Blind Deconvolution With Model Discrepancies

Jan Kotera, *Student Member, IEEE*, Václav Šmídl, *Member, IEEE*, and Filip Šroubek, *Member, IEEE*

***Abstract*— Blind deconvolution is a strongly ill-posed problem comprising of simultaneous blur and image estimation. Recent advances in prior modeling and/or inference methodology led to methods that started to perform reasonably well in real cases. However, as we show here, they tend to fail if the convolution model is violated even in a small part of the image. Methods based on variational Bayesian inference play a prominent role. In this paper, we use this inference in combination with the same prior for noise, image, and blur that belongs to the family of independent non-identical Gaussian distributions, known as the automatic relevance determination prior. We identify several important properties of this prior useful in blind deconvolution, namely, enforcing non-negativity of the blur kernel, favoring sharp images over blurred ones, and most importantly, handling non-Gaussian noise, which, as we demonstrate, is common in real scenarios. The presented method handles discrepancies in the convolution model, and thus extends applicability of blind deconvolution to real scenarios, such as photos blurred by camera motion and incorrect focus.**

***Index Terms*— Blind deconvolution, variational bayes, automatic relevance determination, gaussian scale mixture.**

I. Introduction

NUMEROUS measuring processes in real world are modeled by convolution. The linear operation of convolution is characterized by a convolution (blur) kernel, which is also called a point spread function (PSF), since the kernel is equivalent to an image the device would acquire after measuring an ideal point source (delta function). In devices with classical optical systems, such as digital cameras, optical microscopes or telescopes, image blur caused by camera lenses or camera motion is modeled by convolution. Media turbulence (*e.g.* atmosphere in the case of terrestrial telescopes) generates blurring that is also modeled by convolution. In atomic force microscopy or scanning tunneling microscopy, resulting images are convolved with a PSF, whose shape is related to the measuring tip shape. In medical imaging, *e.g.* magnetic resonance perfusion, pharmacokinetic models consist of convolution with an unknown arterial input function. These are just a few examples of acquisition processes with a convolution model. In many practical applications convolution kernels are unknown. Then the problem of estimating latent data from blurred observations without any knowledge of kernels is called blind deconvolution.

Due to widespread presence of convolution in images, blind deconvolution is an active field of research in image processing and computer vision. However, the convolution model may not hold over the whole image. Various optical aberrations alter images so that only the central part of images follows the convolution model. Physical phenomena such as occlusion, under and overexposure, violate the convolution model locally. It is therefore important to have a methodology that handles such discrepancies in the convolution model automatically.

Let us assume the standard image acquisition model, in which a noisy observed image g is a result of convolution of a latent image u and an unknown PSF h, plus corruption by noise ϵ,

$$g = h * u + \epsilon. \tag{1}$$

The goal of blind image deconvolution is to recover u solely from the given blurry image g.

We follow the stochastic approach and all the variables in consideration are 2D random fields characterized by corresponding probability distributions denoted as $p(h)$, $p(u)$, and $p(\epsilon)$. The Bayesian paradigm dictates that the inference of u and h from the observed image g is done by modeling the posterior probability distribution $p(u, h|g) \propto p(g|u, h)p(u)p(h)$. Estimating the pair $(\hat{u}, \hat{h})$ is then accomplished by maximizing the posterior $p(u, h|g)$, which is commonly referred to as maximum *a posteriori* (MAP) approach, sometimes denoted $\text{MAP}_{u,h}$ to emphasize the simultaneous estimation of image and blur. Levin *et al.* in [1] pointed out that even for image priors $p(u)$ that correctly capture natural-image statistics (sparse distribution of gradients), $\text{MAP}_{u,h}$ approach tends to fail by returning a trivial "no-blur" solution, i.e., the estimated sharp image is equal to the input blurred input g and the estimated blur is a delta function. However, $\text{MAP}_{u,h}$ avoids the "no-blur" solution if we artificially sparsify intermediate images by shock filtering, removing weak edges, overestimating noise levels, etc., as widely used in [2]–[8].

From the Bayesian perspective, a more appropriate approach to blur kernel estimation is by maximizing the posterior marginalized w.r.t. the latent image u, *i.e.* $p(h|g) = \int p(u, h|g)du$. This distribution can be expressed in closed form only for simple image priors (*e.g.* Gaussian) and suitable approximation is necessary in other cases. In the Variational Bayesian (VB) inference, we approximate the posterior $p(u, h|g)$ by a restricted parametrization in factorized form and optimize its Kullback-Leibler divergence to the correct solution. The optimization is tractable and the resulting approximation provides an estimate of the sought marginal distribution $p(h|g)$.

