarantee that, for zero mean noise, the real and observed images have (approximately) the same mean value. In blind image deconvolution, since the original image and the blur are unknown, they have to be estimated from the available data, that is, the degraded observed image. This inverse problem is highly underdetermined since in the estimation of the original image and blur there are more unknowns than available observations, and therefore multiple blur/image pairs will be compatible with the observed data. It is also an ill-posed problem in the sense of Hadamard [2], 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, adequate prior knowledge on the unknowns should be introduced to constrain the solution space and sound estimation procedures are needed to find appropriate solutions.

BID methods can be formulated in either the image or the filter space [1]. While in the image space, the restored image and blur are estimated directly from the observed image, in the filter space several pseudo-observations are first generated by applying a set of high-pass filters to the original image. Notice that these pseudo-observations will be sparse since natural images contain

* Corresponding authors.
E-mail addresses: jgserra@decsai.ugr.es (J.G. Serra), jmd@decsai.ugr.es (J. Mateos), rms@decsai.ugr.es (R. Molina), aggk@eecs.northwestern.edu (A.K. Katsaggelos).

many smooth regions which will be set equal to zero by the high-pass filters, while producing higher intensity values at pixels corresponding to edges. Multiple works have discussed the advantages and disadvantages of each space, see, for example, [3,4,1]. While filter space methods have access to more pseudo-observations to estimate the blur (one per filter) they appear to be more sensitive to noise. Additionally, following the estimation of the blur, a non-blind deconvolution method is always used to recover the original image. Image space approaches, on the other hand, can estimate the image and blur simultaneously [5]. However, restored images often present a cartoon-like appearance since good blur estimates rely on images with steep edges and few details. Hence, a non-blind deconvolution method is commonly used to recover the final sharp image with the blur estimated from the observed image.

Typically, stochastic blind deconvolution methods fall into two main classes, Maximum *a Posteriori* (MAP) estimation and Variational Inference (VI). We can find an interesting comparison of these methods in [6], where it is demonstrated that VI has a better ability to avoid bad local solutions when the prior model is highly sparse. However, variational methods are usually more computationally demanding than MAP-based ones [1] since they have to estimate the full posterior distribution, not only its maximum. Early attempts on BID used to apply MAP or regularization approaches (which are equivalent to MAP [7]) and assumed that the kernel had a given parametric form [8], reducing the blur estimation problem to the estimation of the parameters of the blur function [9], or imposed other constraints, such as non-negativity, symmetry or smoothness [10,11]. To avoid the trivial delta kernel solution, which is a preferred MAP solution [12], carefully designed image and blur priors are needed [13] and *ad hoc* intermediate edge enhanced images are used [14–16]. Recent works make use, for instance, of hyper-Laplacian priors [17] with iteration-wise tuning of parameters [18,15], log-TV priors [19], approximations of the ℓ_0 function (see the work in [4] extended in [16] with the use of a dark channel prior) or combinations of ℓ_0 and ℓ_2 priors on the gradient image and blur kernel [20]. Noticing that insignificant edges and edges of small objects result in inaccurate kernel estimation, a two-phase kernel estimation with salient-edge selection was proposed in [21], producing a first estimate of the blur from heuristically selected informative edges, which is later refined by iterative regularization of the small values in the PSF. Gong et al. [22] also observed that the blur kernel can be robustly estimated from a subset of the image gradients and proposed an automatic algorithm via a cutting-plane method for selecting those gradients. Note that these approaches can be related to the spike-and-slab prior which will be presented later. Some of the most recent methods directly exploit the inherent characteristics of natural images. For instance, in [23] an ℓ_p prior is imposed on the blur and the Gaussian Scale Mixture Field of Experts prior is introduced for the image, whose filters are learned from natural images. Cai et al. [24] use salient edges as well as sparse representations to incorporate learned data as a prior for both image and kernel estimation. The work in [25] utilizes a CNN as a prior that distinguishes between blurry and sharp images. Finally, a Gaussian mixture model based prior, trained on images of the specific class (i.e., text, face or fingerprint) under restoration is used in [26].

Variational Inference BID methods also rely on the use of sparsity promoting image models (see [1] for a review of image models used in VI BID). The use of sparsity and VI is a successful approximation of growing interest since the seminal work of [27], where a mixture of Gaussians (MoG) is used. Notice that the use of VI in image deconvolution was proposed early in [28], see also [5]. This method was generalized by Babacan et al. [29] where a BID VI framework based on Super Gaussian (SG) and Scale Mixture of Gaussian (SMG) priors was presented. Other popular sparse prior models, such as TV, ℓ_p, MoG, and Student-t [30] are included in the above mentioned framework. While most common SG priors lack differentiability around zero, Huber Super Gaussian (HSG) priors, recently proposed in [31], circumvent this problem. Recently, mixtures of exponential distributions, such as the Laplace distribution, have been introduced [32,33]. This mixture, named Power Exponential Scale Mixture (PESM), is a generalization of Gaussian and Laplacian Scale Mixtures used to model prior sparse distributions. Although most of the BID literature concentrates on priors on the image and blur, 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, [34].

Soft-sparse or shrinkage priors, such as the Laplace and other related SMG priors, may not be ideal sparsity-promoting priors [35] since they assign zero probability to the event of a random variable taking exactly the value zero. Recall that in continuous distributions, the probability of a single point is zero. As a result, there is a growing interest in priors that combine discrete and continuous distributions that better approximate the ℓ_0 penalization [36,37]. Such is the case of spike-and-slab priors [38], also named Bernoulli–Gaussian priors [37] since they consist of a product of a Bernoulli (discrete) and a Gaussian (continuous) 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 [39]. Applications of this prior include variable selection [40,41], denoising [42,35], inpainting [35], unsupervised latent variable models [43], hyper-spectral image fusion [44] and sparse signal recovery [45, 37]. In the BID problem at hand, we expect this prior to discard noisy observations and to select only pixels that are relevant to blur estimation.

Alternatively, the spike-and-slab prior can be modeled by a linear convex combination of a Gaussian distribution, the slab, and a delta function, the spike. Although this prior can be approximated by the mixture of two Gaussians [46], one with a very low variance, this is still a mixture of two continuous distributions and thus not truly sparse. Unfortunately, preserving the sparsity enforcing property makes variational inference for spike-and-slab models a very challenging task in several ways, due to the presence of the delta in the prior. First, the exact posterior can not be calculated in closed form and, secondly, the classical mean field variational inference removes essential dependencies in the posterior distribution approximation. Due to these problems, the costly Monte-Carlo Markov Chain (MCMC) sampling has been the typical choice for inference, until recently. For multi-task and multiple Gaussian process learning problems, Titsias et al. [35] propose an alternative Variational EM (VEM) inference model, a kind of VI model, which solves these problems using a simple reparametrization of the prior. This reparametrization eliminates the inconvenient delta function and better matches the combinatorial nature of the posterior, resulting in a simple and efficient algorithm (see also [47] for the use of Expectation Propagation for posterior approximation).

Note that prior distributions, including spike-and-slab, are governed by a set of parameters that need to be set or estimated. The estimation of the model parameters is crucial in BID algorithms since they dramatically affect the quality of the final result. Usually, their values are chosen by trial-and-error, grid search, generalized cross-validation [48], discrepancy principle [49], or measures such as the whiteness measure [50]. However, VI, and VEM specifically, allow for direct parameter estimation within their framework [5,29,31].

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 prior endows our BID method with the capability to distinguish relevant from noisy observations, discarding the latter and thus produces more accurate and robust blur estimates. This paper adapts the variational inference

with the spike-and-slab prior to the blur estimation problem for the first time in the literature. Moreover, the prior model parameters are automatically estimated from the observed data within the Bayesian framework. The VEM 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. A preliminary version of this formulation was presented in [51,52]. In this paper, however, a better parameter estimation procedure, a deep study of the spike-and-slab prior behavior, as well as an extensive experimental validation, have been performed.

The rest of the paper is organized as follows. Section 2 describes the proposed model for blur estimation. In Sec. 3, the inference procedure is detailed and, in Sec. 4, our complete BID algorithm is presented. The performance of the proposed method is assessed in Section 5. Finally, Section 6 concludes the paper and discusses future research directions.

2. Bayesian modeling

The BID problem in (1) is formulated here in the filter space. We generate L pseudo-observations $\mathbf{y}_\gamma \in \mathbb{R}^N$ by filtering the blurred and noisy image $\mathbf{y}$ with a set of high-pass filters $\{\mathbf{f}_\gamma\}_{\gamma=1}^L$ obtaining

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

with $\mathbf{x}_\gamma = \mathbf{F}_\gamma \mathbf{x}$, where $\mathbf{F}_\gamma$ is a block-circulant convolution matrix associated with the filter $\mathbf{f}_\gamma$, which clearly commutes with the block-circulant matrix $\mathbf{H}$.

Assuming independent pseudo-observations [12,1], denoting by $\mathbf{y}_\Gamma = \{\mathbf{y}_\gamma\}_{\gamma=1}^L$, $\mathbf{x}_\Gamma = \{\mathbf{x}_\gamma\}_{\gamma=1}^L$, and β_γ the precision parameter which is assumed to be known, we can write our observation model as

$$p(\mathbf{y}_\Gamma | \mathbf{h}, \mathbf{x}_\Gamma) = \prod_\gamma p(\mathbf{y}_\gamma | \mathbf{h}, \mathbf{x}_\gamma) = \prod_\gamma \mathcal{N}(\mathbf{y}_\gamma | \mathbf{H} \mathbf{x}_\gamma, \beta_\gamma^{-1} \mathbf{I})$$

$$\propto \prod_\gamma \exp\left\{ -\frac{\beta_\gamma}{2} \|\mathbf{y}_\gamma - \mathbf{H} \mathbf{x}_\gamma\|^2 \right\}. \qquad (3)$$

Notice that the images $\{\mathbf{x}_\gamma\}_{\gamma=1}^L$ are typically sparse since they represent high-pass filtered instances of the original image. Following the Bayesian approach, sparsity is enforced by the use of prior distributions on the solution. In this paper we use a spike-and-slab prior on the value of each pixel $x_{\gamma i}$ of $\mathbf{x}_\gamma$, so that

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

where $\delta(\cdot)$ denotes the Dirac delta function. Observe that this is a truly sparse prior, that is, $x_{\gamma i}$ is exactly zero with probability $1 - \pi_\gamma$. However, the delta function in the prior model hampers the inference procedure.

We can overcome this problem with a simple but powerful reparametrization of $x_{\gamma i}$, expressing it as the product of a Gaussian zero-mean random variable $\tilde{x}_{\gamma i} \sim \mathcal{N}(0, \alpha_{\gamma i}^{-1})$ and a Bernoulli random variable $s_{\gamma i} \sim \pi_\gamma^{s_{\gamma i}}(1 - \pi_\gamma)^{1 - s_{\gamma i}}$ [35], i.e.,

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

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

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

where $s_{\gamma i} \in \{0, 1\}$.

