Digital Signal Processing 47 (2015) 116–127

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Variational Bayesian Blind Image Deconvolution: A review ☆

Pablo Ruiz [a], Xu Zhou [b,*], Javier Mateos [a], Rafael Molina [a], Aggelos K. Katsaggelos [c]

[a] Departamento de Ciencias de la Computación e I. A. E.T.S. Ing. Informática y Telecomunicación, Universidad de Granada, 18071 Granada, Spain
[b] Image Processing Center, Beihang University, 100191 Beijing, China
[c] Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208, USA

ARTICLE INFO

Article history:
Available online 4 May 2015

Keywords:
Blind deconvolution
Image deblurring
Image restoration
Variational Bayesian
Bayesian modeling

ABSTRACT

In this paper we provide a review of the recent literature on Bayesian Blind Image Deconvolution (BID) methods. We believe that two events have marked the recent history of BID: the predominance of Variational Bayes (VB) inference as a tool to solve BID problems and the increasing interest of the computer vision community in solving BID problems. VB inference in combination with recent image models like the ones based on Super Gaussian (SG) and Scale Mixture of Gaussians (SMG) representations have led to the use of very general and powerful tools to provide clear images from blurry observations. In the provided review emphasis is paid on VB inference and the use of SG and SMG models with coverage of recent advances in sampling methods. We also provide examples of current state of the art BID methods and discuss problems that very likely will mark the near future of BID.

1. Introduction

Thousands of millions of pictures are taken everyday. If the claim in [1] is right, 880 billion photos were taken in 2014. Every minute, 27,800 pictures are uploaded to Instagram, 208,300 photos are uploaded to Facebook and more than one thousand to Flickr, and the trend, with a digital camera in every mobile phone, is probably exponentially increasing. The quality of these pictures varies widely from professional to amateur, in which case in many instances the images are taken under adverse conditions, such as low lighting or with motion between the camera and the scene, thus resulting in blurred images. While in some cases the introduction of blur in photography is intentional, being a powerful element of visual aesthetics, in most cases it is an undesirable effect degrading the quality of the image. Examples of the intentional introduction of blur includes the silky water effect obtained by using a long exposure when photographing a water flow (Fig. 1(a)), the bokeh effect obtained in parts of the scene lying outside the depth of field (Fig. 1(b)) and used to focus the attention of the viewer on a specific subject, or the motion blur effect (Fig. 1(c)) used to provide a sense of speed. Unintentional blur is caused by a number of causes, the most important ones being: camera or subject motion while the shutter is open (Fig. 1(d)) which leads to motion blur, out-of-focus (Fig. 1(e)) that blurs the whole the image or relevant parts of it or, simply, the presence of the atmosphere (Fig. 1(f)) as is the case with astrophotography.

Not only commercial photography is affected by blur. Modern science makes an intensive use of images in areas such as astronomy, remote sensing, medical imaging and microscopy and, in all of them, imperfections and characteristics of the capture system lead to images degraded during the observation process by blur, noise, and other degradations that diminish the quality and, hence, the value of the captured images.

Image deconvolution is a mature topic that aims at recovering the underlying original image from its blurred and noisy observations. Sometimes, the blur is completely or partially known or can be estimated prior to the deconvolution process. For instance, in astronomical imaging, an accurate representation of the blur can be obtained by imaging a single star first before photographing the astronomical object of interest. In contrast, blind image deconvolution (BID) tackles the restoration problem without knowing the blur in advance, leading to one of the most challenging image processing problems, since many combinations of blur and "true" image can produce the observed image. To start with, deconvolution is an ill posed problem in the Hadamard sense [2], that is, small variations in the data result in large variations in the solution. The problem is exacerbated in the BID problem, since in

☆ This paper has been partially supported by the Spanish Ministry of Economy and Competitiveness under project TIN2013-43880-R, the European Regional Development Fund (FEDER), the CEI BioTic at the Universidad de Granada, the National Natural Science Foundation of China (61233005) and the Department of Energy (DE-NA0002520).

* Corresponding author.
E-mail addresses: mataran@decsai.ugr.es (P. Ruiz), xuzhou@buaa.edu.cn (X. Zhou), jmd@decsai.ugr.es (J. Mateos), rms@decsai.ugr.es (R. Molina), aggk@eecs.northwestern.edu (A.K. Katsaggelos).

http://dx.doi.org/10.1016/j.dsp.2015.04.012
1051-2004/

Fig. 1. Blurred pictures due to intentional blur: (a) silky water effect by Geraint Rowland (https://www.flickr.com/photos/geezaweezer/15327097294), (b) bokeh by Rodrigo Gomez (https://www.flickr.com/photos/rgomez74/2970906336), (c) motion blur by Ernest (https://www.flickr.com/photos/viernest/3380560365). Blurred pictures due to unintentional blur: (d) camera motion by tunguska (https://www.flickr.com/photos/tunguska/103472115), (e) out of focus by Nacho (https://www.flickr.com/photos/gonmi/8193430914), (f) atmosphere by Mike Durkin (https://www.flickr.com/photos/madmiked/43831827).

addition, small variations in the estimated blur can lead to large variations in the restored image.

BID is an underdetermined nonlinear inverse problem, which requires the estimation of many more unknown variables than the available observed data. To find meaningful solutions, not only prior information about the unknowns is crucial, but also a good and sound estimation approach. In this paper, we provide a comprehensive survey of BID methods reported since the publication of the review [3], with a focus on Bayesian approaches. In our opinion, since the publication of [3], Variational Bayes (VB) inference has emerged as a dominant approach for the solution of BID problems. VB inference in combination with recently introduced image models, like the ones based on Super Gaussian (SG) and Scale Mixture of Gaussian (SMG) representation, has led to the development of very general and powerful tools to obtain clear images from blurry observations. We review the recent BID literature with an emphasis on VB inference and the use of SG and SMG models but without ignoring recent advances in sampling methods. We also provide examples of current state of the art BID methods and discuss problems that very likely will mark the near future of BID. The paper is organized as follows. In Section 2, we briefly introduce the BID problem as well as the prior models. Section 3 shows the variational Bayesian methodology and its advantages over other inference approaches. We also present two representation models for variational inference, followed by the final BID algorithm. Section 4 discusses some important outstanding challenges regarding the applications of VB based BID methods and BID as a whole research field. Experimental results are presented in Section 5.

2. Bayesian problem formulation

2.1. Bayesian framework for BID

In BID the image formation model is usually assumed to be:

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

where $\mathbf{y} \in \mathbb{R}^N$ is the observed blurred image (a column vector of N pixels), $\otimes$ represents the convolution operation, $\mathbf{x} \in \mathbb{R}^N$ is the unknown original image, $\mathbf{H} \in \mathbb{R}^{N \times N}$ is the convolution matrix obtained from the also unknown blur kernel $\mathbf{h} \in \mathbb{R}^K$ and $\mathbf{n} \in \mathbb{R}^N$ is a noise term which is assumed to be i.i.d. Gaussian with variance β^{-1}. As discussed in Section 4.4, other degradation models than the Linear and Spatially Invariant model above are also utilized.

Notice that although the BID problem is defined here in the image domain, it can also be easily formulated in transformed domains, such as the derivative, wavelet, and curvelet domains. The use of the filter space has gained popularity recently, however, there are still some open questions which need to be addressed before deciding which one is the right domain to work on, see Section 4.1.

From a Bayesian perspective, given the observed blurred image $\mathbf{y}$, the goal is to infer the latent (hidden) variables $\mathbf{z} = \{\mathbf{x}, \mathbf{h}\}$ and possibly the model parameters denoted by Ω. The image degradation model in Eq. (1) can be written as:

$$\mathrm{p}(\mathbf{y}|\mathbf{z}, \beta) = \mathcal{N}(\mathbf{y}|\mathbf{H}\mathbf{x}, \beta^{-1}\mathbf{I}), \tag{2}$$

where β is the precision parameter of the observation model, and possibly one of the model parameters to be estimated.

It is well known that the inverse problem of Eq. (1) is ill-posed [3]. Therefore, additional information on the latent variables and model parameters must be provided. The Bayesian paradigm introduces this necessary information for the BID problem as a prior distribution $\mathrm{p}(\mathbf{z}|\Omega)$, which models the information on $\mathbf{z}$, and a prior $\mathrm{p}(\Omega)$ on the model parameters. Sometimes the prior on the model parameters is called hyperprior and the elements of Ω are called hyperparameters.

With these ingredients, the global modeling of the BID problem can be written as

$$\mathrm{p}(\mathbf{z}, \Omega, \mathbf{y}) = \mathrm{p}(\mathbf{y}|\mathbf{z}, \Omega)\mathrm{p}(\mathbf{z}|\Omega)\mathrm{p}(\Omega). \tag{3}$$

Before describing how inference is performed, we will now review the image, blur and hyperparameters priors proposed for the BID problem since the publication of [3].

2.2. Image prior models

An unquestionable landmark on the recent history of BID is the paper by Miskin and MacKay [5]. In that work the authors propose the use of a mixture of Laplacians to restore cartoon images and utilize, for the first time in the BID literature, VB inference (to be described later) to restore the observed image. Later, Likas and Galatsanos [6] proposed a Gaussian prior to impose smoothness on the image and blur, see also [7], and Fergus *et al.* [8] proposed a mixture-of-Gaussians (MOG) to impose sparsity.

The 2007 Bishop *et al.* [3] review on BID describes, among others, classical prior models such as Conditional Autoregression (CAR) or Simultaneous Autoregression (SAR) used by Molina *et al.* [9] to impose smoothness, or Total Variation proposed by Rudin, Osher and Fatemi [10] to impose piecewise-smoothness. The TV prior model has been frequently used in BID, see for instance [11–15], see also [16] and [17]. Fergus *et al.* [8] represents the first publication on the use of filtered versions of the original image to estimate image and blur. The use in [12] of majorization methods with variational inference and diagonal covariance approximation led to a new way to approach BID in image processing (not widely acknowledged in the computer vision community). As we will see in the following, the TV prior used in [12] is a particular case of the use of Super Gaussian Distributions in BID.

Since the influential work of Fergus *et al.* [8], sparse prior models have attracted the attention of BID researchers and are, in our opinion, rightly considered to be the state of the art representation in filtered domains. It is a well known fact that when high-pass filters are applied to natural images, the resulting coefficients are sparse; i.e., most of the coefficients are zero or very small while only a small number of them are large (e.g., at the edges). This is a very important characteristic that should be taken into account when restoring natural images.