Manuscript received August 19, 2016; revised January 14, 2017 and February 21, 2017; accepted February 21, 2017. Date of publication March 1, 2017; date of current version April 1, 2017. This work was supported by the Czech Science Foundation under Grant GA13-29225S and Grant GA15-16928S. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Javier Mateos.

The authors are with the Institute of Information Theory and Automation, Czech Academy of Sciences, 182 08 Prague, Czech Republic (e-mail: kotera@utia.cas.cz).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2017.2676981

As soon as the blur h is estimated, the problem of recovering u becomes much easier. It can be usually determined by the very same model, only now we maximize the posterior $p(u|g,h)$, or outsourced to any of the multitude of available non-blind deconvolution methods.

It is important to realize, that the error between the observation and the model, $\epsilon = g - h * u$, may not always be of stochastic uncorrelated zero-mean Gaussian nature – the true noise. In real-world cases, the observation error comes from many sources, *e.g.* sensor saturation, dead pixels or blur space-variance (objects moving in the scene) to name a few. Vast majority of blind deconvolution methods do not take any extra measures to handle model violation and the fragile nature of blind blur estimation typically causes complete failure when more than just a few pixels do not fit the assumed model, which unfortunately happens all too often. A non-identical Gaussian distribution with automatically estimated precision, which is called the Automatic Relevance Determination model (ARD) [9], is simple enough to be computationally tractable in the VB inference and yet flexible enough to handle model discrepancies far beyond the limited scope of Gaussian noise.

In this work, we adopt the probabilistic model of Tzikas *et al.* [10], which is based solely on VB approximation of the posterior $p(u,h)$ and which uses the same ARD model for all the priors $p(u)$, $p(h)$, and importantly also for the noise distribution $p(\epsilon)$. Our main focus is to analyze properties of the VB-ARD model and to elaborate on details of its implementation in real world scenarios, which was not directly considered in the original work of Tzikas. Specifically, we propose several extensions: include global precision for the whole image in the noise distribution $p(\epsilon)$ to decouple the Gaussian and non-Gaussian part of noise, different approximation of the blur covariance matrix, pyramid scheme for the blur estimation, and handling convolution boundary conditions. We demonstrate that VB-ARD with proposed extensions is robust to outliers and in this respect outperforms by a wide margin state-of-the-art methods.

The rest of the paper is organized as follows. Sec. II overviews related work in blind deconvolution, ARD modeling and masking. Sec. III discusses the importance of modeling the data error by ARD. Sec. IV presents the VB algorithm with ARD priors. Experimental validation of robustness to model discrepancies is given in Sec. V and Sec. VI concludes this work.

II. Related Work

First blind deconvolution algorithms appeared in telecommunication and signal processing in early 80's [11]. For a long time, the general belief was that blind deconvolution was not just impossible, but that it was hopelessly impossible. Proposed algorithms usually worked only for special cases, such as astronomical images with uniform (black) background, and their performance depended on initial estimates of PSF's; see [12], [13]. Over the last decade, blind deconvolution experiences a renaissance. The key idea behind the new algorithms is to address the ill-posedness of blind deconvolution by characterizing the prior $p(u)$ using natural image statistics and by a better choice of estimators. A major performance leap was achieved in [14] and [15] by applying VB to approximate the posterior $p(u,h|g)$ by simpler distributions. Other authors [6], [16]–[19] stick to the alternating $\text{MAP}_{u,h}$ approach, yet their methods converge to a correct solution by using appropriate ad hoc steps. It was advocated in [1] that marginalizing the posterior with respect to the latent image u is a proper estimator of the PSF h. The marginalized probability $p(h|g)$ can be expressed in a closed form only for simple priors, otherwise approximation methods such as VB [20] or the Laplace approximation [21] must be used. More recently in [4], [22], and [23], even better results were achieved when the model of natural image statistics was abandoned and priors that force unnaturally sparse distributions were used instead. Such priors belong to the category of strong Super-Gaussian distributions [24]. A formal justification of unnaturally sparse distributions was given in [25] together with a unifying framework for the MAP and VB formulation. An overview of state-of-the-art VB blind deconvolution methods can be found in [26].