We use the notation $\tilde{\mathbf{x}}_\Gamma = \{\tilde{\mathbf{x}}_\gamma\}_{\gamma=1}^L$, $\mathbf{s}_\Gamma = \{\mathbf{s}_\gamma\}_{\gamma=1}^L$, $\boldsymbol{\alpha}_\Gamma = \{\boldsymbol{\alpha}_\gamma\}_{\gamma=1}^L$, $\boldsymbol{\pi}_\Gamma = \{\pi_\gamma\}_{\gamma=1}^L$, $\boldsymbol{\Omega} = \{\mathbf{h}, \boldsymbol{\alpha}_\Gamma, \boldsymbol{\pi}_\Gamma\}$, and $\boldsymbol{\Theta} = \{\tilde{\mathbf{x}}_\Gamma, \mathbf{s}_\Gamma\}$. Then we have

$$p(\mathbf{y}_\Gamma, \tilde{\mathbf{x}}_\Gamma, \mathbf{s}_\Gamma | \boldsymbol{\Omega})$$

$$= \left[\prod_\gamma \prod_i p(\tilde{x}_{\gamma i}, s_{\gamma i} | \alpha_{\gamma i}, \pi_\gamma) \right] \prod_\gamma p(\mathbf{y}_\gamma | \mathbf{h}, \tilde{\mathbf{x}}_\gamma, \mathbf{s}_\gamma). \qquad (7)$$

3. Variational inference for blur estimation

The logarithm of the marginal likelihood, which should be used to estimate the model parameters, is

$$\log p(\mathbf{y}_\Gamma | \boldsymbol{\Omega}) = \log\left(\sum_{\mathbf{s}_\Gamma} \int p(\mathbf{y}_\Gamma, \tilde{\mathbf{x}}_\Gamma, \mathbf{s}_\Gamma | \boldsymbol{\Omega}) d\tilde{\mathbf{x}}_\Gamma \right). \qquad (8)$$

The variational method maximizes the following Jensen's lower bound on the above log marginal likelihood

$$\mathcal{F} = \log\left(\sum_{\mathbf{s}_\Gamma} \int q(\tilde{\mathbf{x}}_\Gamma, \mathbf{s}_\Gamma) \log \frac{p(\mathbf{y}_\Gamma, \tilde{\mathbf{x}}_\Gamma, \mathbf{s}_\Gamma | \boldsymbol{\Omega})}{q(\tilde{\mathbf{x}}_\Gamma, \mathbf{s}_\Gamma)} d\tilde{\mathbf{x}}_\Gamma \right). \qquad (9)$$

Notice that since $\log p(\mathbf{y}_\Gamma | \boldsymbol{\Omega}) \geq \mathcal{F}$, by increasing $\mathcal{F}$ on $\boldsymbol{\Omega}$ we are increasing a lower bound of the log-likelihood. Furthermore, we will have $p(\tilde{\mathbf{x}}_\Gamma, \mathbf{s}_\Gamma | \boldsymbol{\Omega}, \mathbf{y}_\Gamma) \approx q(\tilde{\mathbf{x}}_\Gamma, \mathbf{s}_\Gamma)$.

The factorization $q(\tilde{\mathbf{x}}_\Gamma, \mathbf{s}_\Gamma) = \prod_\gamma q(\tilde{\mathbf{x}}_\gamma, \mathbf{s}_\gamma)$, where $q(\tilde{\mathbf{x}}_\gamma, \mathbf{s}_\gamma) = \prod_i q(\tilde{x}_{\gamma i}) q(s_{\gamma i})$, has been extensively used (see [45], for instance). However, as pointed out in [35], this is a unimodal distribution, which fails to capture the multimodality of the true posterior (observe that the posterior $p(\tilde{\mathbf{x}}_\gamma | \mathbf{y}_\gamma, \boldsymbol{\Omega}) = \sum_{\mathbf{s}_\gamma} p(\tilde{\mathbf{x}}_\gamma, \mathbf{s}_\gamma | \mathbf{y}_\gamma, \boldsymbol{\Omega})$ involves the summation over the 2^N possible values of the binary vector $\mathbf{s}_\gamma$, thus being able to have multiple modes). Instead, 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, and use the factorization

$$q(\tilde{\mathbf{x}}_\gamma, \mathbf{s}_\gamma) = \prod_i q(\tilde{x}_{\gamma i}, s_{\gamma i}). \qquad (10)$$

In the next two subsections we present a variational EM algorithm for the maximization of $\mathcal{F}$.

3.1. E-step

Given $\boldsymbol{\Omega}$, the stationary condition for $q(\tilde{x}_{\gamma i}, s_{\gamma i})$, i.e., the best distribution maximizing $\mathcal{F}$, is

$$q(\tilde{x}_{\gamma i}, s_{\gamma i}) = \frac{1}{\mathcal{Z}} e^{\langle \log p(\mathbf{y}_\gamma | \mathbf{h}, \tilde{\mathbf{x}}_\gamma, \mathbf{s}_\gamma) \rangle_{\boldsymbol{\Theta}_{\tilde{x}_{\gamma i}, s_{\gamma i}}}} \mathcal{N}(\tilde{x}_{\gamma i} | 0, \alpha_{\gamma i}^{-1}) \pi_\gamma^{s_{\gamma i}}$$

$$\times (1 - \pi_\gamma)^{1 - s_{\gamma i}}, \qquad (11)$$

where $\boldsymbol{\Theta}_\theta$ denotes $\boldsymbol{\Theta}$ with θ removed and $\langle \cdot \rangle_{\boldsymbol{\Theta}_\theta}$ denotes the mean calculated using the distribution $q(\boldsymbol{\Theta}_\theta)$.

Furthermore,

$$\langle \log p(\mathbf{y}_\gamma | \mathbf{h}, \tilde{\mathbf{x}}_\gamma, \mathbf{s}_\gamma) \rangle_{\boldsymbol{\Theta}_{\tilde{x}_{\gamma i}, s_{\gamma i}}} = -\frac{\beta_\gamma}{2} \langle \|\mathbf{y}_\gamma - \mathbf{H}(\tilde{\mathbf{x}}_\gamma \odot \mathbf{s}_\gamma)\|^2 \rangle_{\boldsymbol{\Theta}_{\tilde{x}_{\gamma i}, s_{\gamma i}}}$$

$$+ \frac{N}{2} \log \beta_\gamma - \frac{N}{2} \log 2\pi, \qquad (12)$$

with

$$\langle \|\mathbf{y}_\gamma - \mathbf{H}(\tilde{\mathbf{x}}_\gamma \odot \mathbf{s}_\gamma)\|^2 \rangle_{\boldsymbol{\Theta}_{\tilde{x}_{\gamma i}, s_{\gamma i}}}$$

$$= \langle \|\mathbf{y}_\gamma - \sum_{k \neq i} \mathbf{h}_k s_{\gamma k} \tilde{x}_{\gamma k} - \mathbf{h}_i s_{\gamma i} \tilde{x}_{\gamma i}\|^2 \rangle_{\boldsymbol{\Theta}_{\tilde{x}_{\gamma i}, s_{\gamma i}}}$$

$$= -2(\mathbf{y}_\gamma - \sum_{k \neq i} \mathbf{h}_k \langle s_{\gamma k} \tilde{x}_{\gamma k} \rangle)^{\mathrm{T}} \mathbf{h}_i \tilde{x}_{\gamma i} s_{\gamma i} + \|\mathbf{h}\|^2 \tilde{x}_{\gamma i}^2 s_{\gamma i}$$

$$+ \langle \|\mathbf{y}_\gamma - \sum_{k \neq i} \mathbf{h}_k s_{\gamma k} \tilde{x}_{\gamma k}\|^2 \rangle, \qquad (13)$$

where $\mathbf{h}_i$ denotes the ith column of $\mathbf{H}$, $\|\mathbf{h}\|^2 = \mathbf{h}_i^T \mathbf{h}_i, \forall i$, since we assume a spatially invariant blur, $\odot$ represents the Hadamard product, and the mean $\langle s_{\gamma k} \tilde{x}_{\gamma k} \rangle$ is calculated using the distribution $q(\tilde{x}_{\gamma k}, s_{\gamma k})$.

To compute the explicit expression for the posterior approximation we separate the derivations for $q(\tilde{x}_{\gamma i}|s_{\gamma i})$ and $q(s_{\gamma i})$. Plugging (13) into (11) and setting $s_{\gamma i} = 0$ and $s_{\gamma i} = 1$, we obtain

$$q(\tilde{x}_{\gamma i}|s_{\gamma i}=0) \propto \exp\left\{ -\frac{\alpha_{\gamma i}}{2} \tilde{x}_{\gamma i}^2 \right\}, \tag{14}$$

$$q(\tilde{x}_{\gamma i}|s_{\gamma i}=1) \propto \exp\left\{ \beta_\gamma (\mathbf{y}_\gamma - \sum_{k\neq i} \mathbf{h}_k \langle s_{\gamma k} \tilde{x}_{\gamma k} \rangle)^T \mathbf{h}_i \tilde{x}_{\gamma i} \right.$$
$$\left. - \frac{\beta_\gamma}{2} \|\mathbf{h}\|^2 \tilde{x}_{\gamma i}^2 - \frac{\alpha_{\gamma i}}{2} \tilde{x}_{\gamma i}^2 \right\}. \tag{15}$$

We can easily deduce from the quadratic form of (14) and (15) that both distributions are Gaussian,

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

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

where

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

$$\rho_{\gamma i} = \beta_\gamma \|\mathbf{h}\|^2 + \alpha_{\gamma i}, \tag{19}$$

are the mean value and the inverse of the variance of the slabs, respectively. Equation (19) admits a nice interpretation. The variance of the slabs is always smaller than the variance of the prior since it includes information from the data but, more importantly, it depends on the blur. If the norm of the blur is high (close to one) it means that the blur has a small number of significant values, i.e., only a few pixels contribute to the current one, and hence, the estimation of the image can be reliably calculated. As the number of pixels influenced by the blur increases, the norm of the blur will decrease and so will the confidence on the estimation.