The ℓ_p prior has been used in a large number of works such [12,13,18–21]. They use a prior distribution based on the minimization of quasi-norms $\|\cdot\|_p^p$ with $0 < p \leq 1$. Levin *et al.* [18] suggest the use of p in the range $[0.6, 0.8]$ for natural images.

The Super-Gaussianity property presented by Palmer in [22], was used in Babacan *et al.* [23] as the building block to propose a general representation for sparse priors. As we will see, almost all previous and very recently proposed prior models can be represented using SG. This representation is used in the same work [23] to introduce two new image priors log and exp. Recent models like the one proposed by Zhang and Wifp [24], or the Student-t prior recently proposed by Mohammad-Djafari [25] are particular cases of SG distributions.

2.2.1. Sparse general representation

A probability distribution is considered to be sparse when it is Super Gaussian (SG) [22], i.e., compared to the Gaussian distribution, it has heavier tails, it is more peaked, and has a positive excess kurtosis. These distributions are referred to as sparse since most of the distribution mass is located around zero (hence strongly favoring zero values), but the probability of occurrence of large signal values is higher compared to the Gaussian distribution.

Babacan *et al.* [23] propose the use of the following general framework to define the prior model either in the image or the filter space. First they consider L high-pass filters $\{f_\gamma\}_{\gamma=1}^L$ (such as derivatives, wavelets, curvelets, etc.) and define

$$x_\gamma = f_\gamma \otimes \mathbf{x}, \quad \gamma = 1, \ldots, L. \tag{4}$$

Using these filters on the real underlying image the following prior model in the image space can be defined

$$\mathrm{p}(\mathbf{x}) \propto \prod_{\gamma=1}^{L} \prod_{i=1}^{N} \exp\left(-\alpha_\gamma \rho(x_\gamma(i))\right), \tag{5}$$

where $\rho(\cdot)$ is an energy function symmetric around zero with $\rho(\sqrt{s})$ increasing and concave for $s \in (0, \infty)$ [22] and α_γ a scale parameter.

Alternatively, the filtered original images can be assumed to be independent. The following set of independent priors is then considered

$$\mathrm{p}(x_\gamma) \propto \prod_{i=1}^{N} \exp\left(-\rho(x_\gamma(i))\right), \quad \gamma = 1, \ldots, L. \tag{6}$$

In this case, a set of blurred and noisy observations can be defined, associated with the filtered original images

$$y_\gamma = f_\gamma \otimes \mathbf{y} = \mathbf{h} \otimes f_\gamma \otimes \mathbf{x} + f_\gamma \otimes \mathbf{n} = \mathbf{h} \otimes x_\gamma + n_\gamma, \tag{7}$$

where n_γ is assumed to be Gaussian independent noise with precision β. It is important to note that the observations y_γ, $\gamma = 1, \ldots, L$, are assumed to be independent and they provide information on the blur but not exactly on $\mathbf{x}$ but on its filtered versions.

Notice that the most popular recent prior models, such as TV, ℓ_p, or MOG are Super Gaussian distributions (see Fig. 2 for some examples), and therefore can be represented using Eq. (5). Notice also in Fig. 2, that log enforces sparsity very strongly due to its infinite peak at zero and heavy tails.

A sub-class of Super Gaussian distributions is the so called Scale Mixture of Gaussians (SMG), proposed by Andrews and Mallows [26] and used as a general framework for BID in Babacan *et al.* [23]. Here, associated with each filter γ and each pixel i we have

$$\mathrm{p}(x_\gamma(i)) = \int \mathrm{p}(x_\gamma(i)|\xi_\gamma(i))\mathrm{p}(\xi_\gamma(i))\mathrm{d}\xi_\gamma(i), \tag{8}$$

where $\mathrm{p}(x_\gamma(i)|\xi_\gamma(i))$ is a Gaussian distribution with precision $\xi_\gamma(i)$. This model can also benefit from the introduction of a global scale parameter α_γ in Eq. (5).

SMG requires complete monotonicity of $\mathrm{p}(\sqrt{s})$, i.e., $(-1)^n \mathrm{p} \times (\sqrt{s})^n \geq 0$ must be satisfied for all $n = 0, 1, 2, \ldots$. As can be seen in [22], this representation is more strict, in the sense that fewer classes of sparse priors can be represented with it than using Eq. (5). Finding $\mathrm{p}(\xi_\gamma(i))$ is in general a difficult task; however, as we will see in Section 3 its full knowledge is not needed for our purposes. One example of SMG is the Student-t prior proposed by Mohammad-Djafari [25]. It is clear that an MOG is an SMG model and that spike and slab distributions on $z \in \mathbb{R}$, $\mathrm{p}(z) = \lambda\delta(z) + (1-\lambda)\mathcal{N}(z|0, \sigma^2)$ [27] are the limit of MOG models with two components, one of them with very small variance. Inference with these models is complicated due to the image size; however variational inference can still be carried out, see [28]. Notice that sparse promoting spike and slab and Bernoulli–Gaussian [27,29] priors will very likely receive more attention by the BID community especially when estimating the blur in the filter space.

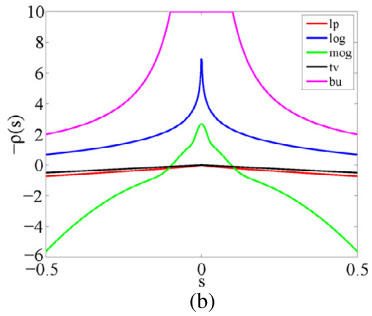
2.3. Blur models

Although the above described prior models were proposed for the image, all of them can also be used for the blur as well. The BID literature also contains specific blur models which we now describe. Molina *et al.* [9] propose a Dirichlet prior for kernel modeling. Since the curvelet representation can take into account both the continuity and sparsity of the motion blur kernel, Cai *et al.* [30] suggest the use of this representation for this type of blur. Oh and Kim [31] propose a piecewise-linear model for motion blur in order to reduce the dimensionality of the solution space and make the kernel estimation process more robust.

Based on the assumption that the power spectrum of natural images drops quadratically as the frequency increases Goldstein

Prior	$\rho(s)$
TV	s
ℓ_p	$\frac{1}{p}\lvert s\rvert^p$
log	$\log\lvert s\rvert$
BU	$-s^{-1}$
MOG	$-\log\sum_j \frac{\pi_j}{\sqrt{2\pi\sigma_j}}\exp\left(-\frac{s^2}{2\sigma_j^2}\right)$

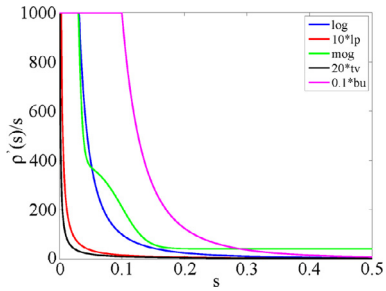
(a)



(b)

Prior	$\rho'(s)/s$
TV	s^{-1}
ℓ_p	$\lvert s\rvert^{p-2}$
log	$\lvert s\rvert^{-2}$
BU	s^{-3}
MOG	$\sum_j \frac{\frac{\pi_j}{\sigma_j}\exp\left(-\frac{s^2}{2\sigma_j^2}\right)}{\sigma_j^2\sum_j \frac{\pi_j}{\sigma_j}\exp\left(-\frac{s^2}{2\sigma_j^2}\right)}$

(c)



(d)

Fig. 2. (a) and (b): examples of penalty functions $\rho(s)$, where the MOG is obtained from Levin *et al.* [4]. (c): their corresponding $\rho'(s)/s$. (d): plots of $\rho'(s)/s$, where an upper bounding is taken for visualization. Note that TV is replaced with anisotropic TV (ℓ_1 prior) since isotropic TV cannot be shown in 1-D function.

and Fattal [32] introduce a power spectrum prior on the blur kernel. Recently, a novel convex blur regularizer based on the spectral properties of the convolution operators can be found in [33]. Since the spectral properties used are based on a linear and shift-invariant model without considering the noise, these methods do not work well for spatially varying blurs and noisy observations.

Let us consider the observation model in Eq. (1) and assume that the original image $\mathbf{x}$ is known. In this case we have N observations and aim at estimating K coefficients, where K is the size of the blur. Since the image size is usually much larger than the blur size, N observations should be sufficient to obtain a good blur estimate, even more so if L filtered observations are used. Based on the fact that usually $K << N$, many authors [15,23,34,35] have recently advocated the use of flat priors on the blur, enforcing only its nonnegativity and normalization constraints.

2.4. Hyperparameters models

So far we have studied the distributions $\mathrm{p}(\mathbf{z}|\Omega)$, $\mathrm{p}(\mathbf{y}|\mathbf{z},\Omega)$ that appear in the Bayesian modeling of the BID problem in Eq. (3). We complete this modeling by studying now the distribution $\mathrm{p}(\Omega)$.

An important problem is the estimation of the vector of parameters Ω when they are unknown. To deal with this estimation problem, the hierarchical Bayesian paradigm introduces a second stage, where the hyperprior $\mathrm{p}(\Omega)$ is also formulated.

For parameters, ω, corresponding to inverses of variances, the gamma distribution is used. It is defined by:

$$\mathrm{p}(\omega) = \Gamma(\omega|a_\omega, b_\omega) = \frac{(b_\omega)^{a_\omega}}{\Gamma(a_\omega)}\omega^{a_\omega-1}\exp\left[-b_\omega\omega\right], \tag{9}$$

where $\omega > 0$ denotes a hyperparameter, $b_\omega > 0$ is the rate parameter, and $a_\omega > 0$ is the shape parameter. These parameters are assumed known. The gamma distribution has the following mean, variance, and mode:

$$\mathbb{E}(\omega) = \frac{a_\omega}{b_\omega}, \mathrm{Var}(\omega) = \frac{a_\omega}{b_\omega^2}, \mathrm{Mode}(\omega) = \frac{a_\omega - 1}{b_\omega}. \tag{10}$$

Note that the mode does not exist when $a_\omega \leq 1$ and that mean and mode do not coincide. The literature also reports the use of non-informative prior models, $\mathrm{p}(\Omega) \propto \textit{constant}$, which can be considered as the limits of the above described hyperpriors.

Finally, we would like to mention here that the SG and SMG formulations turn the parameter estimation into a difficult task, especially when several filtered images are considered, since their partition functions can not usually be calculated.

3. Bayesian inference

Once the observation and prior models have been described, in other words, once the elements of the joint probability model in (3) have been specified, the goal now becomes the drawing of inference of the unknown variables $\Theta = \{\mathbf{z}, \Omega\}$ given the observations.