The ARD model was originally proposed for neural networks in [9]. Each input variable has its associated hyperparameter that controls magnitudes of weights on connections out of that input unit. The weights have then independent Gaussian prior distributions with variance given by the corresponding hyperparameter. This prior distribution is also known as the scale mixture of Gaussian [27]. It is slightly less general than strong Super-Gaussian distributions but its advantage is that it is nicely tractable in the VB inference [22]. Distributions of the ARD type have been successfully used in the context of deconvolution as image or PSF priors [28]–[31], and less frequently also as a noise model [10], [32], [33].

A method of masking was originally proposed for handling convolution boundary conditions [34], however it can be also used for model discrepancies in blind deconvolution as discussed in [35]. A binary mask defines regions where convolution model holds. It allows for a very fast implementation using Fourier transform, which renders the method applicable to large scale problems. However, the main drawback of this approach is that the areas with model discrepancies are not estimated automatically and must determined in advance by other means. Saturated pixels were considered in [36] using the EM algorithm but only for a non-blind scenario. Explicitly addressing the problem of outliers in blind deconvolution was proposed in [2] using $\text{MAP}_{u,h}$ and automatically masking out regions where the convolution model is violated. Handling severe Gaussian noise was *e.g.* proposed in [5] by mitigating noise via projections.

III. Automatic Relevance Determination

In the discrete domain, convolution is expressed as matrix-vector multiplication. Then according to (1) the data error ϵ_i of the i-th pixel is

$$\epsilon_i = g_i - H_i u = g_i - U_i h\,, \quad i = 1, \ldots, N\,, \tag{2}$$

where H and U are convolution matrices performing convolution with the blur and latent image, respectively, and h

and u are column vectors containing lexicographically ordered elements of the corresponding 2D random fields. N is the total number of pixels. Subscript i in vectors denotes the i-th element and in matrices the i-th row. If the subscript is omitted then we mean the whole vector (or matrix).

In the majority of blind deconvolution methods, the data error term is assumed to be i.i.d. zero-mean Gaussian with precision α, i.e.

$$p(\epsilon|\alpha) = \prod_i \mathcal{N}(\epsilon_i|0, \alpha^{-1}). \tag{3}$$

Such assumption leads to the common ℓ^2 data term $\frac{\alpha}{2}\sum_i (g_i - U_i h)^2$. However as we demonstrate below, if this Gaussian-error assumption is slightly violated (e.g. by pixel saturation, model locally doesn't hold, etc.), the ℓ^2 data term gives an incorrect solution. It is therefore desirable to model both the Gaussian and non-Gaussian part of the error, for which the Student's t-distribution is a good choice, as it is essentially a scaled mixture of Gaussians and also plays nicely with the VB framework. As demonstrated earlier for the autoregressive model in [32], we propose using a Gaussian distribution with pixel-dependent factors γ_i modulated by the overall noise precision α. The error model of ϵ is then defined as

$$p(\epsilon|\alpha, \gamma) = \prod_i \mathcal{N}\left(\epsilon_i|0, (\alpha\gamma_i)^{-1}\right), \tag{4}$$

to which we refer as the *ARD model with common precision*. To draw a parallel to the classical formulation, the data term in this case takes the form $\frac{\alpha}{2}\sum_i \gamma_i (g_i - U_i h)^2$. The power of this model lies in determining the precisions α and γ_i automatically. This is covered in the following section, where we formulate the VB inference. For the current discussion, it suffices to state that we need priors also on γ. Let $\mathcal{G}$ denote the standard Gamma distribution, defined as $\mathcal{G}(\xi|a, b) = (1/\Gamma(a))b^a \xi_i^{a-1} \exp(-b\xi)$. We define the γ prior as

$$p(\gamma|\nu) = \prod_i \mathcal{G}(\gamma_i|\nu, \nu). \tag{5}$$

Marginalizing $p(\epsilon|\alpha, \gamma)p(\gamma|\nu)$ over γ gives us the Student's t-distribution with zero mean, precision α and degrees of freedom 2ν. From the above model it follows that the mean of γ_i is equal to $a/b = \nu/\nu = 1$. If ν becomes large then $\mathcal{G}(\gamma_i|\nu, \nu)$ tends to the delta distribution at 1 and the error model will be just a Gaussian distribution. As ν decreases, tails decay more slowly and γ_i will be allowed to adjust and automatically suppress outliers violating the acquisition model.