We can now calculate $q(s_{\gamma i})$ by marginalizing (11) over $\tilde{x}_{\gamma i}$. Let $\omega_{\gamma i} = q(s_{\gamma i} = 1)$, then using simple algebraic calculation we observe that

$$\omega_{\gamma i} = \frac{1}{1 + e^{-u_{\gamma i}}}, \tag{20}$$

with $u_{\gamma i} = \log q(s_{\gamma i} = 1) - \log q(s_{\gamma i} = 0)$. This will liberate us from calculating the partition function $\mathcal{Z}$ of the distribution $q(s_{\gamma i}, \tilde{x}_{\gamma i})$, since it will be canceled out in the subtraction. We have from (11)

$$\log q(s_{\gamma i}=1) = \log \int q(\tilde{x}_{\gamma i}, 1) d\tilde{x}_{\gamma i}$$
$$= \frac{N}{2} \log \beta_\gamma - \frac{N}{2} \log 2\pi + \frac{1}{2} \log \alpha_{\gamma i}$$
$$- \frac{\beta_\gamma}{2} \langle \|\mathbf{y}_\gamma - \sum_{k\neq i} \mathbf{h}_k x_{\gamma k}\|^2 \rangle + \log \pi_\gamma - \frac{1}{2} \log \rho_{\gamma i}$$
$$+ \frac{\beta_\gamma^2}{2\rho_{\gamma i}} \left(\mathbf{h}_i^T (\mathbf{y}_\gamma - \sum_{k\neq i} \langle s_{\gamma k} \tilde{x}_{\gamma k} \rangle \mathbf{h}_k) \right)^2 - \log \mathcal{Z}, \tag{21}$$

and

$$\log q(s_{\gamma i}=0) = \log \int q(\tilde{x}_{\gamma i}, 0) d\tilde{x}_{\gamma i} = \frac{N}{2} \log \beta_\gamma - \frac{N}{2} \log 2\pi$$
$$- \frac{\beta_\gamma}{2} \langle \|\mathbf{y}_\gamma - \sum_{k\neq i} \mathbf{h}_k x_{\gamma k}\|^2 \rangle + \log(1 - \pi_\gamma)$$
$$- \log \mathcal{Z}, \tag{22}$$

which produces

$$u_{\gamma i} = \log \frac{\pi_\gamma}{1 - \pi_\gamma} + \frac{1}{2} \log \frac{\alpha_{\gamma i}}{\rho_{\gamma i}}$$
$$+ \frac{\beta_\gamma^2}{2\rho_{\gamma i}} \left(\mathbf{h}_i^T (\mathbf{y}_\gamma - \sum_{k\neq i} \langle s_{\gamma k} \tilde{x}_{\gamma k} \rangle \mathbf{h}_k) \right)^2, \tag{23}$$

and allows us to calculate $\omega_{\gamma i} = q(s_{\gamma i} = 1)$ using (20). From this equation it is clear that the value of $\omega_{\gamma i}$ will be close to one if the error of the current estimation $(\mathbf{y}_\gamma - \sum_{k\neq i} \langle s_{\gamma k} \tilde{x}_{\gamma k} \rangle \mathbf{h}_k)$ is high and the estimation will be discarded, focusing on the prior. If the estimation is good, the value of $\omega_{\gamma i}$ will depend on the relation between π_γ, $\alpha_{\gamma i}$, and $\rho_{\gamma i}$ but more probably it will take values close to zero, valuing the importance of the estimation.

Finally, we can write the full posterior $q(\tilde{x}_{\gamma i}, s_{\gamma i})$ as

$$q(\tilde{x}_{\gamma i}, s_{\gamma i}) = q(s_{\gamma i})q(\tilde{x}_{\gamma i}|s_{\gamma i}) = \omega_{\gamma i}^{s_{\gamma i}} (1 - \omega_{\gamma i})^{1-s_{\gamma i}}$$
$$\times \mathcal{N}(\tilde{x}_{\gamma i}|s_{\gamma i}\mu_{x_{\gamma i}}, s_{\gamma i}\rho_{\gamma i}^{-1} + (1 - s_{\gamma i})\alpha_{\gamma i}^{-1}). \tag{24}$$

Furthermore, using $q(s_{\gamma i}, \tilde{x}_{\gamma i})$ we have

$$\langle x_{\gamma i} \rangle = \langle s_{\gamma i} \tilde{x}_{\gamma i} \rangle = \sum_{s_{\gamma i}} \int s_{\gamma i} \tilde{x}_{\gamma i} q(\tilde{x}_{\gamma i}, s_{\gamma i}) d\tilde{x}_{\gamma i}$$
$$= \int \tilde{x}_{\gamma i} q(\tilde{x}_{\gamma i}, 1) d\tilde{x}_{\gamma i}$$
$$= \omega_{\gamma i} \int \tilde{x}_{\gamma i} q(\tilde{x}_{\gamma i}|s_{\gamma i}=1) d\tilde{x}_{\gamma i} = \omega_{\gamma i} \mu_{x_{\gamma i}}, \tag{25}$$

$$\langle x_{\gamma i}^2 \rangle = \langle s_{\gamma i}^2 \tilde{x}_{\gamma i}^2 \rangle = \sum_{s_{\gamma i}} \int s_{\gamma i}^2 \tilde{x}_{\gamma i}^2 q(\tilde{x}_{\gamma i}, s_{\gamma i}) d\tilde{x}_{\gamma i}$$
$$= \int \tilde{x}_{\gamma i}^2 q(\tilde{x}_{\gamma i}, 1) d\tilde{x}_{\gamma i}$$
$$= \omega_{\gamma i} \int \tilde{x}_{\gamma i}^2 q(\tilde{x}_{\gamma i}|s_{\gamma i}=1) d\tilde{x}_{\gamma i}$$
$$= \omega_{\gamma i} (\mu_{x_{\gamma i}}^2 + \rho_{\gamma i}^{-1}), \tag{26}$$

where $\mu_{x_{\gamma i}}$, $\rho_{\gamma i}$ and $\omega_{\gamma i}$ have been defined in (18), (19) and (20) respectively.

3.2. M-step

Fixing $q(\tilde{\mathbf{x}}_\Gamma, \mathbf{s}_\Gamma)$, we now proceed to calculate $\mathbf{\Omega} = \{\mathbf{h}, \boldsymbol{\alpha}_\Gamma, \boldsymbol{\pi}_\Gamma\}$ by maximizing the lower bound $\mathcal{F}$ in (9) on $\mathbf{\Omega}$.

In order to update π_γ we have to maximize the function

$$\mathcal{F}(\pi_\gamma) = \sum_i \left[\langle s_{\gamma i} \rangle \log \pi_\gamma + (1 - \langle s_{\gamma i} \rangle) \log(1 - \pi_\gamma) \right] \tag{27}$$

where $0 < \pi_\gamma < 1$, which produces

$$\pi_\gamma = \frac{\sum_i \omega_{\gamma i}}{N}. \tag{28}$$

To find $\alpha_{\gamma i}$ we have to maximize

$$\mathcal{F}(\alpha_{\gamma i}) = \omega_{\gamma i} \int \mathcal{N}(\tilde{x}_{\gamma i}|\mu_{x_{\gamma i}}, \rho_{\gamma i}^{-1}) \log \mathcal{N}(\tilde{x}_{\gamma i}|0, \alpha_{\gamma i}^{-1}) d\tilde{x}_{\gamma i}$$
$$= \omega_{\gamma i} \left[\frac{1}{2} \log \alpha_{\gamma i} - \frac{1}{2} \alpha_{\gamma i} \langle \tilde{x}_{\gamma i}^2 \rangle_{\mathcal{N}(\tilde{x}_{\gamma i}|\mu_{x_{\gamma i}}, \rho_{\gamma i}^{-1})} \right] + \text{const}, \tag{29}$$

which produces

$$\alpha_{\gamma i}^{-1} = \mu_{x_{\gamma i}}^2 + \rho_{\gamma i}^{-1}. \tag{30}$$

Finally, we have to minimize

$$f(\mathbf{h}) = \sum_\gamma \langle \|\mathbf{y}_\gamma - \mathbf{H}\mathbf{x}_\gamma\|^2 \rangle_\Theta = \sum_\gamma \left[\|\mathbf{y}_\gamma - \mathbf{H}\langle\mathbf{x}_\gamma\rangle\|^2 + \mathbf{h}^{\mathrm{T}}\mathbf{D}_\gamma\mathbf{h} \right],$$

(31)

constrained to $h_i \geq 0$, $\sum_i h_i = 1$, where $\mathbf{D}_\gamma$ is a diagonal matrix of the form $\mathbf{D}_\gamma = \sum_i (\langle x_{\gamma i}^2 \rangle - \langle x_{\gamma i} \rangle^2)\mathbf{I}$. This estimation problem can be efficiently solved using quadratic programming, see [29].

4. 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.

Algorithm 1 Blur estimation algorithm.

Require: $\mathbf{y}_\Gamma$, $\boldsymbol{\beta}_\Gamma$ and initial values $\mathbf{h}^{(0)}$, $\boldsymbol{\pi}_\Gamma^{(0)}$, $\boldsymbol{\alpha}_\Gamma^{(0)}$ and $\langle\mathbf{x}_\Gamma\rangle^{(0)}$.
Set $k = 0$.
repeat
    Update $\mu_{x_{\gamma i}}^{(k)}$, $\rho_\gamma^{(k)}$, and $\omega_{\gamma i}^{(k)}$ $\forall\gamma$, $\forall i$, from (18), (19), and (20) respectively.
    Update $\langle x_{\gamma i}\rangle^{(k)}$ $\forall i$, from (25).
    Update $\boldsymbol{\pi}_\Gamma^{(k+1)}$ from (28), and $\boldsymbol{\alpha}_\Gamma^{(k+1)}$ from (30).
    Update $\mathbf{h}^{(k+1)}$ using (31).
    Set $k = k+1$.
until convergence.
return $\hat{\mathbf{h}} = \mathbf{h}^{(k)}$.

An important issue in BID is the initialization of the blur kernel since most algorithms, as already pointed out in [27], are subject to local minima, particularly for large blur kernels. It is common to use a multiscale approach for blur kernel estimation (see [27, 29,31]) since it is easier to produce accurate blur estimates at coarser scales, which are later used as initialization for the next scale. However, we perform blur estimation only at the finest scale, initializing the blur with the final estimate produced by [31]. The method in [31] is a multiscale BID algorithm that imposes sparsity on the images of gradients by means of the Huber Super-Gaussian priors. These priors circumvent the differentiability problems of Super-Gaussian priors around zero. Image and blur kernel estimations make use of the Alternating Direction Method of Multipliers (ADMM), which allows for a fast variational inference due to FFT computations. Since the method in [31] takes advantage of the FFT to compute the filtered images estimations it is much faster than our algorithm which iterates on each image pixel sequentially. So, we utilize this speed to obtain a sensible blur initialization for our algorithm.

Notice that whereas the proposed 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

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

(32)

using the fast iterative method in [53,31]. The value of p is usually selected so that $p \in [0.6, 0.8]$ [31]. Following [54,55,31], we fixed $p = 0.8$.

Finally, note that for color images a single blur kernel is estimated for all the image color bands from the luminance image. Afterwards, the estimated PSF is used in the above non-blind image restoration algorithm to restore each of the RGB color bands separately.

5. Experimental results

In this section, we present a series of experiments to analyze the performance of the blind deconvolution algorithm presented in this paper. Subsection 5.1 delves into the behavior of the spike-and-slab prior and its role in the image estimation process. Subsection 5.2 is devoted to compare our method with three state-of-the-art blind deconvolution techniques: firstly, a set of experiments is carried out in a controlled set-up, secondly, the algorithm is tested on real images and, finally, a succinct time analysis is carried out.

Unless otherwise stated, the algorithm is initialized as follows. We use the filters $\mathbf{f}_1 = [1\ 0\ -1; 1\ 0\ -1; 1\ 0\ -1]$, $\mathbf{f}_2 = \mathbf{f}_1^{\mathrm{T}}$ to obtain the pseudo-observations from the observed blurred image, according to (2). We empirically determined that the addition of more filters has no significant impact on the blur estimation. However, it clearly increases the computational burden of the algorithm. Initial blur is set to the final estimation from the algorithm presented in [31]. The value for $\langle\mathbf{x}_\Gamma\rangle$ is initialized with the pseudo-observation $\mathbf{y}_\Gamma$. Additionally, all π_γ are initialized at 0.01 and $\alpha_{\gamma i} = 32$, $\forall i$. Estimating $\alpha_{\gamma i}^{-1}$ according to (30) produces noisy estimations that are smoothed with the filter $[0\ 1\ 0; 1\ 8\ 1; 0\ 1\ 0]/12$. The precision parameters β_γ are considered known and initialized to $\beta_\gamma = \beta/\|\mathbf{f}_\gamma\|^2$, $\forall\gamma$, with $\beta = 10^4$, following [12]. Notice that accurate estimates of the noise variance of the images can be obtained from flat regions of the image or with any of the various methods in the literature (see [56] and the references therein and [57–59] for more recent publications). Nevertheless, we experimentally found a low sensitivity to this parameter.