In the Bayesian framework Θ is inferred calculating (or approximating) the posterior distribution $\mathrm{p}(\Theta|\mathbf{y})$, expressed using the Bayes' rule as

$$\mathrm{p}(\Theta|\mathbf{y}) = \frac{\mathrm{p}(\Theta,\mathbf{y})}{\mathrm{p}(\mathbf{y})} = \frac{\mathrm{p}(\mathbf{y}|\Theta)\mathrm{p}(\mathbf{z}|\Omega)\mathrm{p}(\Omega)}{\mathrm{p}(\mathbf{y})}. \tag{11}$$

Unfortunately, since the integral $\mathrm{p}(\mathbf{y}) = \int \mathrm{p}(\Theta,\mathbf{y})\mathrm{d}\Theta$ is not tractable, the above posterior cannot be analytically calculated. Different estimation methods have been proposed to address this problem in the BID context and we will now review them.

Probably the most widely used method in the literature is Maximum a Posteriori (MAP). Since $\mathrm{p}(\Theta|\mathbf{y}) \propto \mathrm{p}(\Theta,\mathbf{y})$ the maximum of the posterior distribution can be obtained by maximizing the joint distribution $\mathrm{p}(\Theta,\mathbf{y})$ with respect to Θ. However, as pointed out in the landmark papers by Levin *et al.* [4,34], MAP is not a suitable estimation procedure in BID problems, because the associated cost function favors flat images for many sparse priors and leads to a delta blur estimate. To avoid the delta blur solution, Perrone and Favaro [14] show that a delayed normalization [11] should be used while other authors [36–38] suggest using non-dimensional sparsity measures.

Another very popular inference method is $\mathrm{MAP}_{\mathbf{h}}$ [24,34,39,40]. Unlike MAP, this method integrates the joint distribution with respect to $\mathbf{x}$ before estimating $\mathbf{h}$ and Ω, that is, blur and parameters are estimated by maximizing the evidence [41]. The restored image is finally calculated by maximizing the joint distribution, using

Table 1
Comparison of inference methods.

	MAP	$\text{MAP}_{\mathbf{h}}$	VB	MCMC
Has full posterior	no	partial	yes	yes
Has point estimates	yes	yes	yes	yes
Has uncertainty info	no	partial	yes	yes
Allows hidden data	no	yes	yes	yes
Complexity	low	low	medium	high

the estimate values of $\mathbf{h}$ and Ω through the above integration on the image.

Variational Bayesian inference has been widely used in BID (see [4–8,13,20,23,24,40]). VB generalizes MAP and $\text{MAP}_{\mathbf{h}}$ (see [42] for a proof) providing approaches for estimation of the posterior distributions of $\mathbf{x}$, $\mathbf{h}$ and Ω.

Together with the well established use of VB inference in BID, Markov Chain Monte Carlo (MCMC) methods are also gaining popularity. MCMC is the most general method used to approximate a posterior distribution, see [43–45] for details. The model in Eq. (3) is used to generate thousands of samples of $\mathrm{p}(\mathbf{z}, \Omega|\mathbf{y})$, which are used to infer the posterior distribution. In theory, sampling methods can find the exact form of the posterior distribution, but in practice they are computationally intensive (especially for multidimensional signals such as images) and their convergence is hard to establish.

In computationally cost terms, VB is much more efficient than MCMC, and more expensive than MAP or $\text{MAP}_{\mathbf{h}}$. The features of each method are summarized in Table 1.

We now describe the application of VB and MCMC to BID.

3.1. Variational inference in the image space for Super Gaussian priors

Since SG distributions are flexible enough to represent most of the image models used in the BID literature, we restrict, without loss of generality, the VB description to this representation. Furthermore we will formulate the inference in the image space; a detailed account of the use of SMG for the filter representation can be found in [23].

As it has already been explained above, the posterior distribution cannot be calculated analytically using the Bayes' rule in Eq. (11). To approximate $\mathrm{p}(\mathbf{z}, \Omega|\mathbf{y})$, VB minimizes the following Kullback–Leibler divergence

$$\begin{aligned}\mathrm{KL}(\mathrm{q}(\Theta)\|\mathrm{p}(\Theta|\mathbf{y})) &= \int \mathrm{q}(\Theta)\log\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta|\mathbf{y})}\mathrm{d}\Theta \\ &= \int \frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta,\mathbf{y})}\mathrm{d}\Theta + const.\end{aligned} \tag{12}$$

The Kullback–Leibler divergence is always non negative and is zero if and only if $\mathrm{q}(\Theta) = \mathrm{p}(\Theta|\mathbf{y})$. Since the minimizer $\mathrm{q}(\Theta) = \mathrm{p}(\Theta|\mathbf{y})$ cannot be calculated, some assumptions on $\mathrm{q}(\Theta)$ have to be made. One possible assumption is that $\mathrm{q}(\Theta)$ has a specific parametric form, e.g., a Gaussian distribution. Another widely used assumption is that $\mathrm{q}(\Theta)$ factorizes into disjoint groups, i.e.,

$$\mathrm{q}(\Theta) = \mathrm{q}(\mathbf{x})\mathrm{q}(\mathbf{h})\mathrm{q}(\Omega). \tag{13}$$

This factorized form of variational inference is called mean field theory in physics [46].

Using Eq. (13), the KL divergence can be minimized with respect to each factor while holding the other factors fixed. The optimal solution for each factor is then [47]

$$\log \mathrm{q}(\theta) = \mathbb{E}\left[\ln \mathrm{p}(\Theta,\mathbf{y})\right]_{\mathrm{q}(\bar{\Theta})} + const, \tag{14}$$

where $\bar{\Theta} = \Theta \setminus \theta$ is the set of unknowns excluding θ and $\mathbb{E}\left[\ln \mathrm{p}(\Theta,\mathbf{y})\right]_{\mathrm{q}(\bar{\Theta})}$ denotes the expectation taken with respect to all the approximating factors $\bar{\Theta}$. This system of equations is solved by an alternating minimization procedure, where each distribution $\mathrm{q}(\theta)$ is iteratively updated using the latest distributions of all the other factors. Since the KL divergence (12) is convex with respect to $\mathrm{q}(\theta)$ [48], the convergence of this alternating minimization procedure is guaranteed.

The penalty function $\rho(\cdot)$ defined in (5) can be represented as (see [49])

$$\rho(s) = \inf_{\xi>0} \frac{1}{2}\xi s^2 - \rho^*\left(\frac{1}{2}\xi\right), \tag{15}$$

where $\rho^*(\xi/2)$ is the concave conjugate function

$$\rho^*\left(\frac{1}{2}\xi\right) = \inf_{s} \frac{1}{2}\xi s^2 - \rho(s). \tag{16}$$

Furthermore, the infimum in (15) is achieved at $\xi = \rho'(s)/s$, as shown in [23]. Directly applying VB inference using $\mathrm{p}(\mathbf{x},\mathbf{h},\mathbf{y})$ is unfeasible, since the expectation of the logarithm of the joint distribution with respect to $\mathrm{q}(\mathbf{x})$ is intractable.

Since $\rho(s)$ is the penalty associated to a SG distribution we can write

$$\begin{aligned}&\mathrm{p}(\mathbf{x}) \geq Z\prod_{\gamma=1}^{L}\prod_{i=1}^{N}\exp(-\alpha_\gamma(\frac{\xi_\gamma(i)}{2}x_\gamma^2(i) - \rho^*(\frac{1}{2}\xi_\gamma(i)))), \\ &\forall \xi_\gamma(i) > 0,\end{aligned} \tag{17}$$

where Z is a constant. This Gaussian like lower bound will allow the expectation of the joint distribution to be calculated analytically. We have

$$\begin{aligned}\mathrm{p}(\mathbf{x},\mathbf{h},\mathbf{y}) &\geq \mathrm{p}(\mathbf{y}|\mathbf{x},\mathbf{h})\mathrm{p}(\mathbf{h})Z\prod_{\gamma=1}^{L}\prod_{i=1}^{N}\exp(-\alpha_\gamma(\frac{\xi_\gamma(i)}{2}x_\gamma^2(i) \\ &\quad - \rho^*(\frac{1}{2}\xi_\gamma(i)))) \\ &= M(\mathbf{y},\mathbf{x},\mathbf{h},\boldsymbol{\xi}),\end{aligned} \tag{18}$$

where $\boldsymbol{\xi} = \{\xi_\gamma(i), \gamma = 1,\ldots,L, i = 1,\ldots,N\}$ with all component positive.

We then have

$$\begin{aligned}&\int\int \mathrm{q}(\mathbf{x})\mathrm{q}(\mathbf{h})\log\frac{\mathrm{q}(\mathbf{x})\mathrm{q}(\mathbf{h})}{\mathrm{p}(\mathbf{x},\mathbf{h},\mathbf{y})}d\mathbf{x}d\mathbf{h} \\ &\leq \int\int \mathrm{q}(\mathbf{x})\mathrm{q}(\mathbf{h})\log\frac{\mathrm{q}(\mathbf{x})\mathrm{q}(\mathbf{h})}{M(\mathbf{y},\mathbf{x},\mathbf{h},\boldsymbol{\xi})}d\mathbf{x}d\mathbf{h} \\ &= \int\int\int \mathrm{q}(\boldsymbol{\xi})\mathrm{q}(\mathbf{x})\mathrm{q}(\mathbf{h})\log\frac{\mathrm{q}(\boldsymbol{\xi})\mathrm{q}(\mathbf{x})\mathrm{q}(\mathbf{h})}{M(\mathbf{y},\mathbf{x},\mathbf{h},\boldsymbol{\xi})}d\mathbf{x}d\mathbf{h}d\boldsymbol{\xi},\end{aligned} \tag{19}$$

where $\mathrm{q}(\boldsymbol{\xi})$ is a degenerate distribution on $\boldsymbol{\xi}$.