The *conventional ARD* model used e.g. in [10] is

$$p^*(\epsilon|\gamma) = \prod_i \mathcal{N}\left(\epsilon_i|0, \gamma_i^{-1}\right),$$
$$p^*(\gamma_i) = \mathcal{G}(\gamma_i|a_\gamma, b_\gamma). \tag{6}$$

The marginal distribution of this prior over γ is a Student's t-distribution with $2a_\gamma$ degrees of freedom. It is possible to choose the number of degrees of freedom as a priori known – a common approach is to choose a_γ, b_γ as small as possible, yielding Student's t-prior with infinite variance. Estimation of the hyperparameters a_γ, b_γ via a numerical MAP method has been proposed in [10].

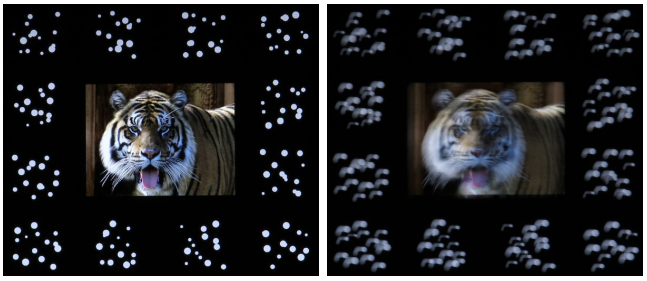

Fig. 1. Sharp (left) and intentionally blurred (right) image pair acquired for accurate calculation of the blur PSF from the known patterns surrounding the image.

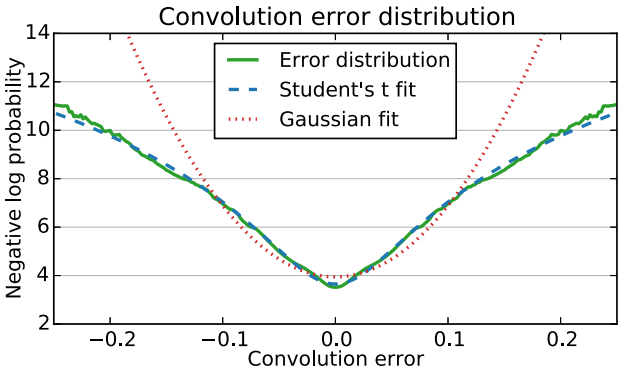

Fig. 2. Convolution error distribution in the case of real motion blur (solid green). It is much more heavy-tailed than the usually assumed Gaussian (dotted red, $\alpha = 0.6 \cdot 10^3$), while the Student's t-distribution (dashed blue, $2\nu = 3.5$, $\alpha = 1.2 \cdot 10^5$) is a perfect fit.

The ARD model is valuable in real scenarios even when there are no visible local discrepancies of the convolutional model. We conjecture that under real image acquisition conditions there exists no convolution kernel h such that the distribution of ϵ in (2) is strictly Gaussian. Different factors inherently present in the acquisition process, such as lens imperfections, camera sensor discretization and quantization, contribute to the violation of the convolution model. To verify our conjecture, we acquired several pairs of sharp–blurred images (u, g) with intentional slight camera motion during exposure. Except for this, we carefully avoided any other kinds of error like pixel saturation or space-invariance of the blur and worked strictly with raw data from the camera. For each of these pairs we estimated the blur PSF h following the procedure suggested in [37], which uses patterns printed around the image and designed to make the blur identification stable; see example in Fig. 1. For this data, we measured the error of the convolution (2) and plotted its distribution (negative log) in Fig. 2. The distribution is far from Gaussian, as the maximum likelihood estimate of the Gaussian distribution clearly provides a very poor approximation, especially in the tails. The Student's t-distribution, on the other hand, approximates the error distribution correctly and thus justifies the ARD choice for $p(\epsilon)$. It is interesting to note, that we performed a similar analysis on Levin's dataset [1] and obtained the same Student's t-distribution of the convolution error. Another justification of the ARD model provided in [10]