5.1. Spike-and-slab prior behavior

The spike-and-slab prior for BID we introduced in this paper is able to shrink information that is not relevant to blur estimation. In this section we study how this prior acts on the image estimation process and its effect on the image.

Fig. 1 depicts the parameters and image estimations from the vertical ($\gamma = 1$) and horizontal ($\gamma = 2$) pseudo-observations assuming that the blur is known. The prior in (4) is a product of distributions which assumes that the values of $x_{\gamma i}$ come from a zero-mean Gaussian with variance $\alpha_{\gamma i}^{-1}$ with probability π_γ or take exactly the value 0, with probability $1 - \pi_\gamma$. For the image in Fig. 1 when $\gamma = 1$, $\pi_1 = 0.057$, this means that a given pixel in the background will take exactly the value 0 with probability 0.943 and a value from a zero-mean Gaussian with variance 260^{-1} with probability 0.057. Even more, the full posterior distribution in (24) follows the same distribution as the prior in (6) but with its parameters updated with the observed data. The mean of the posterior (see (33)) is depicted in Fig. 1 with the labels $\langle\mathbf{x}_1\rangle$ and $\langle\mathbf{x}_2\rangle$, this figure shows that the proposed algorithm recovers images with sharp edges.

Fig. 1 also shows the estimated values of $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ which consist of a large number of pixels with a value very close to zero and a few pixels with a value very close to one. Taking into account that $\pi_\Gamma = \{0.057, 0.028\}$, it means that more than 95% of the pixels are set to zero in the estimation. Note that most of these pixels correspond to flat regions which do not provide information on the blur, or to highly textured regions where it is difficult to discriminate noise from texture. It is also important to note that some prominent edges in the image have also been shrunk. Such is the case with pixels where noise is dominant or affected by ringing artifacts as, for instance, the pixels around the head or the elbow in $\mu_{\mathbf{x}_1}$ in Fig. 1 that have been zeroed in $\langle\mathbf{x}_1\rangle$. Therefore, only the pixels that provide the most reliable information for blur estimation are selected by the spike-and-slab prior and they are mildly
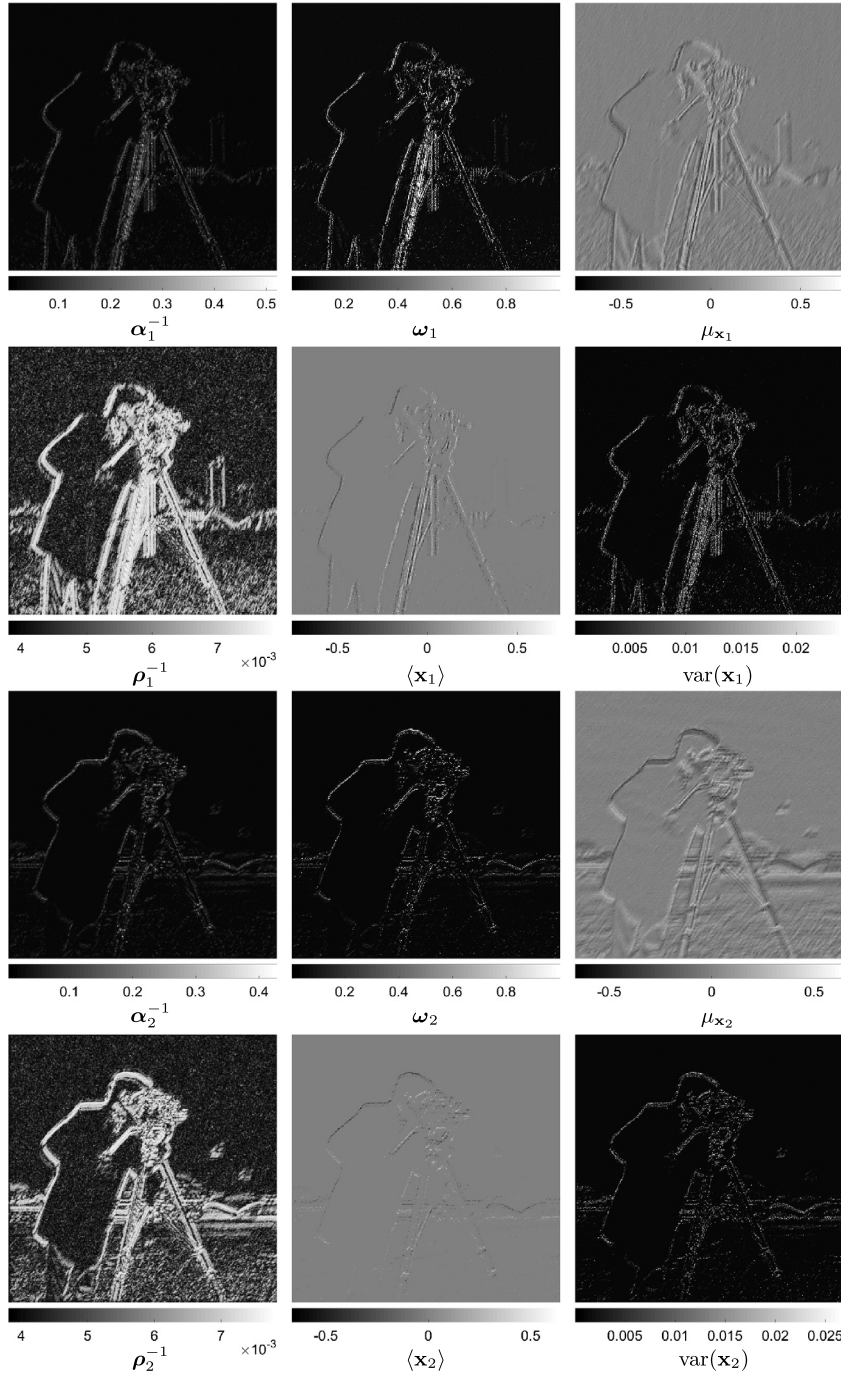
Fig. 1. Values of the parameters and image estimations. $\pi_{\Gamma} = \{0.057, 0.028\}$.

smoothed in the estimation of $\langle \mathbf{x}_{\gamma i} \rangle$, hence removing the noise and making the estimation more robust.

The variance of the full posterior in (24), $\mathrm{var}(x_{\gamma i}) = \omega_{\gamma i} \rho_{\gamma i}^{-1} + (1 - \omega_{\gamma i}) \omega_{\gamma i} \mu_{x_{\gamma i}}^2$, is shown as $\mathrm{var}(\mathbf{x}_1)$ and $\mathrm{var}(\mathbf{x}_2)$ in Fig. 1. When $\omega_{\gamma i}$ is zero, the variance is also zero as the pixel corresponds to the spike in the prior. When $\omega_{\gamma i}$ is one, the variance is $\rho_{\gamma i}^{-1}$ and, when $\omega_{\gamma i}$ is between 0 and 1, the variance is increased being larger with larger absolute values of $\mu_{\gamma i}$ and $\omega_{\gamma i}$ is closer to 0.5. That is, the variance increases as we go away from zero in the estimation and as we are more in doubt whether the pixel comes from a spike or a slab. This behavior is clear if we study the posterior distribution of $\tilde{x}_{\gamma i}$. Marginalizing (24), we obtain

$$q(\tilde{x}_{\gamma i}) = \sum_{s_{\gamma i} \in \{0,1\}} q(s_{\gamma i}) q(\tilde{x}_{\gamma i} | s_{\gamma i})$$
$$= \omega_{\gamma i} \mathcal{N}(\tilde{x}_{\gamma i} | \mu_{x_{\gamma i}}, \rho_{\gamma i}^{-1}) + (1 - \omega_{\gamma i}) \mathcal{N}(\tilde{x}_{\gamma i} | 0, \alpha_{\gamma i}^{-1}), \quad (33)$$

that is, a convex combination of two Gaussians, one centered at $\mu_{x_{\gamma i}}$ and another one centered at zero. The distribution in (33) has the following mean and variance

$$\langle \tilde{x}_{\gamma i} \rangle = \omega_{\gamma i} \mu_{x_{\gamma i}}, \tag{34}$$

$$\mathrm{var}(\tilde{x}_{\gamma i}) = \omega_{\gamma i} \rho_{\gamma i}^{-1} + (1 - \omega_{\gamma i}) \alpha_{\gamma i}^{-1} + (1 - \omega_{\gamma i}) \omega_{\gamma i} \mu_{x_{\gamma i}}^2. \tag{35}$$

Fig. 2 illustrates the result of this combination on selected pixels from different regions of the image. In flat regions, Fig. 2(a), the value of $\omega_{\gamma i}$ is very close to zero and hence the prior pulls
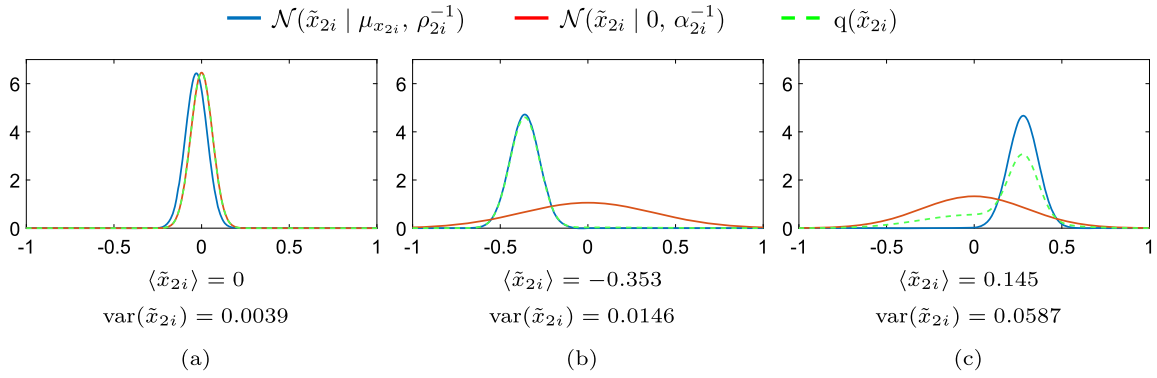
Fig. 2. Behavior of the posterior $q(\tilde{x}_{\gamma i})$ for different pixels in the image for $\gamma = 2$. (a) Pixel in flat area, $\mu_{x_{2i}} = -0.03$, $\rho_{2i}^{-1} = 0.0038$, $\alpha_{2i}^{-1} = 0.0038$, $\omega_{2i} = 0.008$. (b) Edge pixel at the elbow $\mu_{x_{2i}} = -0.36$, $\rho_{2i}^{-1} = 0.0071$, $\alpha_{2i}^{-1} = 0.1429$, $\omega_{2i} = 0.97$. (c) Edge pixel below the elbow $\mu_{x_{2i}} = 0.28$, $\rho_{2i}^{-1} = 0.0073$, $\alpha_{2i}^{-1} = 0.0909$, $\omega_{2i} = 0.58$.
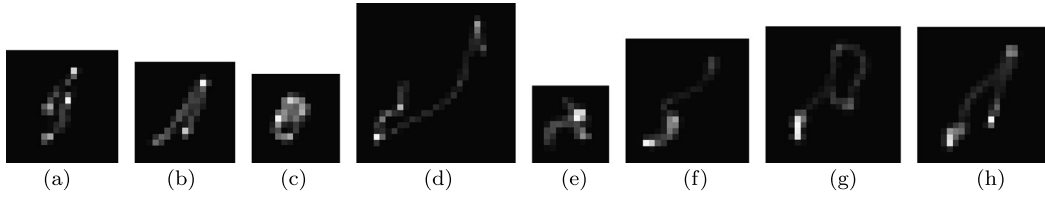