We then minimize the above integral on $\mathrm{q}(\boldsymbol{\xi})$, $\mathrm{q}(\mathbf{x})$, and $\mathrm{q}(\mathbf{h})$ assuming that $\mathrm{q}(\mathbf{h})$ is degenerate. According to (14), we obtain

$$\hat{\mathbf{h}} = \arg\max_{\mathbf{h}} \mathbb{E}[\log(M(\mathbf{y},\mathbf{x},\mathbf{h},\hat{\boldsymbol{\xi}}))]_{\hat{\mathrm{q}}(\mathbf{x})}, \tag{20}$$

$$\hat{\mathrm{q}}(\mathbf{x}) \propto M(\mathbf{y},\mathbf{x},\hat{\mathbf{h}},\hat{\boldsymbol{\xi}}), \tag{21}$$

$$\hat{\boldsymbol{\xi}} = \arg\max_{\boldsymbol{\xi}>0} \mathbb{E}[\log(M(\mathbf{y},\mathbf{x},\hat{\mathbf{h}},\boldsymbol{\xi}))]_{\hat{\mathrm{q}}(\mathbf{x})}. \tag{22}$$

3.2. Estimation of blur, image, and variational parameters

For the latent image, we obtain from (21),

$$\log\hat{\mathrm{q}}(\mathbf{x}) = -\frac{\beta}{2}\|\hat{\mathbf{H}}\mathbf{x} - \mathbf{y}\|_2^2 - \frac{1}{2}\sum_{\gamma=1}^{L}\alpha_\gamma x_\gamma^T diag(\hat{\xi}_\gamma)x_\gamma, \tag{23}$$

which is a multivariate Gaussian with precision matrix

$$\mathbf{C}_{\mathbf{x}}^{-1} = \beta \hat{\mathbf{H}}^T \hat{\mathbf{H}} + \sum_{\gamma=1}^{L} \alpha_\gamma \mathbf{F}_\gamma^T diag(\hat{\xi}_\gamma) \mathbf{F}_\gamma, \tag{24}$$

where $\mathbf{F}_\gamma$ is an $N \times N$ convolution matrix formed by the filter f_γ, and $\hat{\mathbf{H}}$ is an $N \times N$ convolution matrix obtained from $\hat{\mathbf{h}}$. The mean value $\hat{\mathbf{x}}$ is used as the estimate for $\mathbf{x}$, which is obtained by solving the following system of linear equations

$$\mathbf{C}_{\mathbf{x}}^{-1} \hat{\mathbf{x}} = \beta \hat{\mathbf{H}}^T \mathbf{y}. \tag{25}$$

For the variational parameter $\hat{\boldsymbol{\xi}}$, we obtain from (22)

$$\hat{\xi}_\gamma(i) = \frac{\rho'(v_\gamma(i))}{v_\gamma(i)}, \tag{26}$$

where $v_\gamma(i) = \sqrt{\mathbb{E}[x_\gamma^2(i)]}$, $1 \le i \le N$, with the expected value calculated using the distribution $\hat{\mathrm{q}}(\mathbf{x})$.

To estimate the blur we have

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}} \|\mathbf{H}\hat{\mathbf{x}} - \mathbf{y}\|_2^2 + \mathbf{h}^T \mathbf{D}_{\mathbf{x}} \mathbf{h}, \tag{27}$$

$$\text{subject to } \mathbf{h}(i) \ge 0, \sum_{i=1}^{K} \mathbf{h}(i) = 1, \tag{28}$$

where $\mathbf{D}_{\mathbf{x}}$ is a $K \times K$ matrix given by

$$\mathbf{D}_{\mathbf{x}}(m,n) = \sum_{j=1}^{N} \mathbf{C}_{\mathbf{x}}(m+j, n+j). \tag{29}$$

To estimate the variational parameters in Eq. (26) and the blur in Eq. (27), the matrix $\mathbf{C}_{\mathbf{x}}$ is required. This means that the $N \times N$ matrix $\mathbf{C}_{\mathbf{x}}^{-1}$ has to be inverted, a very time and memory demanding task. Following [4] and [23], we approximate $\mathbf{C}_{\mathbf{x}}$ as the inverse of the diagonal of $\mathbf{C}_{\mathbf{x}}^{-1}$. This inverse approximation is commented on in the open issues Section 4.3.

The prior image model can also be made dependent on a global parameter, which models the general scale behavior of the prior model. Its estimation is a very hard problem which can be approached under some assumptions on the prior model, see [20].

3.3. Algorithm

The VB based blind deconvolution algorithm is presented in Algorithm 1. However, as pointed out by Fergus *et al.* [8], directly applying it to estimate the blur may end up in the local minima, especially when the kernel support is large. To handle the large blur support problem, they suggest using a multiscale approach, namely building an image pyramid and then applying the BID method at each scale, which has proved to be very effective in BID problems. The rationale is that at the coarsest level, the blur is reduced significantly, so that it is easy to estimate a kernel from the downsampled image. At the next finer level, this kernel estimate is upsampled and can be used as a good initial guess for the single scale BID. Repeating this process until the finest level, we can obtain a better kernel estimate. After kernel estimation, we reconstruct the final sharp image using a non-blind deconvolution method (e.g., [18,50,51]).

The computation in Algorithm 1 is dominated by the solution of Eqs. (25) and (27). Since the most time consuming part when solving these two equations is the 2-D convolution, the computational complexity is $O(NK)$ or $O(N\log(N))$, depending on the usage of spatial convolution or FFT, respectively. We should mention that the computational complexity increases to an extremely large number, $O(N^3)$, if $\mathbf{C}_{\mathbf{x}}^{-1}$ is inverted exactly. The number of iterations required for convergence depends on the image priors. For example, the use of log prior leads to faster convergence than the use of $\ell_{0.8}$ prior, because the log prior is more edge preserving and sparsity promoting than the $\ell_{0.8}$ prior. It is also shown in [38] that, for the same optimization method and parameter settings, the use of the normalized ℓ_1 prior [38] results in fewer iterations for convergence than the ℓ_1/ℓ_2 prior [36] in the kernel estimation step. Due to the use of the covariance matrix, the VB BID method is slower than the MAP method [36–38], and much slower than the edge prediction based methods [52,53].

Algorithm 1 Single scale Bayesian blind deconvolution using Super Gaussian priors.

Require: Observation $\mathbf{y}$, noise level β, penalty $\rho(s)$, prior weight α.
1: Initialization $\mathbf{x} = \mathbf{y}$, $\mathbf{C}_{\mathbf{x}} = 0$,
2: **repeat**
3: Initialize $\boldsymbol{\xi}$
4: **while** not converge **do**
5: Update $\mathbf{x}$ by solving the linear system 25
6: Update $\boldsymbol{\xi}$ using Eq. (26)
7: Approximate $\mathbf{C}_{\mathbf{x}}(i,i)$ with $1/\mathbf{C}_{\mathbf{x}}^{-1}(i,i)$
8: **end while**
9: Update $\mathbf{h}$ by solving the quadratic programming problem in Eq. (27)
10: **until** Convergence

3.4. Sampling methods

Since the posterior distribution is not analytically available, sampling methods can be used to draw a large number of samples from it, and Monte Carlo integration techniques provide tools which allow performing inference on this dataset.

To simulate the posterior distribution Markov chains are used to develop different sampling methods. Perhaps the most widely used sampling method is Gibbs sampling described by the Geman and Geman in [54]. More recent methods are the Metropolis adjusted Langevin algorithms [55] and Hamiltonian Monte Carlo [56].

To better understand the sampling methods let us see an example of Gibbs sampling. If we can write down analytic expressions for the conditional distributions of all the unknowns we wish to estimate, given the others, we simply draw samples from each of the distributions in turn, conditioned on the most recently generated samples values for the other parameters. In our case we want to simulate $\mathrm{p}(\mathbf{x}, \mathbf{h}, \Omega|\mathbf{y})$; the iterations would proceed as follows:

$$\begin{aligned}
\text{First iteration:}\quad & \mathbf{x}^{(1)} \leftarrow \mathrm{p}(\mathbf{x}|\mathbf{h}^{(0)}, \Omega^{(0)}, \mathbf{y}) \\
& \mathbf{h}^{(1)} \leftarrow \mathrm{p}(\mathbf{h}|\mathbf{x}^{(1)}, \Omega^{(0)}, \mathbf{y}) \\
& \Omega^{(1)} \leftarrow \mathrm{p}(\Omega|\mathbf{x}^{(1)}, \mathbf{h}^{(1)}, \mathbf{y}) \\
\text{Second iteration:}\quad & \mathbf{x}^{(2)} \leftarrow \mathrm{p}(\mathbf{x}|\mathbf{h}^{(1)}, \Omega^{(1)}, \mathbf{y}) \\
& \mathbf{h}^{(2)} \leftarrow \mathrm{p}(\mathbf{h}|\mathbf{x}^{(2)}, \Omega^{(1)}, \mathbf{y}) \\
& \Omega^{(2)} \leftarrow \mathrm{p}(\Omega|\mathbf{x}^{(2)}, \mathbf{h}^{(2)}, \mathbf{y}) \\
& \vdots
\end{aligned} \tag{30}$$

where the symbol $\leftarrow$ means that the values are drawn form the distribution on the right. Once enough samples have been collected, point estimates and other statistics of the distribution may be found using Monte Carlo integration, for example the Minimum Squared Error estimator of the mean can be obtained as $\hat{\mathbf{x}} = \frac{1}{J}\sum_{j=1}^{J} \mathbf{x}^{(j)}$, where J is the number of drawn samples.

Due to the expensive computational cost (which is even worse when it is applied to high-dimensional data, such as images), the use of sampling methods in BID is not very extended. The works in this field are focused on developing more efficient algorithms. Ge *et al.* [57] or Kail *et al.* [58] propose modified versions of the Gibbs

sampling, and Pereyra [35] uses the Langevin algorithm which uses convex analysis to simulate efficiently the distributions.

4. Open issues

Before presenting some BID examples we would like to comment here on some open problems, either on the Variational Bayesian BID (VBBID) or BID itself, that we believe will very likely be explored in the near future:

4.1. Image space versus filter space

VB methods can be formulated in either the image or the filter space. Levin *et al.* [4] state that the filter space has better performance for the MOG prior. Xu *et al.* [37] indicate that using the image space formulation for latent image estimation and filter space formulation for kernel estimation is better than using the same spaces. In our opinion additional work is needed to establish the best spaces for image and kernel estimation. The image space appears to be less sensitive to noise since the noise is amplified in the filter space. Furthermore, the filter space is probably more computationally expensive than the image space. On one hand, utilizing the same number of iterations and L derivative filters, the total computation time in the filter space is about L times that of the image space. On the other hand, it is shown in Cho and Lee [52] that the kernel estimation in the image space requires more iterations to converge than in the filter space, since the symmetric matrix $\hat{\mathbf{X}}^T\hat{\mathbf{X}}$ of the quadratic program in Eq. (27), is not as diagonally dominant as $\sum_\gamma \hat{X}_\gamma^T\hat{X}_\gamma$, where $\hat{\mathbf{X}}$ and $\hat{X}_\gamma$ denote the convolution matrices formed by $\hat{\mathbf{x}}$ and $\hat{x}_\gamma$, respectively. Based on the above two factors, it is conceivable that the image space is computationally less expansive than the filter space, when $L \geq 2$. Additionally, filter space methods have access to more "observations" to estimate the blur, although with an unrealistic independence assumption on them. The pros and cons of both approaches should be carefully analyzed.