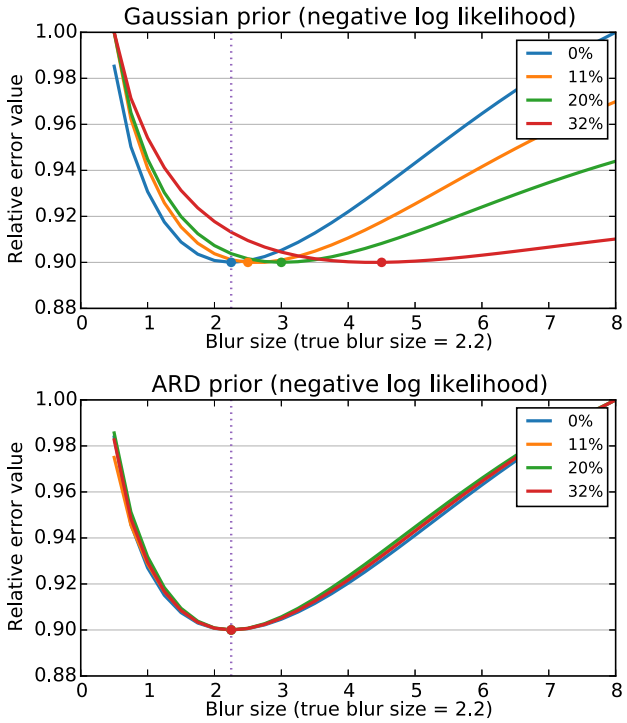

Fig. 3. Relative data term value (negative log likelihood) for commonly used Gaussian (top) and ARD (bottom) priors (low value means high probability of the particular PSF) for image blurred with PSF of size 2.2. Different data series correspond to different percentages of non-Gaussian error in the input. The curves' minima (indicated by dots) should correspond to the true PSF (vertical line). The Gaussian prior favors larger and larger blurs as the non-Gaussian error increases, while the ARD prior remains virtually unaffected and correctly identifies the true PSF.

is that a small error in the blur estimation also produces heavy-tailed $p(\epsilon)$.

Having demonstrated that the model error may have a significantly non-Gaussian distribution, the logical next step is to further analyze how this influences the solution accuracy. We conducted an experiment to answer a question: Is the ubiquitous standard-issue quadratic data term (Gaussian model) the right choice in the presence of non-Gaussian input error? As one expects, the answer is no. The problem is that in the presence of non-Gaussian error, the quadratic data term attains its minimum at a wrong point, therefore we get a solution, however not the solution we sought.

The setup of our experiment was as follows. We took a sharp image u and set a certain percentage of randomly selected pixels to over-exposed values to represent non-Gaussian error. We then blurred the saturated image with the "true" PSF h_{t}, added mild Gaussian noise and clipped the image intensities to obtain the final image g. We then proceeded to measure the goodness of several PSF candidates h by evaluating the corresponding data terms (more precisely, $-\log(\cdot)$ of the assumed noise distribution) of the classical Gaussian model (3) and the ARD model with common precision (4).

The whole experiment is graphically documented in two plots in Fig. 3. The top plot corresponds to the Gaussian model (quadratic data term) and the bottom plot corresponds to the ARD model used by our method. Individual line series represent different percentage of pixels intentionally corrupted by non-Gaussian error. The y-axis shows the data-term value for different blurs h as a function of their size (x-axis). The minimum of each line is marked by a small dot. The actual blur size of h_{t} is depicted by a vertical line around $x = 2.2$. For the Gaussian model (top plot), the minimum is reached for the correct blur size as long as the error is Gaussian-only (0% non-Gaussian). As the presence of non-Gaussian error increases, the data-term minimum shifts further and further away from the true point, effectively eliminating the chances of successful blur estimation. The ARD model (bottom plot), however, is unaffected by whatever amount of non-Gaussian error is thrown at it. We can conclude that when non-Gaussian input error can be expected the Gaussian presents a poor choice for the likelihood, a choice which compromises the chances of successful sharp image restoration.

IV. Variational Bayesian Inference

There are many examples of the VB inference applied to blind deconvolution in the literature; see *e.g.* [10], [14], [15], [20], [22]. They approximate the posterior $p(u, h|g)$ by a factorized distribution $q(u, h) = q(u)q(h)$. We follow the same path and use the ARD model with common precision for the error ϵ and the conventional ARD model for image and blur priors. The common precision in the image and blur priors is in our opinion superfluous, since it lacks any relation to real phenomena as opposed to the error ϵ where the common precision models white Gaussian noise.