Fig. 3. Blur kernels used in the Levin's dataset.

the value of the pixel to zero, removing the noise in the observation. Note that in this region, the variance is $\alpha_{\gamma i}^{-1}$. In edge regions, where $\omega_{\gamma i}$ is very close to one, however, the outcome is obtained mainly from the observation (see Fig. 2(b)). Note that the variance of the distribution here is close to $\rho_{\gamma i}^{-1}$. There are a few pixels where $\omega_{\gamma i}$ is not clearly 0 or 1, as depicted in Fig. 2(c), and the resulting mean value is obtained from the observed data smoothed by the prior. In this case, the variance is higher than $\rho_{\gamma i}^{-1}$, reflecting our uncertainty on the value of $\tilde{x}_{\gamma i}$. In summary, the spike-and-slab prior finds relevant pixels in image edges, which help to better estimate the blur. Note that, if we followed the MAP approach instead of approximating the full posterior distribution, we would have no uncertainty and the value of $\omega_{\gamma i}$ would be either 0, if $q(\tilde{x}_{\gamma i}|s_{\gamma i} = 0) < q(\tilde{x}_{\gamma i}|s_{\gamma i} = 1)$ or 1 otherwise, and hence the value of $\tilde{x}_{\gamma i}$ would be either 0, if $\omega_{\gamma i} = 0$, or exactly $\mu_{\gamma i}$, if $\omega_{\gamma i} = 1$.

The idea of selecting just a few sharp edges of the image that are good for kernel estimation is also applied in the MAP based method in [22] where an *ad hoc* modification of the degradation model is proposed to select the most active gradients, which are then used for kernel estimation, while the rest of the gradients are set to zero. The regularization parameters, as well as the number of pixels to select are set manually. Other MAP based algorithms use heuristics to prune out small edges such as bilateral filtering [60] or to threshold the edge response [21]. Our method, however, introduces the pixel selection as a prior on the edges and rigorously estimates, within the variational framework, all involved parameters. The proposed model can be extended to consider the same prior variance for different pixels of the image, for instance for pixels on flat regions or edge regions, and prior information relating those parameters can also be introduced into the model formulation by means of hyperpriors. Furthermore, some heuristics such as forcing $\omega_{\gamma i}$ to take a given value (say, zero for saturated pixels of the image) can be considered.

5.2. Comparative study

We first evaluate the performance of the proposed algorithm on Levin's dataset [12]. This dataset consists of a set of 4 images

blurred with 8 different PSFs, displayed in Fig. 3, adding up to a total of 32 images. All images are 256×256 and blur kernel sizes range from 13×13 to 27×27.

The proposed method is compared with the VI algorithm in [31] (VI, Huber Super-Gaussian priors), namely Zhou17, and two recent MAP approaches with different image priors, Pan17 [16] (MAP, dark channel prior) and Perrone15 [19] (MAP, logarithmic image priors). These methods can be considered to represent current state-of-the-art algorithms in BID. Algorithms were named using the last name of the first author plus year of publication. For a fair comparison, once the PSF is estimated with the corresponding method, the final image restoration is obtained by the non-blind restoration method in [53].

We compare the performance of these algorithms based on two reference metrics: PSNR and SSIM. Results are depicted graphically in Figs. 4 (PSNR) and 5 (SSIM). Each graph represents, for each kernel, the results for each method on the 4 images of the dataset. Table 1 shows a summary of the obtained results. For each tested method, the left column reports the mean PSNR values obtained per PSF, this is, the average PSNR over the four different images. Likewise, the right column shows the corresponding SSIM values. The standard deviation of each measure is shown between brackets below the mean value. The proposed algorithm performs better in terms of PSNR than the rest in 4 out of the 8 tested PSFs and SSIM values follow a similar behavior. The overall superiority of the spike-and-slab deconvolution technique is reflected in the average PSNR and SSIM values, which also have a smaller standard deviation than all the competing methods. A one-on-one comparison of the proposed method with Zhou17 shows that in most of the images and kernels tested we perform better, see Figs. 4 and 5, and we obtain a considerable gain in terms of both PSNR and SSIM (7 out of 8 PSFs). In fact, this gain has an average value of $+0.28$ dB and a lower standard deviation, with very small loss in the few cases where the performance is worsened and a mean gain of $+0.51$ dB in the rest. Compared with Perrone15, the proposed method obtains higher PSNR and SSIM values in almost all the cases. Pan17 seems to be the less robust one, since it obtains a standard deviation in PSNR of 2.93 and 0.094 in SSIM. This is
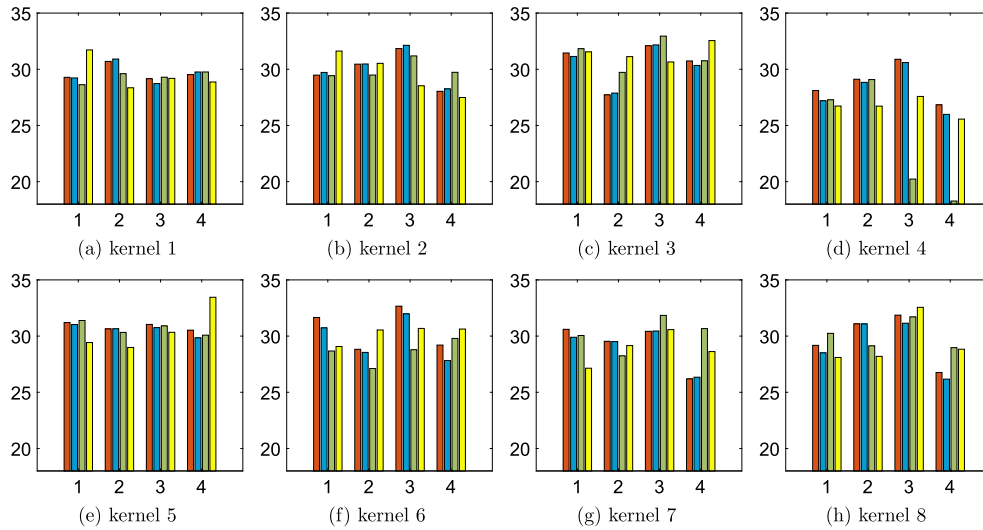
Fig. 4. PSNR values for all images per kernel. Image number is displayed on the horizontal axis. Each bar represents a method: ■ Proposed, ■ Zhou17, ■ Pan17, and ■ Perrone15.
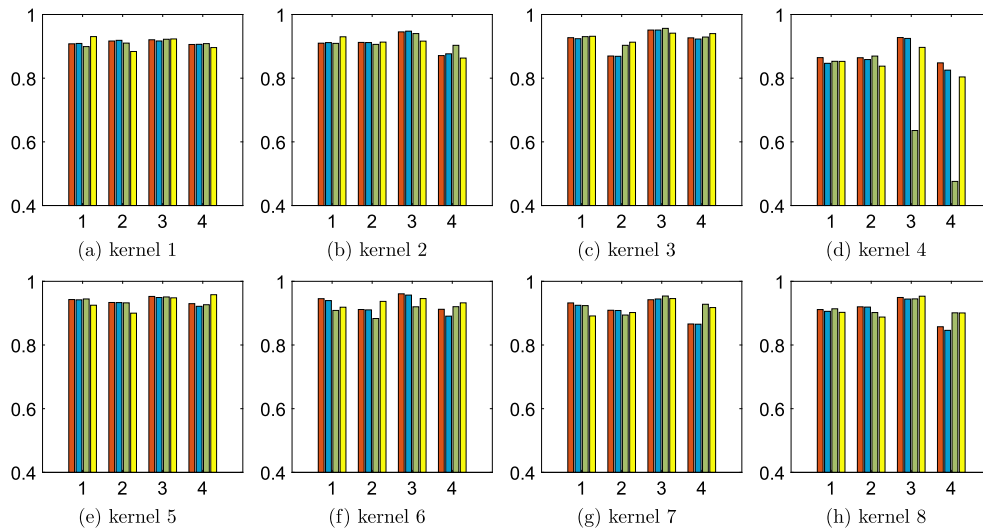

Fig. 5. SSIM values for all images per kernel. Image number is displayed on the horizontal axis. Each bar represents a method: ■ Proposed, ■ Zhou17, ■ Pan17, and ■ Perrone15.

Table 1
Mean (standard deviation) PSNR and SSIM per PSF.