4.2. Bottom-up approach

The bottom-up approach, first proposed by Babacan *et al.* [23], refers to formulating a weight update scheme $\phi(v) = \rho'(v)/v$ for the Gaussian prior approximation (without knowing explicitly the penalty function) provided that $\phi(v)$ is decreasing on $(0, +\infty)$. A crucial and very challenging question is how to choose a good penalty function ρ or ϕ for VB blind deconvolution.

Wipf and Zhang [59] state that the preferred distribution is not the one reflecting the accurate statistics of the latent image, but the one that is most likely to guide VB iterations to high quality global solutions by strongly differentiating between blurry and sharp images. This implies choosing a ϕ that strongly discriminates sharp and blurry images. To that end, ϕ should be strongly sparsity promoting and also very edge preserving, such as $\phi(v) = v^{-p}$ with $(p \geq 2)$.

We believe that a trade-off between preserving edges and promoting sparsity should be achieved when dealing with noisy images. If noise is high, a very edge preserving $\phi(v)$ cannot suppress it. Babacan *et al.* [23] also suggest a more general form $\phi(v) = (Fv)^{-p}$, where F is a linear operator (e.g., nonlocal mean filter [60]). A variety of heuristics can easily be embedded through F to combat noise and increase robustness. Finally, we emphasize that given a ϕ, the value of α_γ should be chosen properly, as we will show in the experimental section.

4.3. Covariance approximation and general optimization issues

The covariance matrix $\mathbf{C_x}$ plays a very important role not only in the image estimation step but also in the kernel estimation step. This matrix makes VB methods different from MAP methods. Intuitively, the introduction of $\mathbf{C_x}$ in the estimation of the weights $\boldsymbol{\xi}$ makes their estimated values slightly smaller than when the covariance is not considered. As a result, the edges are better preserved. Besides, in the kernel estimation step, it provides an adaptive smoothness promoting regularization term which helps avoid the delta kernel estimates.

Unfortunately, due to the high computational cost, $\mathbf{C_x}$ is approximated by the inverse of the diagonal of the weighted deconvolution matrix $\mathbf{C_x}^{-1}$. Since $\mathbf{C_x}^{-1}$ is not diagonal, the diagonal approximation definitely introduces an error. The diagonal approximation is only reliable when β^{-1} and $\boldsymbol{\xi}$ are relatively large. If both β^{-1} and $\boldsymbol{\xi}$ are small, this approximation is not that reliable. Another alternative is the mean value approximation proposed by Babacan *et al.* [61], which replaces the weights $\boldsymbol{\xi}_\gamma$ with the average $\sum_{i=1}^N \xi_\gamma(i)/N$, so that $\mathbf{C_x}$ is a circulant matrix associated with the kernel $\mathbf{h}_{\mathbf{C_x}} = \mathcal{F}^{-1}\Lambda_{\mathbf{h}}^{-1}$, where $\mathcal{F}$ denotes the 2-D DFT and $\Lambda_{\mathbf{h}}$ is a column vector formed by the eigenvalues of $\mathbf{C_x}^{-1}$. $\mathbf{h}_{\mathbf{C_x}}$ has a large but finite support thanks to regularization and can be computed efficiently with the use of an FFT. This approximation takes the non-diagonal elements information into account, but the important information on the spatially variant weights is lost. The consequences of the use of the diagonal and mean value approximation remain an open question. Better but also feasible approximations to $\mathbf{C_x}$ should also be explored.

Notice that the image estimation step involves solving a nonconvex problem when the penalty function is nonconvex, e.g., $\rho(s) = s^p/p$ $(0 < p < 1)$. Assuming that the covariance term in $\boldsymbol{\xi}$ is ignored, it has been shown in [21] that the IRLS method which alternatively solves the linear equations in (25) and updates the weights by (26), definitely converges to a stationary point. Since the problem is nonconvex, the initial weights can make a difference in the final result, especially for the extremely nonconvex functions like log. A typical choice for the initial weights is the use of a large constant (e.g., 10^4, see [4,23]). It is conceivable that a $\boldsymbol{\xi}$ whose large values are located at the blur region may lead to a good stationary point, as the blur will be removed accurately. Since it is hard to know the blur region, finding such a good initial weights is not a trivial task in BID. Finally, we would like to mention that the linear equations (25) can be efficiently solved by ADMM [21] (Alternating Direction Method of Multipliers, see [62] for a comprehensive review), provided that the blur is spatially invariant.

Together with the IRLS method [21], other nonconvex optimization methods including variable splitting and look-up-table based method [63], ℓ_1-decomposition based method [64], and recently the smoothing trust region methods [65,66] have also been applied to image deconvolution.

4.4. Spatially varying blur and other modeling problems

In this paper we have assumed that the blur is the same across the image. However, as shown in [67], even the camera optical system generates a considerable amount of spatially varying (SV) blur. In general, spatially varying degradation can be modeled as

$$y(s) = \sum_u h(s, s-u)x(u) + n(s), \tag{31}$$

where $y(s)$ is the value of the observed image at position s, $x(u)$ is the value of the unknown ideal image at position u, $h(s, s-u)$ is the blur affecting the image, that depends on each image pixel position, and $n(s)$ is the noise. When SV BID is addressed, some restrictions are applied to the way the blur varies in Eq. (31) in order to make the problem feasible. Such restrictions include the assumption that the blur is piecewise-invariant or piecewise smooth

spatially varying, that is, the blur varies smoothly in the image, or that the blur is piecewise constant and location dependent, that is, different regions in the image have different blurs but the blur is spatially-invariant in each region. Another typical restriction is to assume that the type of the blur is known, for example, it is due to camera shake, or to consider images of a certain type, such as images with text [68] or star fields [69].

One approach to SV BID is to divide the image into non-overlapping patches where the blur is assumed to be stationary, apply a BID method on each patch independently, and merge the restored patches to obtain the final image. If the patches are not predefined, this approach casts the SV BID into a segmentation problem [70] in which the simpler case is to consider just two regions, a focused foreground and an out-of-focus background [71]. If the patches overlap and the blur varies smoothly on the image, the degradation model in Eq. (31) can be approximated as

$$y(s) = \sum_r \sum_u h_r(s-u) w_r(u) x_r(u) + n(s), \tag{32}$$

where $w_r(u) \geq 0$, $\sum_r w_r(u) = 1$, are weights allowing the smooth blending of the overlapping patches [72]. The advantage of this model is that it allows for an efficiently implementation using the Efficient Flow Filter (EFF) method [73] and also for different types of blur. On the other hand, its accuracy depends on the accuracy of the estimated kernel and may produce large errors if the kernels are not precisely estimated. In [74] the EFF method is extended to handle TV priors and a method to detect and replace erroneous blurs is proposed making it more robust. Another method to estimate smoothly varying blurs, with applications to star field images, is proposed in [69]. The method estimates the blur at certain image positions and uses SVD to remove outliers and estimate a smooth PSF field from the individual PSFs.

If only camera-shake blur is considered, the Projective Motion Path approach [75] models the SV degradation as the average of multiple sharp images, each one corresponding to one of all possible camera poses, that is,

$$y(s) = \sum_i x(H_i u) + n(s), \tag{33}$$

where H_i are homographies, that is, combinations of rotations and translations, that project the sharp image given a camera orientation. The homographies can be obtained from auxiliary sensors attached to the camera, such as gyroscopes (see [76] and Section 4.5), and high speed low resolution cameras [77], or they are estimated with the image [78].

A similar approximation is considered in [79] where the SV degradation is modeled as a weighted sum of sharp images obtained at all possible camera poses, that is,

$$\mathbf{y} = \sum_i w(i) \mathbf{C}_i \mathbf{x} + \mathbf{n}, \tag{34}$$

where $\mathbf{C}_i$ is the matrix that applies the homography H_i to the image $\mathbf{x}$ and $w(i)$ weights the i-th projection depending on the time spent by the camera at the i-th pose during the capture time. In [80], VB is used to estimate both the image $\mathbf{x}$ and the weights $w(\cdot)$. The drawback of this approach is that is resource demanding since it has to compute and store all the projections. To alleviate this problem, [81] proposes an iterative method that, at each iteration, restricts the solution space to a small set of camera poses which the camera motion trajectory is most likely to belong to.

Despite all these advances, more research is still needed to solve the general SV BID problem as described by Eq. (31).

Even without mentioning the spatially variant nature of the blur, the linear model in Eq. (1), utilized by most BID methods, is not a realistic one for real-life images. Common violations include the presence of defective sensor pixels, saturated pixels [82], a nonlinear camera response curve [83], or non additive white Gaussian noise [84,85] which, if not properly handled, may generate ringing artifacts when restoring the image even if the blur is accurately estimated [86]. We believe that developing methods that explicitly handle such model violations will improve the applicability of BID to real problems.

These modeling problems are alleviated with the use of more than one images. Considering color and, in general, multichannel images, remedies, to a great extend, the ill-posed nature of blind deconvolution [87]. Using image pairs with different properties facilitates blur estimation and helps handle saturated pixels and other camera imperfections. For instance, in [88] a near-infrared image is captured together with a visible blurred image and, in [89,90] a low exposure sharp but noisy image is used to improve the restoration results. If video is available, techniques can take into account the motion between frames [91–93] to tackle the deblurring problem. Of interest is also the approach in [94] where a single high-quality image is obtained from a sequence of images distorted by atmospheric turbulence. Having several images also allows blind image deconvolution to be addressed simultaneously with other problems, such as, super-resolution (see, for instance, [95] or [96]) or high dynamic rage (HDR) imaging [97].

Finally, to conclude this section on modeling, we would like to mention the need to model what a good restoration is. We believe that more BID software applications will be developed if the quality of a restored image can be assessed, without human intervention, before presenting it to the user.

4.5. Deconvolution in mobile devices

The ubiquity of mobile devices, such as smartphones and tablets, and the not-so-high quality of their cameras make the restoration of images taken with those devices a succulent market. Running the deconvolution process on mobile devices is, nevertheless, difficult given their limited computational power. Some commercial applications that claim to remove blur from images are available for the different platforms (see *DeblurIt Pro* or *Photo Fix de Blur* for Android or *Photo Doctor* for iOS). However they seem to deal only with out-of-focus blur with a manually selected radius and implement simple deconvolution algorithms.