Let us first define the individual distributions. Substituting from (2) into the ARD model in (4), the conditional probability distribution of the blurred image is

$$p(g|u, h, \alpha, \gamma) = \mathcal{N}(g|Hu, (\alpha\Gamma)^{-1}) = \prod_i \mathcal{N}(g_i|H_i u, (\alpha\gamma_i)^{-1}) \propto \prod_i (\alpha\gamma_i)^{1/2} \exp\left(-\frac{\alpha\gamma_i}{2}(g_i - H_i u)^2\right), \quad (7)$$

where Γ^{-1} is a diagonal covariance matrix having the inverse of the precision vector γ on the main diagonal, $\Gamma = \mathrm{diag}(\gamma)$. Let us recall that the precision γ_i is in general different for every pixel and it is determined from the data, which allows for automatic detection and rejection of outliers violating the acquisition model.

The image prior $p(u)$ is defined over image features (derivatives) and takes the form

$$p(u|\lambda) = \mathcal{N}(Du|0, \Lambda^{-1}) = \prod_i \mathcal{N}(D_i u|0, \lambda_i^{-1}) \propto \prod_i \lambda_i^{1/2} \exp\left(-\frac{\lambda_i}{2}(D_i u)^2\right), \quad (8)$$

where D_i is the first order difference at the i-th pixel and $\Lambda = \mathrm{diag}(\lambda)$. The operator D can be replaced by any sparsifying image transform, like wavelet transform or other set of high-pass filters. The prior ability to capture sparse features (edges) comes from the automatically determined precisions λ_i's.

It was advocated in [1], [19], and [22] to use flat priors on the blur and enforce only non-negativity, $h_i \geq 0$, and constant energy, $\sum_i |h_i| = 1$. This reasoning stems from the fact that the blur size is by several orders of magnitudes smaller

than the image size and therefore inferring the blur from the posterior is driven primarily by the likelihood function (7) and less by the prior $p(h)$. However, if the image estimation u is inaccurate, which is typically the case in the initial stages of any blind deconvolution algorithm, then a more informative prior $p(h)$ is likely to help in avoiding local maxima and/or speeding up the convergence. To keep the approach coherent, we apply the ARD model on blur intensities

$$p(h|\beta) = \mathcal{N}(h|0, B^{-1}) = \prod_i \mathcal{N}(h_i|0, \beta_i^{-1}) \propto \prod_i \beta_i^{1/2} \exp\left(-\frac{\beta_i}{2} h_i^2\right), \tag{9}$$

where $B = \mathrm{diag}(\beta)$.

The ARD models in (7), (8), and (9) are conditioned to unknown precision parameters $(\alpha, \gamma_i, \lambda_i, \beta_i)$. The conjugate distributions of precisions are Gamma distributions and thus for image and blur precisions we have

$$\begin{aligned} p(\lambda_i) &= \mathcal{G}(\lambda_i|a_\lambda, b_\lambda), \\ p(\beta_i) &= \mathcal{G}(\beta_i|a_\beta, b_\beta)\,, \end{aligned} \tag{10}$$

and for the error precisions according to (4) and (5) we have

$$\begin{aligned} p(\alpha) &= \mathcal{G}(\alpha|a_\alpha, b_\alpha), \\ p(\gamma_i|\nu) &= \mathcal{G}(\gamma_i|\nu, \nu)\,, \\ p(\nu) &= \mathcal{G}(\nu|a_\nu, b_\nu). \end{aligned} \tag{11}$$

The hyperparameters $a_{(\cdot)}$ and $b_{(\cdot)}$ are user-defined constants.

Let $\mathcal{Z} = \{u, h, \alpha, \nu, \{\gamma_i\}, \{\lambda_i\}, \{\beta_i\}\}$ denote all the unknown variables and $\mathcal{Z}_k$ its particular member indexed by k. Using the above defined distributions, the posterior $p(\mathcal{Z}|g)$ is proportional to

$$p(g|u, h, \alpha, \gamma)p(\alpha)p(\gamma|\nu)p(\nu)p(u|\lambda)p(\lambda)p(h|\beta)p(\beta)\,.$$

The VB inference [38] approximates the posterior $p(\mathcal{Z}|g)$ by the factorized distribution $q(\mathcal{Z})$,

$$p(\mathcal{Z}|g) \approx q(\mathcal{Z}) = q(u)q(h)q(\alpha)q(\nu)q(\gamma)q(\lambda)q(\beta). \tag{12}$$

This is done by minimizing the Kullback-Leibler divergence, which provides a solution for individual factors

$$\log q(\mathcal{Z}_k) \propto \mathbb{E}_{l\neq k}\left[\log p(\mathcal{Z}|g)\right], \tag{13}$$

where $\mathbb{E}_{l\neq k}$ denotes expectation with respect to all factors $q(\mathcal{Z}_l)$ except $q(\mathcal{Z}_k)$. Formula (13) gives implicit solution, because each factor $q(\mathcal{Z}_k)$ depends on moments of other factors. We must therefore resort to an iterative procedure and update the factors q in a loop.