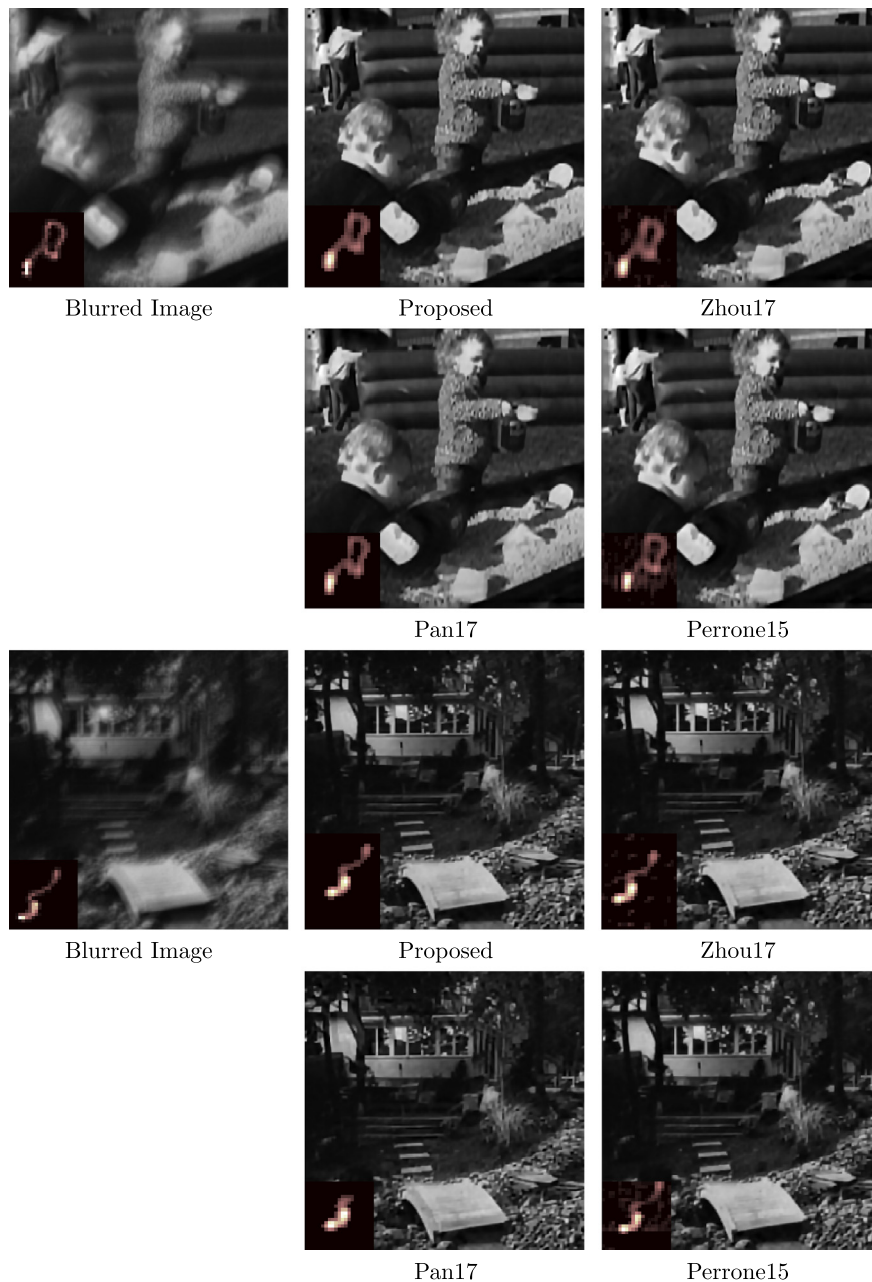| PSF	Proposed		Zhou17		Pan17		Perrone15	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	**29.67**	**0.913**	29.66	**0.913**	29.32	0.910	29.53	0.908
	(0.71)	**(0.007)**	(0.94)	**(0.006)**	(0.50)	(0.010)	(1.50)	(0.022)
2	29.95	0.910	**30.15**	0.912	29.96	**0.915**	29.54	0.906
	(1.60)	(0.031)	**(1.61)**	(0.029)	(0.83)	**(0.017)**	(1.87)	(0.029)
3	30.50	0.919	30.38	0.917	31.32	0.930	**31.47**	**0.931**
	(1.93)	(0.035)	(1.83)	(0.035)	(1.39)	(0.022)	**(0.81)**	**(0.013)**
4	**28.74**	**0.876**	28.16	0.864	23.72	0.708	26.65	0.848
	(1.71)	**(0.035)**	(2.01)	(0.043)	(5.27)	(0.188)	(0.83)	(0.039)
5	**30.85**	**0.940**	30.57	0.936	30.68	0.939	30.55	0.933
	(0.32)	**(0.010)**	(0.012)	(0.51)	(0.59)	(0.011)	(2.01)	(0.026)
6	**30.59**	0.932	29.77	0.924	28.59	0.908	30.23	**0.933**
	(1.87)	(0.025)	(1.85)	(0.030)	(1.50)	(0.018)	(1.42)	**(0.011)**
7	29.19	0.912	29.05	0.911	**30.20**	**0.925**	28.87	0.914
	(2.05)	(0.034)	(1.85)	(0.034)	**(1.50)**	**(0.024)**	(1.42)	(0.024)
8	29.72	0.909	29.23	0.904	**30.01**	**0.915**	29.42	0.911
	(2.27)	(0.038)	(2.38)	(0.042)	**(1.26)**	**(0.021)**	(2.12)	(0.029)
avg.	**29.90**	**0.914**	29.62	0.910	29.23	0.894	29.53	0.911
	(1.64)	**(0.031)**	(1.70)	(0.034)	(2.93)	(0.094)	(1.89)	(0.035)

Fig. 6. Sample of results on Levin's dataset. (Top) Image 1 blurred with kernel 7. (Bottom) Image 2 blurred with kernel 6.

mainly due to the results with kernel 4 where results of Pan17 method are much worse than any other method.

We also show a small sample of the results for visual comparison in Figs. 6 and 7, where the inset kernels have been drawn using a nonlinear color scheme to enhance the lower values' visibility. The proposed method achieves sharp-looking images with no visible artifacts provided the size of the PSF is not too large. Large PSFs affect large areas of the blurred images, which hinders blur estimation, naturally worsening all methods' performance. See, for instance, the results for PSF 4 in Fig. 7 bottom, where the PSF estimations for all methods are less precise and hence the restoration. This is also reflected in the corresponding PSNR and SSIM values in Table 1. However, for this large PSF, the proposed method obtains better results than the competing methods both visually, with sharper edges and less artifacts, and numerically, with higher PSNR and SSIM values. Typically, Zhou17 and Perrone15 produce noisier PSFs, whereas Pan17's method produces a less robust, albeit clean,

estimation of the blur kernel (see kernels in Figs. 6 bottom and 7 bottom). Concretely, regarding the image in Fig. 6 bottom, we clearly notice that the PSF produced by the proposed method is notably cleaner and closer to the original one (depicted with the blurred image in the figure).

The result of a second set of experiments on real blurred images is shown in Figs. 8 and 9. In this scenario, all competing algorithms produce high quality images. The proposed algorithm matches the performance of the other state-of-the-art algorithms with little visual differences, although it produces slightly sharper restorations since it estimates less noisy PSFs. Again, we see that the proposed method produces cleaner PSFs than the methods of Zhou17 and Perrone15. In Fig. 9 we see that, in the central area of the image, all algorithms performed well. However, ringing artifacts appear at the borders of the image (see, for instance, the fire hydrant in the lower part of the image) due to the spatially variant nature of the blur in this image.

Fig. 7. Sample of results on Levin's dataset. (Top) Image 3 blurred with kernel 5. (Bottom) Image 4 blurred with kernel 4.

It is also important to study the impact of the noise level on the blur estimation process. For this purpose, we create a synthetic set of images generated from the Lena image following the degradation model in (1). This is, we convolve this image using one of the kernels from the dataset, kernel 7, and add i.i.d. Gaussian noise with different variances producing a set of images with blurred signal-to-noise ratios (BSNR) of $40, 35, \ldots, 10$ dB. The degraded images along with the estimated filters produced by our method are shown in Fig. 10. For low to mid-high noise levels, 40 to 25 dB, the algorithm produces very accurate reconstructions of the blur kernel. For higher noise levels, BSNR 20 dB, the effect of noise starts to manifest itself. As noise level increases, a greater number of $\omega_{\gamma i}$ shift towards zero, discarding a greater number of potential edge pixels, which, in turn, leads to lower values of π_γ. For very high noise levels, 15 dB and below, the noise hampers the blur estimation, which is gradually aggravated as the BSNR value

drops. We therefore notice a strong robustness of the algorithm to the presence of noise.

To conclude this section we carry out an analysis of the computational time required for the compared algorithms for the estimation of the blur. All algorithms were run with nonoptimized nonparallelized code on Matlab® on an Intel® Core i7-3960X CPU @ 3.30 GHz with 64 GB RAM. The results in Table 2 show the mean execution time over ten runs for a set of degraded images with different image and kernel sizes. Note that the times shown for the proposed algorithm include the time needed to compute the initial estimation using [31]. The proposed algorithm is much faster than the MAP based algorithms in Perrone15 [19], which spends a large amount of time computing the Total Variation Gradient, and only needs two thirds of the time of Pan17 [16], which spends most of the time on the dark channel computation. However, the proposed algorithm is much slower than the variational based method in Zhou17 [31] since it estimates sequentially the
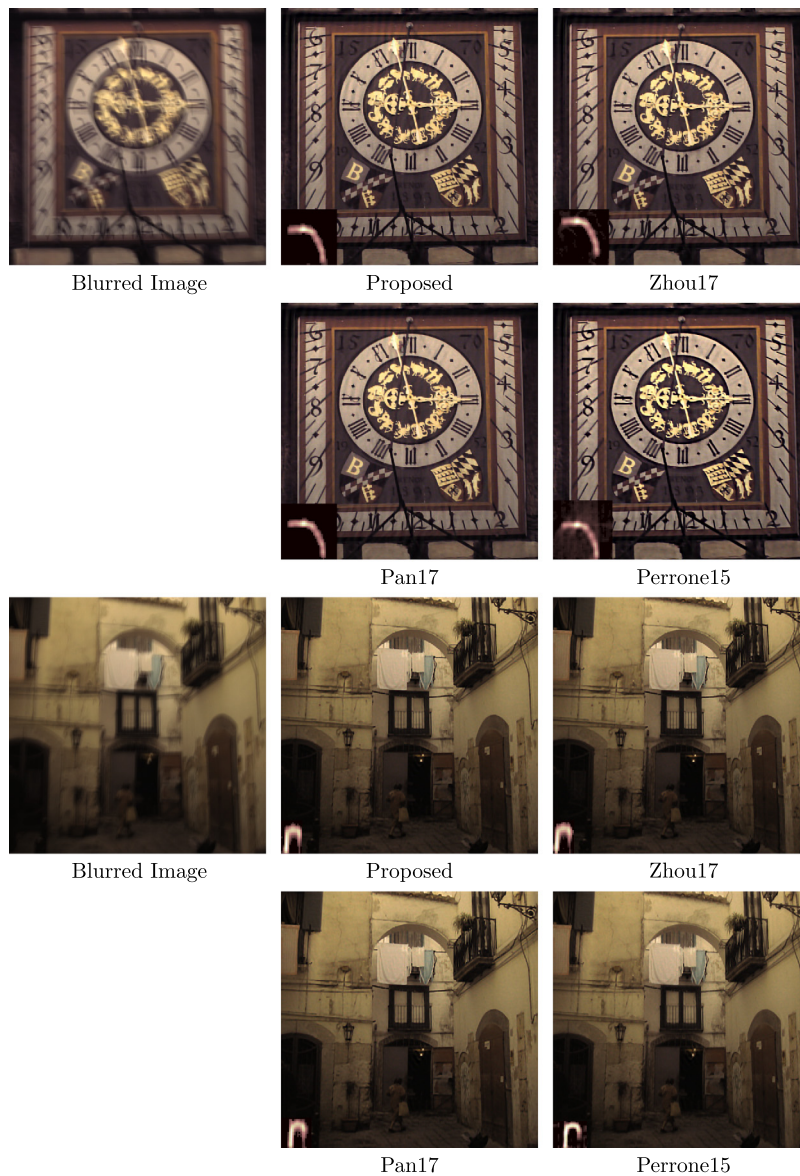
Fig. 8. Comparison of deblurring results for two real and blurred images from the dataset in [61].

image pixels, for all the filtered images, preventing the use of the FFT as in Zhou17 [31] and, hence, slowing down the algorithm. Further studies are needed to clarify if it is possible to estimate all the pixels simultaneously which would allow the use of FFT to perform image estimation.