Smartphones and tablets are more than simple cameras. They usually have other built-in sensors to capture the device position and trajectory and the capability of processing images. Hence, some methods are being proposed to perform deconvolution on those devices. In an effort to take advantage of the sensors present on the mobile phones, Šindelář and Šroubek [98] used the information provided by the gyroscope to keep track of the device motion while taking the picture and, hence, obtain an estimation of the blur by rendering the camera trajectory on the image plane. This blur estimate is used to deconvolve the image by a simple Wiener filter. An extension considering spatially variant blur and rolling shutter compensation is presented in [99].

A similar approach was used in [100] where the blur is obtained from a combination of the kernel estimated from the fusion of gyroscope, magnetometer and accelerometer measurements and a Gaussian kernel with small variance to take into account the out-of-focus blur due to the motion of the camera from the finger movement on pressing on the screen. Additionally, to minimize artifacts on faces, a face detection algorithm is applied to the image and an SVM classifier is trained and used to select between deblurring followed by denoising or sharpening the image, depending on the face characteristics.

Using a developer tablet modified by attaching a USB connected external gyroscope, a multi-image deconvolution which captures

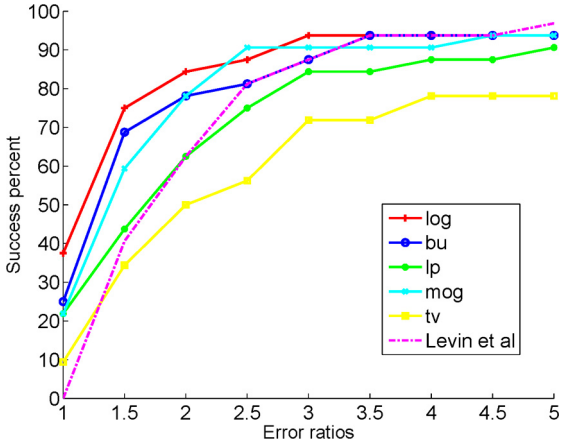


Fig. 3. Cumulative histograms of the error ratios across the dataset [39].

and combines multiple frames in order to make deblurring more robust and tractable is proposed in [101]. Blur is obtained from the gyroscope data and multi-image deconvolution is performed by minimizing

$$\sum_{i=1}^{n} \|\mathbf{y}_i - \mathbf{H}_i\mathbf{x}\|^2 + \lambda\|\Delta\mathbf{x}\|^p, \tag{35}$$

where λ is a regularization parameter and Δ is the gradient operator. The authors conclude that this deconvolution procedure outperforms, in most situations, the align-and-average strategy, that is, averaging multiple noisy images captured using a short exposure time, and hence blur-free, aligned using the gyroscope data. The optimization problem in Eq. (35) was first utilized in the work by Katsaggelos [102].

4.6. Implementation issues

Since BID methods need to estimate both image and blur, they usually take a significant amount of time. Apart from developing mathematically efficient methods to compute blur and image estimates, efficient implementations are needed to speed up the algorithms. Nowadays, most computers are equipped with graphical processing units (GPUs) that have several GFLOPS of computing power. Massive computing using these GPU or hybrid CPU + GPU computing can dramatically improve the speed of the algorithms. Most of the deconvolution implementations using GPUs are based on their capability to accelerate an FFT, with the CUDA framework and the CUFFT library[103] being the most popular implementation.

Some BID methods have been implemented using GPUs with great success as proved in [104] where the time needed to blindly deconvolve an 8 MPixel image using the method in [105] is reduced from 55.6 s to 13.8 s. The EFF spatially variant blind image deconvolution method in [73] runs about ten times faster using GPU than using only CPU.

Several efforts have also been carried out to use GPU computing in non-blind image deconvolution. For instance, Zhang *et al.* [106] performed real-time high definition 720p video processing with a Wiener filter using an NVIDIA GeForce GTX 460 GPU and Holder *et al.* [107] obtained an acceleration of 1:5 compared to CPU of the Richardson–Lucy algorithm on an NVIDIA GeForce GT640M. The GPU implemention of the non-blind Krishnan–Fergus [63] algorithm presented in [108] runs at 15 frames per second on 710×470 pixels color images on an NVIDIA GeForce GTX 260.

5. Experiments

We test the performance of 5 image priors, including log, $\ell_{0.8}$, MOG, TV and bu3, where the parameters for MOG are borrowed from Levin *et al.* [4] and bu3 is referred to the bottom-up approach [20,23] with $\phi(v) = v^{-3}$ (corresponding to $\rho(x) = -x^{-1}$). We choose the widely used dataset [39] which consists of 32 images generated by 4 groundtruth images with 8 motion blurs. For the priors log, $\ell_{0.8}$, MOG, TV and bu3, we set α_γ to 1, 10, 1, 20 and 0.1 respectively. After obtaining the kernels, we use the non-blind deconvolution method [18] with the same parameters used in [4] to reconstruct the final image.

Fig. 3 presents the success percent of 6 methods (ours with 5 different priors and Levin *et al.* [4]) in the sense of error ra-

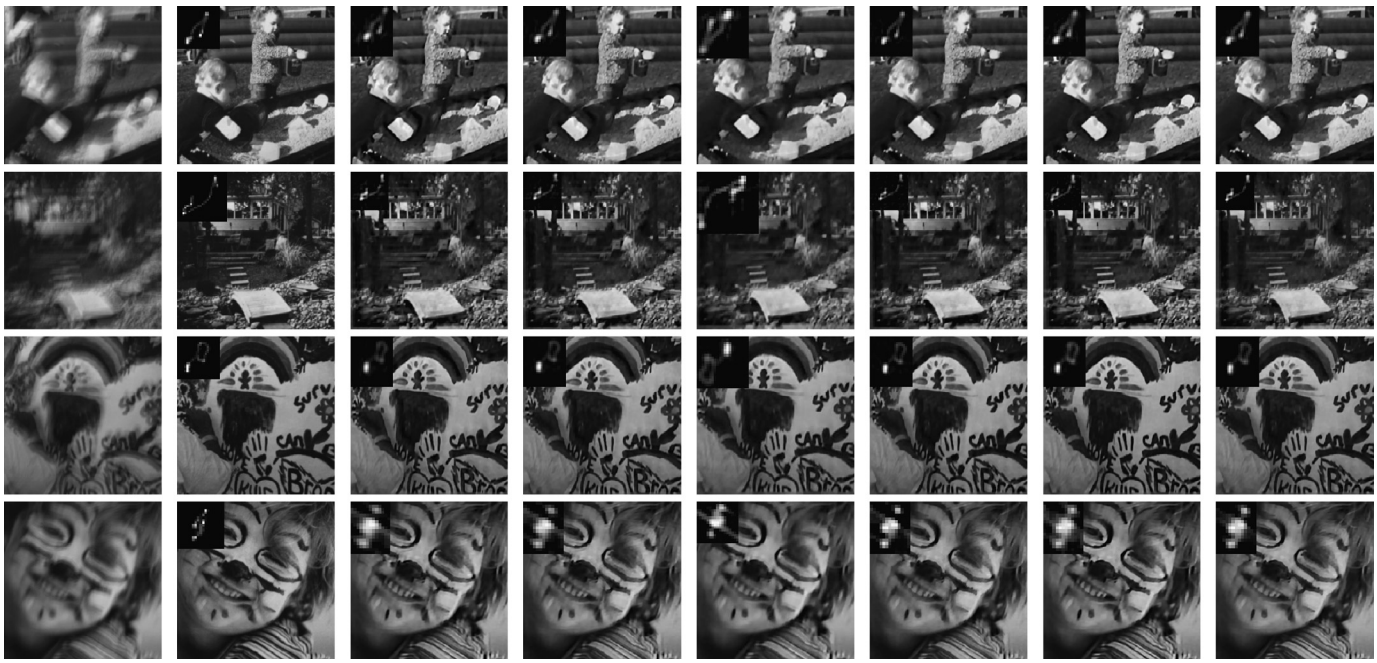


Fig. 4. Selected results on dataset [39] for visual comparison. (a) Blurred. (b) Groundtruth. (c) TV. (d) $\ell_{0.8}$. (e) Levin *et al.* [4]. (f) MOG. (g) bu3. (h) log.

tio metric (ratio between sum of squared difference errors of the restoration with the estimated kernel and the restoration with the groundtruth kernel, see [39] for more details). As we can see, the log prior has the best performance, with over 80% good restorations (error ratio ≤ 2) and 90% successful restorations (error ratio ≤ 3 is regarded as successful restoration, according to Levin *et al.* [4]), followed by *bu* and MOG. $\ell_{0.8}$ and TV also have good performance with about 80% and 70% successful restorations. It should be emphasized that, a suitable α_γ is crucial for the different priors to work well. Fig. 4 shows some selected results for visual evaluation.

References

[1] S. Horaczek, How many photos are uploaded to the internet every minute? [Online]. Available: http://www.popphoto.com/news/2013/05/how-many-photos-are-uploaded-to-internet-every-minute, May 2013.

[2] J. Hadamard, Lectures on Cauchy's Problem in Linear Partial Differential Equations, Yale University Press, New Haven, CT, 1923.

[3] T.E. Bishop, S.D. Babacan, B. Amizic, A.K. Katsaggelos, T. Chan, R. Molina, Blind Image Deconvolution: Problem Formulation and Existing Approaches, CRC Press, 2007.

[4] A. Levin, Y. Weiss, F. Durand, W.T. Freeman, Efficient marginal likelihood optimization in blind deconvolution, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2011, pp. 2657–2664.

[5] J. Miskin, D. MacKay, Ensemble Learning for Blind Image Separation and Deconvolution, Springer, 2000.

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

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

[8] R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, W.T. Freeman, Removing camera shake from a single photograph, in: ACM Transactions on Graphics (TOG) – Proceedings of ACM SIGGRAPH 2006, vol. 25, 2006, pp. 787–794, no. 3.

[9] R. Molina, A. Katsaggelos, J. Abad, J. Mateos, A Bayesian approach to blind deconvolution based on Dirichlet distributions, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, vol. 4, 1997, pp. 2809–2812.

[10] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D 60 (1992) 259–268.

[11] T.F. Chan, C.-K. Wong, Total variation blind deconvolution, IEEE Trans. Image Process. 7 (3) (1998) 370–375.