A detailed derivation of update equations can be found in [10] as the model is similar to ours. The interested reader is also referred to [15] for better understanding of the derivation. In the following subsections we therefore only state the update equations yet analyze their properties in detail.

A. Likelihood

The important feature is automatic estimation of the non-Gaussian part of the error modeled by precision γ. Utilizing the combination of VB inference and ARD prior, we are able to detect and effectively reject outliers from the estimation and achieve unprecedented robustness of the blur estimation, much needed in practical applications.

Using (13), $q(\gamma)$ becomes a Gamma distribution with a mean value

$$\overline{\gamma}_i = \frac{1 + 2\,\overline{\nu}}{\overline{\alpha}\,\mathbb{E}_{u,h}\left[(g_i - H_i u)^2\right] + 2\,\overline{\nu}}\,, \tag{14}$$

where $\overline{(\cdot)}$ denotes a mean value. Relating the inference to the classical minimization of energy function $-\log p(u, h|g)$, the precision γ_i corresponds to the weight of the i-th pixel in data fidelity term. The above equation shows that this weight is inversely proportional to the (expected) reconstruction error at that pixel (up to the relaxation by $\overline{\nu}/\overline{\alpha}$) and it is updated during iterations, as the image and blur estimates change. This technique is similar to the method of iteratively reweighted least squares (IRLS), where the quadratic data terms are reweighted according to the error at the particular data point to achieve greater robustness to outliers, but here it arises naturally as part of the VB framework. We demonstrate how the method behaves with respect to outliers in the experimental section.

According to (14), the mean value of γ depends, apart from u and h, only on the mean values $\overline{\alpha}$ and $\overline{\nu}$. Using again the VB inference formula (13), one can deduce that both $q(\alpha)$ and $q(\nu)$ are Gamma distributions with mean values

$$\overline{\alpha} = \frac{N + 2a_\alpha}{\sum_{i=1}^N \overline{\gamma_i}\,\mathbb{E}_{u,h}\left[(g_i - H_i u)^2\right] + 2b_\alpha} \tag{15}$$

and

$$\overline{\nu} = \frac{N + 2a_\nu}{2\sum_{i=1}^N (\overline{\gamma_i} - \mathbb{E}_{\gamma_i}\left[\log(\gamma_i)\right] - 1) + 2b_\nu}\,. \tag{16}$$

The update equation of $\overline{\nu}$ requires Stirling's approximation; see e.g. [32] for detailed derivation. Note that the above update equations (14), (15) and (16) are easy to compute.

Precision α is expected to be inversely proportional to the level of Gaussian noise in the input image. It is therefore interesting to observe how α behaves during iterations. After the initialization, when the reconstruction error is high, the weight α is correspondingly low and thus the role of priors (regularization) is increased in the early stages of estimation. During subsequent iterations, as the estimation improves, α increases and the effect of priors is attenuated. It has long been observed that this adjustment of data-term weight during iterations is highly beneficial, if not necessary, for the success of blind blur estimation, otherwise the optimization tends to get stuck in a local minimum. Many state-of-the-art blind deconvolution methods therefore perform some kind of heuristic adjustment of the relative data-term/regularizer weight [7], often of the form of geometric progression $\alpha^{k+1} = r\alpha^k$, where k denotes k-th iteration. The drawback of this approach is that the optimal constant r must be determined by trial and error and, more importantly, the progression must