6. Conclusion and future work

A new BID method formulated in the filter space has been presented. The main novelty of the method is the introduction of the spike-and-slab prior on the high-pass filtered image which allows to shrink image information that is irrelevant for blur estimation. A simple but powerful reparametrization of the spike-and-slab prior allows for the use of variational inference and a sensible factorization allows to capture the multimodality of the true posterior distribution. We have developed an efficient and accurate variational EM algorithm for the blur kernel estimation. The proposed algorithm also estimates automatically all the prior model parameters within the variational framework. Experimental results show the competitiveness of the proposed method. The proposed method provides better kernel estimates on the tested images, re-

sulting in higher PSNR with lower PSNR variance, and requires lower running time than most of its competitors. The method also presents a good tolerance to the noise in the images. A study of the spike-and-slab prior behavior provides a deeper understanding of the pixel selection capabilities of this prior which, using a well-founded theoretical framework, help to better estimate the blur.

Extensions of the proposed method include the estimation of the noise variance, the inclusion of other prior distributions on the model's parameters or the use of heuristics to include or discard pixels within the estimation procedure. For instance, model discrepancies can be handled by setting $\omega_{\gamma i} = 0$ in pixels that do not fit the degradation model, such as saturated pixels. A hyperprior can be introduced to include prior information that relates the $\omega_{\gamma i}$. This would allow, for instance, to consider the same prior variance for pixels on flat regions or edge regions. Additionally, since image estimation is performed pixel-wise, it would be straightforward to adapt the algorithm to handle space-variant blur. In this case, the efficient shift-variant image restoration method in [62,63] can be, in principle, adopted for the blur estimation process.
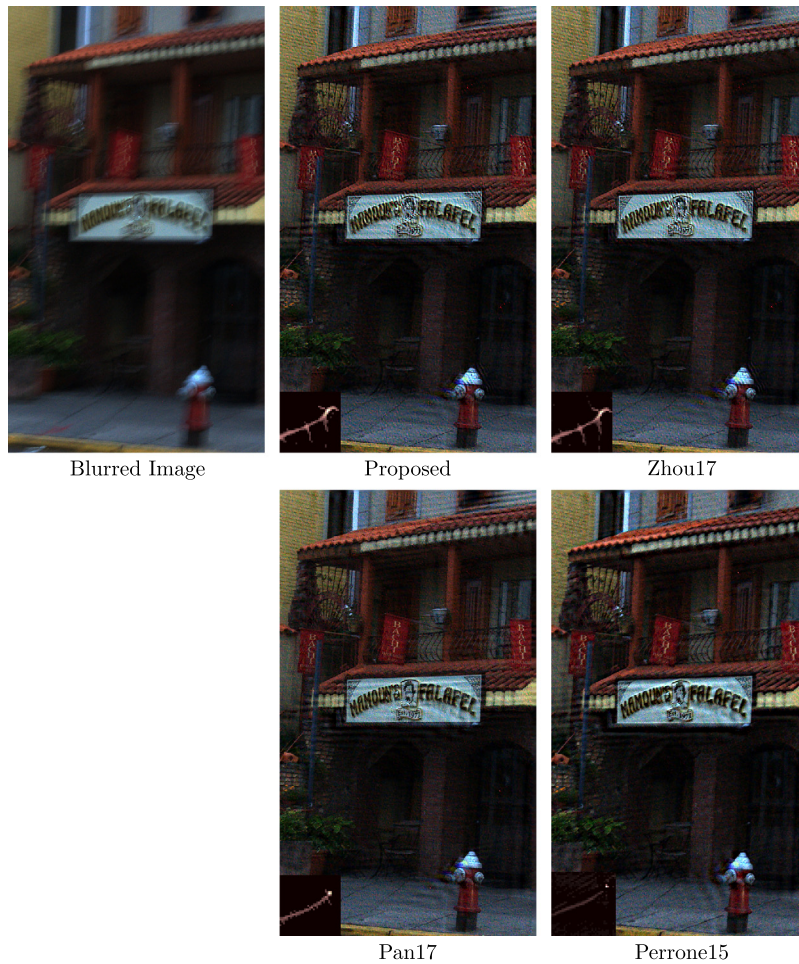
Fig. 9. Comparison of deblurring results on another blurred and noisy image.
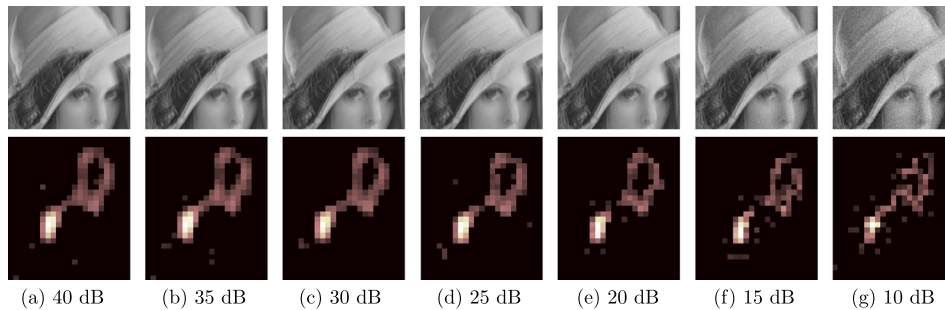

Fig. 10. Degraded images along with the corresponding estimated PSFs for different BSNR values. Zoom in for realistic visualization of the noise intensity in degraded images.

Table 2
Mean CPU time (in seconds) for different BID methods.

Proposed	Pan17	Zhou17	Perrone15
148	220	7	1031

Conflict of interest statement

None declared.

References

[1] P. Ruiz, X. Zhou, J. Mateos, R. Molina, A. Katsaggelos, Variational Bayesian blind image deconvolution: a review, Digit. Signal Process. 47 (2015) 116–127.

[2] M. Bertero, P. Boccacci, Introduction to Inverse Problems in Imaging, CRC Press, 1998.

[3] A. Levin, Y. Weiss, F. Durand, W.T. Freeman, Efficient marginal likelihood optimization in blind deconvolution, in: 2011 IEEE Conf. Comput. Vision and Pattern Recognition (CVPR), 2011, pp. 2657–2664.

[4] L. Xu, S.C. Zheng, J.Y. Jia, Unnatural l0 sparse representation for natural image deblurring, in: 2013 IEEE Conf. Comput. Vision and Pattern Recognition (CVPR), 2013, pp. 1107–1114.

[5] R. Molina, J. Mateos, A.K. Katsaggelos, Blind deconvolution using a variational approach to parameter, image, and blur estimation, IEEE Trans. Image Process. 15 (12) (2006) 3715–3727.

[6] D. Wipf, H. Zhang, Revisiting Bayesian blind deconvolution, J. Mach. Learn. Res. 15 (2014) 3775–3814.

[7] R. Molina, A.K. Katsaggelos, J. Mateos, Bayesian and regularization methods for hyperparameter estimation in image restoration, IEEE Trans. Image Process. 8 (2) (1999) 231–246.

[8] M.R. Banham, A.K. Katsaggelos, Digital image restoration, IEEE Signal Process. Mag. 14 (2) (1997) 24–41.

[9] D. Kundur, D. Hatzinakos, Blind image deconvolution, IEEE Signal Process. Mag. 13 (3) (1996) 43–64.

[10] Y. Yang, N.P. Galatsanos, H. Stark, Projection-based blind deconvolution, J. Opt. Soc. Am. A 11 (9) (1994) 2401–2409.

[11] R. Molina, A.K. Katsaggelos, J. Abad, J. Mateos, A Bayesian approach to blind deconvolution based on Dirichlet distributions, in: 1997 IEEE Int. Conf. Acoust., Speech and Signal Process (ICASSP), 1997, pp. 2809–2812.

[12] A. Levin, Y. Weiss, F. Durand, W.T. Freeman, Understanding and evaluating blind deconvolution algorithms, in: 2009 IEEE Conf. Comput. Vision and Pattern Recognition (CVPR), 2009, pp. 1964–1971.

[13] S. Chaudhuri, R. Velmurugan, R. Rameshan, Blind Image Deconvolution, Springer, 2014.

[14] D. Krishnan, T. Tay, R. Fergus, Blind deconvolution using a normalized sparsity measure, in: 2011 IEEE Conf. Comput. Vision and Pattern Recognition (CVPR), 2011, pp. 233–240.

[15] W. Zuo, D. Ren, S. Gu, L. Lin, L. Zhang, Discriminative learning of iteration-wise priors for blind deconvolution, in: 2015 IEEE Conf. Comput. Vision and Pattern Recognition (CVPR), 2015, pp. 3232–3240.

[16] J. Pan, D. Sun, H. Pfister, M.H. Yang, Deblurring images via dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2018) 2315–2328.

[17] D. Krishnan, R. Fergus, Fast image deconvolution using hyper-Laplacian priors, in: Advances in Neural Inform. Process. Sys. (NIPS), 2009, pp. 1033–1041.

[18] J. Kotera, F. Šroubek, P. Milanfar, Blind deconvolution using alternating maximum a posteriori estimation with heavy-tailed priors, in: 2013 Int. Conf. Comp. Anal. Imag. and Patterns (CAIP), 2013, pp. 59–66.

[19] D. Perrone, R. Diethelm, P. Favaro, Blind deconvolution via lower-bounded logarithmic image priors, in: International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), 2015, pp. 112–125.

[20] W.-Z. Shao, H.-B. Li, M. Elad, Bi-l0-l2-norm regularization for blind motion deblurring, J. Vis. Commun. Image Represent. 33 (2015) 42–59.

[21] L. Xu, J. Jia, Two-phase kernel estimation for robust motion deblurring, in: K. Daniilidis, P. Maragos, N. Paragios (Eds.), Computer Vision – ECCV 2010, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 157–170.

[22] D. Gong, M. Tan, Y. Zhang, A.v.d. Hengel, Q. Shi, Blind image deconvolution by automatic gradient activation, in: 2016 IEEE Conf. Comput. Vision and Pattern Recognition (CVPR), 2016, pp. 1827–1836.

[23] S. Tao, W. Dong, Z. Tang, Q. Wang, Blind image deconvolution using the Gaussian scale mixture fields of experts prior, in: 2017 International Conference on Progress in Informatics and Computing (PIC), 2017, pp. 190–195.

[24] Y. Cai, J. Pan, Z. Su, Blind image deblurring via salient structure detection and sparse representation, in: S. Satoh (Ed.), Image and Video Technology, Springer International Publishing, Cham, 2018, pp. 283–299.

[25] L. Li, J. Pan, W.-S. Lai, C. Gao, N. Sang, M.-H. Yang, Learning a discriminative prior for blind image deblurring, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 6616–6625.

[26] M. Ljubenović, M.A.T. Figueiredo, Blind image deblurring using class-adapted image priors, in: 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 490–494.

[27] R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, W.T. Freeman, Removing camera shake from a single photograph, ACM Trans. Graph. 25 (3) (2006) 787–794.

[28] C.L. Likas, N.P. Galatsanos, A variational approach for Bayesian blind image deconvolution, IEEE Trans. Signal Process. 52 (8) (2004) 2222–2233.

[29] S. Babacan, R. Molina, M. Do, A.K. Katsaggelos, Bayesian blind deconvolution with general sparse image priors, in: 2012 Eur. Conf. Comput. Vision (ECCV), 2012, pp. 341–355.

[30] 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.

[31] X. Zhou, M. Vega, F. Zhou, R. Molina, A.K. Katsaggelos, Fast Bayesian blind deconvolution with Huber super Gaussian priors, Digit. Signal Process. 60 (2017) 122–133.