[12] S. Babacan, R. Molina, A. Katsaggelos, Variational Bayesian blind deconvolution using a total variation prior, IEEE Trans. Image Process. 18 (1) (Jan. 2009) 12–26.

[13] B. Amizic, R. Molina, A.K. Katsaggelos, Sparse Bayesian blind image deconvolution with parameter estimation, EURASIP J. Image Video Process. 2012 (1) (Nov. 2012).

[14] D. Perrone, P. Favaro, Total variation blind deconvolution: the devil is in the details, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2014, pp. 2909–2916.

[15] D. Perrone, P. Favaro, A clearer picture of blind deconvolution [Online]. Available: http://arxiv.org/abs/1412.0251, Nov. 2014.

[16] G. Chantas, N.P. Galatsanos, A. Likas, M. Saunders, Variational Bayesian image restoration based on a product of t-distributions image prior, IEEE Trans. Image Process. 17 (10) (Oct. 2008) 1795–1805.

[17] G. Chantas, N. Galatsanos, R. Molina, A. Katsaggelos, Variational Bayesian image restoration with a product of spatially weighted total variation image priors, IEEE Trans. Image Process. 19 (2) (Feb. 2010) 351–362.

[18] A. Levin, R. Fergus, F. Durand, W.T. Freeman, Image and depth from a conventional camera with a coded aperture, in: ACM Transactions on Graphics (TOG) – Proceedings of ACM SIGGRAPH 2007, vol. 26, 2007, p. 70, no. 3.

[19] Q. Shan, J. Jia, A. Agarwala, High-quality motion deblurring from a single image, in: ACM Transactions on Graphics (TOG) – Proceedings of ACM SIGGRAPH 2008, vol. 27, 2008, p. 73, no. 3.

[20] M. Vega, R. Molina, A. Katsaggelos, Parameter estimation in Bayesian blind deconvolution with super Gaussian image priors, in: European Signal Processing Conference, EUSIPCO, 2014, pp. 1632–1636.

[21] X. Zhou, R. Molina, F. Zhou, A. Katsaggelos, Fast iteratively reweighted least squares for l_p regularized image deconvolution and reconstruction, in: IEEE International Conference on Image Processing, ICIP, Oct. 2014, pp. 1783–1787.

[22] J.A. Palmer, K. Kreutz-Delgado, S. Makeig, Strong sub- and super-Gaussianity, in: V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval, E. Vincent (Eds.), Latent Variable Analysis and Signal Separation, in: Lect. Notes Comput. Sci., vol. 6365, Springer, Berlin, Heidelberg, 2010, pp. 303–310.

[23] S.D. Babacan, R. Molina, M.N. Do, A.K. Katsaggelos, Bayesian blind deconvolution with general sparse image priors, in: European Conference on Computer Vision, ECCV, 2012, pp. 341–355.

[24] H. Zhang, D. Wifp, Non-uniform camera shake removal using a spatially-adaptive sparse penalty, in: Advances in Neural Information Processing Systems, 2013, pp. 1556–1564.

[25] A. Mohammad-Djafari, Bayesian blind deconvolution of images comparing JMAP, EM and BVA with a student-t a priori model, in: International Workshops on Electrical Computer Engineering Subfields, 2014, pp. 98–103.

[26] D.F. Andrews, C.L. Mallows, Scale mixtures of normal distributions, J. R. Stat. Soc. B 36 (1) (1974) 99–102.

[27] K.P. Murphy, Machine Learning: A Probabilistic Perspective, Adapt. Comput. Mach. Learn. Ser., MIT Press, Cambridge (Mass.), 2012.

[28] M. Rattray, O. Stegle, K. Sharp, J. Winn, Inference algorithms and learning theory for Bayesian sparse factor analysis, in: Journal of Physics: Conference Series, vol. 197, 2009.

[29] M. Lavielle, Bayesian deconvolution of Bernoulli–Gaussian processes, Signal Process. 33 (1) (Jul. 1993) 67–79.

[30] J. Cai, H. Ji, C. Liu, Z. Shen, Blind motion deblurring from a single image using sparse approximation, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2009, pp. 104–111.

[31] S. Oh, G. Kim, Robust estimation of motion blur kernel using a piecewise-linear model, IEEE Trans. Image Process. 23 (3) (Mar. 2014) 1394–1407.

[32] A. Goldstein, R. Fattal, Blur-kernel estimation from spectral irregularities, in: European Conference on Computer Vision, ECCV, 2012, pp. 622–635.

[33] G. Liu, S. Chang, Y. Ma, Blind image deblurring using spectral properties of convolution operators, IEEE Trans. Image Process. 23 (12) (Dec. 2014) 5047–5056.

[34] A. Levin, Y. Weiss, F. Durand, W. Freeman, Understanding blind deconvolution algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (Dec. 2011) 2354–2367.

[35] M. Pereyra, Proximal Markov chain Monte Carlo algorithms [Online]. Available: http://arxiv.org/abs/1306.0187, Jun. 2013.

[36] D. Krishnan, T. Tay, R. Fergus, Blind deconvolution using a normalized sparsity measure, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2011, pp. 233–240.

[37] L. Xu, S. Zheng, J. Jia, Unnatural L0 sparse representation for natural image deblurring, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2013, pp. 1107–1114.

[38] X. Zhou, F. Zhou, X. Bai, Blind deconvolution using a nondimensional Gaussianity measure, in: IEEE International Conference on Image Processing, ICIP, 2013, pp. 877–881.

[39] A. Levin, Y. Weiss, F. Durand, W.T. Freeman, Understanding and evaluating blind deconvolution algorithms, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2009, pp. 1964–1971.

[40] C. Wang, Y. Yue, F. Dong, Y. Tao, X. Ma, G. Clapworthy, X. Ye, Enhancing Bayesian estimators for removing camera shake, Comput. Graph. Forum 32 (6) (Sep. 2013) 113–125.

[41] W.J. Fitzgerald, The Bayesian approach to signal modelling, in: IEE Colloquium on Non-Linear Signal and Image Processing (Ref. No. 1998/284), 1998, pp. 9/1–9/5.

[42] Z. Chen, S.D. Babacan, R. Molina, A.K. Katsaggelos, Variational Bayesian methods for multimedia problems, IEEE Trans. Multimed. 16 (4) (2014) 1000–1017.

[43] J.J.K.O. Ruanaidh, W.J. Fitzgerald, Numerical Bayesian Methods Applied to Signal Processing, Springer Verlag, 1996.

[44] W.J. Fitzgerald, Markov chain Monte Carlo methods with applications to signal processing, Signal Process. 81 (1) (2001) 3–18.

[45] S. Gulam-Razul, W.J. Fitzgerald, C. Andrieu, Bayesian deconvolution in nuclear spectroscopy using RJMCMC, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, vol. 2, 2002, pp. 1309–1312.

[46] G. Parisi, Statistical Field Theory, Addison-Wesley, 1988.

[47] C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[48] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[49] R.T. Rockafellar, Convex Analysis, Princeton University Press, 1996.

[50] 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.

[51] D. Zoran, Y. Weiss, From learning models of natural image patches to whole image restoration, in: IEEE International Conference on Computer Vision, ICCV, 2011, pp. 479–486.

[52] S. Cho, S. Lee, Fast motion deblurring, in: ACM Transactions on Graphics (TOG) – Proceedings of ACM SIGGRAPH Asia 2009, vol. 28, 2009, p. 145, no. 5.

[53] L. Xu, J.Y. Jia, Two-phase kernel estimation for robust motion deblurring, in: European Conference on Computer Vision, ECCV, 2010, pp. 157–170.

[54] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-6 (Nov. 1984) 721–741, no. 6.

[55] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, Springer-Verlag, 2004.

[56] R.M. Neal, MCMC using Hamiltonian dynamics [Online]. Available: http://arxiv.org/abs/1206.1901, Jun. 2012.

[57] D. Ge, J. Idier, E. Le Carpentier, Enhanced sampling schemes for MCMC based blind Bernoulli–Gaussian deconvolution, Signal Process. 91 (4) (2011) 759–772.
[58] G. Kail, J.Y. Tourneret, F. Hlawatsch, N. Dobigeon, Blind deconvolution of sparse pulse sequences under a minimum distance constraint: a partially collapsed Gibbs sampler method, IEEE Trans. Signal Process. 60 (6) (Jun. 2012) 2727–2743.
[59] D. Wipf, H. Zhang, Revisiting Bayesian blind deconvolution [Online]. Available: http://arxiv.org/abs/1305.2362, 2013.
[60] A. Buades, B. Coll, J. Morel, A non-local algorithm for image denoising, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, vol. 2, 2005, pp. 60–65.
[61] S.D. Babacan, R. Molina, A.K. Katsaggelos, Parameter estimation in TV image restoration using variational distribution approximation, IEEE Trans. Image Process. 17 (3) (2008) 326–339.
[62] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (2011) 1–122.
[63] D. Krishnan, R. Fergus, Fast image deconvolution using hyper-Laplacian priors, in: Advances in Neural Information Processing Systems, NIPS, 2009, pp. 1033–1041.
[64] M. Nikolova, M. Ng, C.P. Tam, Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction, IEEE Trans. Image Process. 19 (12) (2010) 3073–3088.
[65] M. Hintermüller, T. Wu, Nonconvex TV^q-models in image restoration: analysis and a trust-region regularization based superlinearly convergent solver, SIAM J. Imaging Sci. 6 (3) (2013) 1385–1415.
[66] X. Chen, L. Niu, Y. Yuan, Optimality conditions and a smoothing trust region Newton method for nonLipschitz optimization, SIAM J. Optim. 23 (3) (2013) 1528–1552.
[67] E. Kee, S. Paris, S. Chen, J. Wang, Modeling and removing spatially-varying optical blur, in: IEEE International Conference on Computational Photography, ICCP, 2011, pp. 1–8.
[68] X. Cao, W. Ren, W. Zuo, X. Guo, H. Foroosh, Scene text deblurring using text-specific multiscale dictionaries, IEEE Trans. Image Process. 24 (4) (Apr. 2015) 1302–1314.
[69] D. Miraut, J. Ball, J. Portilla, Efficient shift-variant image restoration using deformable filtering (Part II): PSF field estimation, EURASIP J. Adv. Signal Process. 2012 (1) (Aug. 2012) 1–19.
[70] F. Couzinie-Devy, J. Sun, K. Alahari, J. Ponce, Learning to estimate and remove non-uniform image blur, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2013, pp. 1075–1082.
[71] S. Chan, T. Nguyen, Single image spatially variant out-of-focus blur removal, in: IEEE International Conference on Image Processing, ICIP, 2011, pp. 677–680.
[72] S. Harmeling, M. Hirsch, B. Schölkopf, Space-variant single-image blind deconvolution for removing camera shake, in: Advances in Neural Information Processing Systems, vol. 23, 2010, pp. 829–837.
[73] M. Hirsch, C. Schuler, S. Harmeling, B. Schölkopf, Fast removal of non-uniform camera shake, in: Proceedings of the IEEE International Conference on Computer Vision, 2011, pp. 463–470.
[74] X. Yu, F. Xu, S. Zhang, L. Zhang, Efficient patch-wise non-uniform deblurring for a single image, IEEE Trans. Multimed. 16 (6) (2014) 1510–1524.
[75] Y.-W. Tai, P. Tan, M. Brown, Richardson–Lucy deblurring for scenes under a projective motion path, IEEE Trans. Pattern Anal. Mach. Intell. 33 (8) (2011) 1603–1618.
[76] N. Joshi, S.B. Kang, C.L. Zitnick, R. Szeliski, Image deblurring using inertial measurement sensors, ACM Trans. Graph. 29 (4) (2010) 1.
[77] Y.-W. Tai, H. Du, M. Brown, S. Lin, Correction of spatially varying image and video motion blur using a hybrid camera, IEEE Trans. Pattern Anal. Mach. Intell. 32 (6) (2010) 1012–1028.
[78] X. Zhang, F. Sun, Blind nonuniform deblur under projection motion path, J. Electron. Imaging 22 (3) (2013) 033034.
[79] O. Whyte, J. Sivic, A. Zisserman, J. Ponce, Non-uniform deblurring for shaken images, Int. J. Comput. Vis. 98 (2) (2012) 168–186.
[80] O. Whyte, J. Sivic, A. Zisserman, J. Ponce, Non-uniform deblurring for shaken images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 491–498.
[81] Z. Hu, M. hsuan Yang, Fast non-uniform deblurring using constrained camera pose subspace, in: Proceedings of the British Machine Vision Conference, 2012, pp. 136.1–136.11.
[82] O. Whyte, J. Sivic, A. Zisserman, Deblurring shaken and partially saturated images, Int. J. Comput. Vis. 110 (2) (2014) 185–201.
[83] S. Kim, Y.-W. Tai, S.J. Kim, M. Brown, Y. Matsushita, Nonlinear camera response functions and image deblurring, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Jun. 2012, pp. 25–32.
[84] M.A.T. Figueiredo, J.M. Bioucas-Dias, Restoration of Poissonian images using alternating direction optimization, IEEE Trans. Image Process. 19 (12) (Dec. 2010) 3133–3145.
[85] P. Rodriguez, R. Rojas, B. Wohlberg, Mixed Gaussian-impulse noise image restoration via total variation, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2012, pp. 1077–1080.
[86] S. Cho, J. Wang, S. Lee, Handling outliers in non-blind image deconvolution, in: IEEE International Conference on Computer Vision, ICCV, Nov. 2011, pp. 495–502.
[87] F. Šroubek, P. Milanfar, Robust multichannel blind deconvolution via fast alternating minimization, IEEE Trans. Image Process. 21 (4) (Apr. 2012) 1687–1700.
[88] W. Li, J. Zhang, Q.-H. Dai, Robust blind motion deblurring using near-infrared flash image, J. Vis. Commun. Image Represent. 24 (8) (2013) 1394–1413.
[89] S.D. Babacan, J. Wang, R. Molina, A.K. Katsaggelos, Bayesian blind deconvolution from differently exposed image pairs, IEEE Trans. Image Process. 19 (11) (Nov. 2010) 2874–2888.
[90] M. Tallón, J. Mateos, S.D. Babacan, R. Molina, A.K. Katsaggelos, Space-variant blur deconvolution and denoising in the dual exposure problem, Inf. Fusion 14 (4) (2013) 396–409.
[91] J. Brailean, A. Katsaggelos, Simultaneous recursive displacement estimation and restoration of noisy-blurred image sequences, IEEE Trans. Image Process. 4 (9) (Sep. 1995) 1236–1251.
[92] X. Deng, Y. Shen, M. Song, D. Tao, J. Bu, C. Chen, Video-based non-uniform object motion blur estimation and deblurring, Neurocomputing 86 (2012) 170–178.
[93] Y. Xu, X. Hu, S. Peng, Blind motion deblurring using optical flow, Optik 126 (1) (Jan. 2015) 87–94.
[94] X. Zhu, P. Milanfar, Removing atmospheric turbulence via space-invariant deconvolution, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (Jan. 2013) 157–170.
[95] W.-Z. Shao, M. Elad, Simple, accurate, and robust nonparametric blind super-resolution [Online]. Available: http://arxiv.org/abs/1503.03187, Mar. 2015.
[96] H. Zhang, L. Carin, Multi-shot imaging: joint alignment, deblurring, and resolution-enhancement, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014, pp. 2925–2932.
[97] C. Vijay, C. Paramanand, A. Rajagopalan, R. Chellappa, Non-uniform deblurring in HDR image reconstruction, IEEE Trans. Image Process. 22 (10) (Oct. 2013) 3739–3750.
[98] O. Šindelář, F. Šroubek, Image deblurring in smartphone devices using built-in inertial measurement sensors, J. Electron. Imaging 22 (1) (2013) 011003.
[99] O. Šindelář, F. Šroubek, P. Milanfar, A smartphone application for removing handshake blur and compensating rolling shutter, in: IEEE International Conference on Image Processing, ICIP, 2014.
[100] W. Jiang, D. Zhang, H. Yu, Sensor-assisted image deblurring of consumer photos on smartphones, in: 2014 IEEE International Conference on Multimedia and Expo, ICME, Jul. 2014, pp. 1–6.
[101] S.H. Park, M. Levoy, Gyro-based multi-image deconvolution for removing handshake blur, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2014, pp. 3366–3373.
[102] A.K. Katsaggelos, A multiple input image restoration approach, J. Vis. Commun. Image Represent. 1 (1990) 93–103.
[103] NVIDIA Corporation, NVIDIA CUDA fast Fourier transform [Online]. Available: http://docs.nvidia.com/cuda/cufft/index.html, 2015.
[104] T. Mazanec, A. Hermanek, J. Kamenicky, Blind image deconvolution algorithm on NVIDIA CUDA platform, in: 2010 IEEE 13th International Symposium on Design and Diagnostics of Electronic Circuits and Systems, DDECS, April 2010, pp. 125–126.
[105] F. Šroubek, J. Flusser, Multichannel blind deconvolution of spatially misaligned images, IEEE Trans. Image Process. 14 (7) (Jul. 2005) 874–883.
[106] Y. Zhang, J. He, J. Yuan, A video deblurring optimization algorithm based on motion detection, in: The 3rd International Conference on Multimedia Technology, ICMT-13, 2013.
[107] S. Holder, G. Lin, Acceleration of image restoration algorithms for dynamic measurements in coordinate metrology by using OpenCV GPU framework, 2014.
[108] J. Klosowski, S. Krishnan, Real-time image deconvolution on the GPU, in: SPIE Conference: Parallel Processing for Imaging Applications, Jan. 2011.

Pablo Ruiz received the M.S. degree in mathematics and the Masters degree in multimedia technologies in 2008 and 2009, respectively. Currently, he is a Ph.D. student of the Visual Information Processing group, at the Department of Computer Science and Artificial Intelligence of the University of Granada, where he enjoys a pre-doctoral fellowship. He has worked during 6 months in the Image and Video Processing Laboratory (IVPL) which is headed by Prof. Katsaggelos at Northwestern University in Illinois, USA. His research interest focuses in using Bayesian modeling and inference to solve different inverse problems like Image Restoration, Blind Image Deconvolution, Lightfields Acquisition, Active Learning and Classification.

Xu Zhou received the B.S. degree in applied mathematics from the Central South University, Changsha, China, 2009. Since 2011, he was a Ph.D. student at the Image Processing Center, Beihang University, Beijing, China. During the period from Oct. 2013 to Oct. 2014, sponsored by Chinese Scholarship Council, he was a visiting student of the Visual Information Processing group, at the Department of Computer Science and Artificial Intelligence of the University of Granada, Granada, Spain. Currently, he is primarily working on image deblurring.

Javier Mateos was born in Granada, Spain, in 1968. He 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 was an Assistant Professor with the Department of Computer Science and Artificial Intelligence, University of Granada, from 1992 to 2001, and then he became a permanent Associate Professor. 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 was finalist for the IEEE International Conference on Image Processing Best Student Paper Award (2010) and recipient of the IEEE International Conference on Signal Processing Algorithms, Architectures, Arrangements, and Applications Best Paper Award (2013). He is conducting research on image and video processing, including image restoration, image, and video recovery, superresolution 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), Institute of Electrical and Electronics Engineers (IEEE), and International Association for Pattern Recognition (IAPR).

Rafael Molina was born in 1957. He received the Degree in mathematics (statistics) in 1979 and the Ph.D. degree in optimal design in linear models in 1983. He became a Professor of Computer Science and Artificial Intelligence with University of Granada, Spain in 2000. He was a Former Dean of the Computer Engineering School, University of Granada (1992–2002), and the Head of the Computer Science and Artificial Intelligence Department, University of Granada (2005–2007). His current research interests include using Bayesian modeling and inference in problems like image restoration (applications to astronomy and medicine), super resolution of images and video, blind deconvolution, passive millimeter wave image threat detection, computational photography, source recovery in medicine, compressive sensing, low rank matrix decomposition, active learning, and classification.

Dr. Molina serves the IEEE and other Professional Societies. He was an Associate Editor of the IEEE Transactions on Image Processing (2005–2007), an Associate Editor of Progress in Artificial Intelligence (2010), an Associate Editor of Digital Signal Processing (2011), and an Area Editor (2011). He is a recipient of the IEEE ICIP Paper Award in 2007, the ISPA Best Paper Award in 2009, and the EUSIPCO Best Student Paper Award in 2013. He co-authored a paper that received 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 AT&T chair. He was previously the holder of the Ameritech Chair of Information Technology (1997–2003). He is also the Director of the Motorola Center for Seamless Communications, 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 200 journal papers, 500 conference papers and 40 book chapters) and he is the holder of 21 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, Optimization, and Sparsity (Cambridge University Press, forthcoming).

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 he is a current Member of the Award Board of the IEEE Signal Processing Society. He is a Fellow of the IEEE (1998) and SPIE (2009) and 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).