[32] R. Giri, B. Rao, Type I and type II Bayesian methods for sparse signal recovery using scale mixtures, IEEE Trans. Signal Process. 64 (2016) 3418–3428.

[33] R. Giri, B. Rao, Learning distributional parameters for adaptive Bayesian sparse signal recovery, IEEE Comput. Intell. Mag. 11 (2016) 14–23.

[34] X. Zhou, J. Mateos, F. Zhou, R. Molina, A.K. Katsaggelos, Variational Dirichlet blur kernel estimation, IEEE Trans. Image Process. 24 (2015) 5127–5139.

[35] M.K. Titsias, M. Lázaro-Gredilla, Spike and slab variational inference for multi-task and multiple kernel learning, in: Advances in Neural Inform. Process. Sys. (NIPS), 2011, pp. 2339–2347.

[36] L. Chaari, J.-Y. Toumeret, H. Batatia, Sparse Bayesian regularization using Bernoulli–Laplacian priors, in: Eur. Signal Process. Conf. (EUSIPCO), 2013, pp. 1–5.

[37] L. Chaari, J.Y. Toumeret, C. Chaux, Sparse signal recovery using a Bernoulli generalized Gaussian prior, in: Eur. Signal Process. Conf. (EUSIPCO), 2015, pp. 1711–1715.

[38] H. Ishwaran, J.S. Rao, Spike and slab variable selection: frequentist and Bayesian strategies, Ann. Stat. 33 (2) (2005) 730–773.

[39] S.A. Zilqurnain Naqvi, Efficient Sparse Bayesian Learning Using Spike-and-Slab Priors, Ph.D. thesis, Purdue University, 2016.

[40] G. Malsiner-Walli, H. Wagner, Comparing spike and slab priors for Bayesian variable selection, Aust. J. Stat. 40 (2011) 241–264.

[41] X. Huang, J. Wang, F. Liang, A variational Bayesian algorithm for variable selection, ArXiv e-prints.

[42] X. Lu, Y. Yuan, P. Yan, Sparse coding for image denoising using spike and slab prior, Neurocomputing 106 (2013) 12–20.

[43] S. Mohamed, K.A. Hellerand, Z. Ghahramani, Bayesian and l_1 approaches for sparse unsupervised learning, in: Int. Conf. Mach. Learn. (ICML), 2012, pp. 751–758.

[44] B. Lin, X. Tao, S. Li, L. Dong, J. Lu, Variational Bayesian image fusion based on combined sparse representations, in: 2016 IEEE Int. Conf. Acoust., Speech and Signal Process (ICASSP), 2016, pp. 1432–1436.

[45] M. Zhou, H. Chen, L. Ren, G. Sapiro, L. Carin, J.W. Paisley, Non-parametric Bayesian dictionary learning for sparse image representations, in: Advances in Neural Inform. Process. Sys. (NIPS), 2009, pp. 2295–2303.

[46] V. Ročková, E.I. George, Negotiating multicollinearity with spike-and-slab priors, Metron 72 (2014) 217–229.

[47] J.M. Hernández-Lobato, D. Hernández-Lobato, A. Suárez, Expectation propagation in linear regression models with spike-and-slab priors, Mach. Learn. 99 (2015) 437–487.

[48] H. Liao, M.K. Ng, Blind deconvolution using generalized cross-validation approach to regularization parameter estimation, IEEE Trans. Image Process. 20 (2011) 670–680.

[49] Y.W. Wen, R.H. Chan, Parameter selection for total-variation-based image restoration using discrepancy principle, IEEE Trans. Image Process. 21 (2012) 1770–1781.

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

[51] J. Serra, J. Mateos, R. Molina, A. Katsaggelos, Spike and slab variational inference for blind image deconvolution, in: IEEE International Conference on Image Processing, ICIP 2017, Beijing, China, 2017, pp. 3765–3769.

[52] J. Serra, J. Mateos, R. Molina, A. Katsaggelos, Parameter estimation in spike and slab variational inference for blind image deconvolution, in: European Signal Processing Conference, EUSIPCO 2017, Kos, Greece, 2017, pp. 1540–1544.

[53] X. Zhou, R. Molina, F. Zhou, A.K. Katsaggelos, Fast iteratively reweighted least squares for l_p regularized image deconvolution and reconstruction, in: 2014 IEEE Int. Conf. Image Process (ICIP), 2014, pp. 1783–1787.

[54] A. Levin, R. Fergus, F. Durand, W.T. Freeman, Image and depth from a conventional camera with a coded aperture, ACM Trans. Graph. 26 (3) (2007) 70.

[55] X. Zhou, F. Zhou, X. Bai, B. Xue, A boundary condition based deconvolution framework for image deblurring, J. Comput. Appl. Math. 261 (2014) 14–29.

[56] X. Liu, M. Tanaka, M. Okutomi, Single-image noise level estimation for blind denoising, IEEE Trans. Image Process. 22 (12) (2013) 5226–5237.

[57] C. Sutour, J.F. Aujol, C.A. Deledalle, Automatic estimation of the noise level function for adaptive blind denoising, in: 2016 24th European Signal Processing Conference (EUSIPCO), 2016, pp. 76–80.

[58] M. Rakhshanfar, M.A. Amer, Estimation of Gaussian, Poissonian–Gaussian, and processed visual noise and its level function, IEEE Trans. Image Process. 25 (9) (2016) 4172–4185.

[59] L. Dong, J. Zhou, Y.Y. Tang, Noise level estimation for natural images based on scale-invariant kurtosis and piecewise stationarity, IEEE Trans. Image Process. 26 (2) (2017) 1017–1030.

[60] S. Cho, S. Lee, Fast motion deblurring, ACM Trans. Graph. 28 (5) (2009) 145:1–145:8.

[61] R. Köhler, M. Hirsch, B. Mohler, B. Schölkopf, S. Harmeling, Recording and playback of camera shake: benchmarking blind deconvolution with a real-world database, in: A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, C. Schmid (Eds.), Computer Vision – ECCV 2012, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 27–40.

[62] D. Miraut, J. Portilla, Efficient shift-variant image restoration using deformable filtering (part I), EURASIP J. Adv. Signal Process. 2012 (2012) 100.

[63] D. Miraut, J. Ballé, J. Portilla, Efficient shift-variant image restoration using deformable filtering (part II): PSF field estimation, EURASIP J. Adv. Signal Process. 2012 (2012) 193.

Juan G. Serra received the degree in telecommunications engineering from the Universitat Politècnica de Valéncia in 2014 and the M.S. degree in data science and computer engineering from the University of Granada in 2016, where he is currently pursuing the Ph.D. degree, under the supervision of Prof. Molina, being a member of the Visual Information Processing Group, Department of Computer Science and Artificial Intelligence. His research interests focus on the use of Bayesian modeling and inference to solve different inverse problems related to dictionary learning, image restoration and machine learning. During his research, he has addressed several problems, such as blind image deconvolution, image denoising, and inpainting and multispectral image classification.

Javier Mateos received the degree in computer science in 1991 and the Ph.D. degree in computer science in 1998, both from the University

of Granada. He joined the Department of Computer Science and Artificial Intelligence, University of Granada, in 1992 where he is currently a Professor. He is the deputy head of the Department of Computer Science and Artificial Intelligence since 2014. He is coauthor of "Superresolution of Images and Video" (Claypool, 2006) and "Multispectral Image fusion Using Multiscale and Super-resolution Methods" (VDM Verlag, 2011). He received the IEEE International Conference on Signal Processing Algorithms, Architectures, Arrangements, and Applications best paper award (2013) and was finalist for the IEEE International Conference on Image Processing Best Student Paper Award (2010). He is conducting research on image and video processing, including image restoration, image and video recovery, super-resolution from (compressed) stills and video sequences, pansharpening and image classification. Dr. Mateos is a member of the Asociación Española de Reconocimento de Formas y Análisis de Imágenes (AERFAI) and International Association for Pattern Recognition (IAPR) and senior member of the IEEE. He serves as an editor of "Digital Signal Processing" (2011–present) and as an Associate Editor for the "IEEE Transactions on Image Processing" (2014–2017).

Rafael Molina received the degree in mathematics (statistics) and the Ph.D. degree in optimal design in linear models from the University of Granada, Granada, Spain, in 1979 and 1983, respectively. He became Professor of Computer Science and Artificial Intelligence at the University of Granada, Granada, Spain, in 2000. He is the former Dean of the Computer Engineering School at the University of Granada (1992–2002) and Head of the Computer Science and Artificial Intelligence department of the University of Granada (2005–2007). His research interest focuses mainly on using Bayesian modeling and inference in problems like image restoration (applications to astronomy and medicine), superresolution of images and video, blind deconvolution, computational photography, source recovery in medicine, compressive sensing, low-rank matrix decomposition, machine learning, active learning, classification, fusion, and crowdsourcing. Prof. Molina serves as an Associate Editor of Applied Signal Processing (2005–2007); the IEEE Transactions on Image Processing (2010–2014); and Progress in Artificial Intelligence (2011–present); and an Area Editor of Digital Signal Processing (2011–present). He is the recipient of an IEEE International Conference on Image Processing Paper Award (2007), an ISPA Best Paper Award (2009), and an European Signal Processing Conference award (2013). He is a coauthor of a paper awarded the runner-up prize at Reception for early-stage researchers at the House of Commons.

Aggelos K. Katsaggelos received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Greece, in 1979, and the M.S. and Ph.D. degrees in Electrical Engineering from the Georgia Institute of Technology, in 1981 and 1985, respectively. In 1985, he joined the Department of Electrical Engineering and Computer Science at Northwestern University, where he is currently a Professor holder of the Joseph Cummings chair. He was previously the holder of the Ameritech Chair of Information Technology and the AT&T chair. He is also a member of the Academic Staff, NorthShore University Health System, an affiliated faculty at the Department of Linguistics and he has an appointment with the Argonne National Laboratory. He has published extensively in the areas of multimedia signal processing and communications (over 250 journal papers, 600 conference papers and 40 book chapters) and he is the holder of 25 international patents. He is the co-author of Rate-Distortion Based Video Compression (Kluwer, 1997), Super-Resolution for Images and Video (Claypool, 2007), Joint Source-Channel Video Transmission (Claypool, 2007), and Machine Learning Refined (Cambridge University Press, 2016). He has supervised 56 Ph.D. theses so far. Among his many professional activities Prof. Katsaggelos was Editor-in-Chief of the IEEE Signal Processing Magazine (1997–2002), a BOG Member of the IEEE Signal Processing Society (1999–2001), a member of the Publication Board of the IEEE Proceedings (2003–2007), and a Member of the Award Board of the IEEE Signal Processing Society. He is a Fellow of the IEEE (1998), SPIE (2009), EURASIP (2017), and OSA (2018). He is the recipient of the IEEE Third Millennium Medal (2000), the IEEE Signal Processing Society Meritorious Service Award (2001), the IEEE Signal Processing Society Technical Achievement Award (2010), an IEEE Signal Processing Society Best Paper Award (2001), an IEEE ICME Paper Award (2006), an IEEE ICIP Paper Award (2007), an ISPA Paper Award (2009), and a EUSIPCO paper award (2013). He was a Distinguished Lecturer of the IEEE Signal Processing Society (2007–2008).