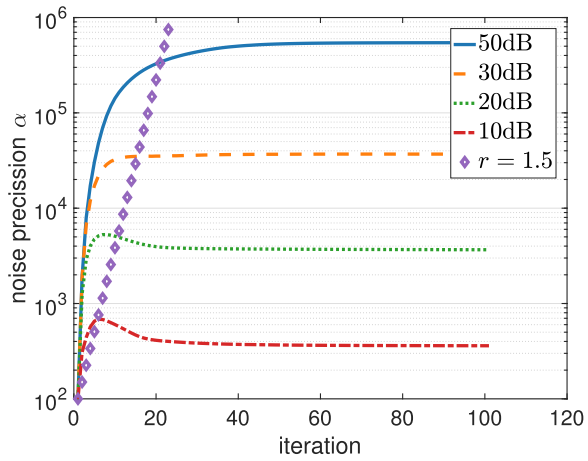

Fig. 4. Estimated noise precision as a function of iterations: The Variational Bayesian algorithm updates the noise precision in every iteration. The curves depict its typical development for different image SNRs; 50dB through 10dB. The *diamond markers* show the fixed update using geometric progression $\alpha^k = 1.5\alpha^{k-1}$.

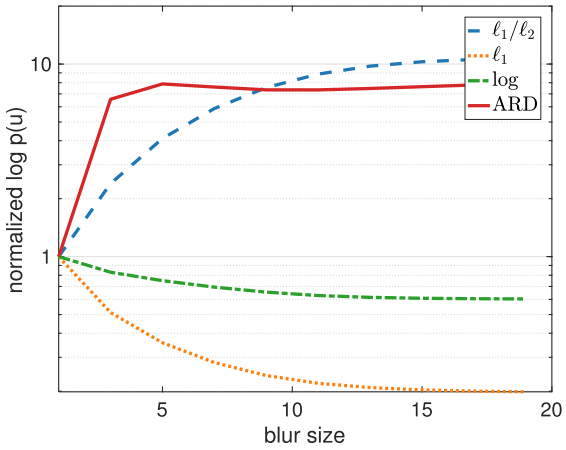

Fig. 5. Comparison of priors: The graph shows $-\log p(u)$ of priors as a function of amount of blurring. The ℓ^1 prior (*dotted red line*) decreases and so does the log prior (*dash-dotted yellow line*), which is the marginalized ARD prior. On the other hand, the ARD prior (*solid purple line*) with precisions estimated from the sharp image steeply increases and flattens out for large blurs. The normalized prior ℓ^1/ℓ^2 (*dashed blue line*) increases more slowly but steadily. The value of priors are normalized to give 1 on sharp images (1 *blur size*). The curves show mean values calculated on various images (photos of nature, human faces, buildings).

stop when the correct α (corresponding to the true noise level) is reached, which is not determined automatically but must be specified by the user. The VB framework has an indisputable advantage over more straightforward MAP methods – not only does it give us the optimal update equation for the data-term precision, it also provides automatic saturation when the correct noise level is reached, as we can see in Fig. 4. During the early iterations the precision sharply increases and then levels out at the correct value. For comparison we also show the fixed geometric progression for $r = 1.5$ (diamond markers).

B. Image Prior

The factors associated with the image are $q(u)$ and $q(\lambda)$. Applying (13), we get (up to a constant)

$$\log q(u) = -\mathbb{E}_{h,\alpha,\gamma,\lambda}\left[\alpha(g-Hu)^T\Gamma(g-Hu)+u^TD^T\Lambda Du\right], \quad (17)$$

where the terms independent of u are omitted. The distribution $q(u)$ is a normal distribution. The mean $\overline{u}$ and covariance $\mathrm{cov}(u)$ are obtained by taking the first and second order derivatives of $\log(q(u))$, respectively, and solving for zero. The update equation for the mean is a linear system

$$\left(\mathbb{E}_h\left[H^T\overline{\Gamma}H\right]+\overline{\alpha}^{-1}D^T\overline{\Lambda}D\right)\overline{u} = \overline{H}^T\overline{\Gamma}g \quad (18)$$

and for the covariance we get

$$\mathrm{cov}\, u = \left(\overline{\alpha}\mathbb{E}_h\left[H^T\overline{\Gamma}H\right]+D^T\overline{\Lambda}D\right)^{-1}. \quad (19)$$

The mean pixel precisions $\overline{\lambda}_i$ form the diagonal matrix $\overline{\Lambda}$. They are calculated from $q(\lambda)$, which is a Gamma distribution with the mean

$$\overline{\lambda}_i = \frac{1+2a_\lambda}{\mathbb{E}_u\left[(D_iu)^2\right]+2b_\lambda}. \quad (20)$$

The parameter b_λ plays the role of relaxation, as it prevents division by zero in the case $D_iu = 0$.

It was demonstrated in [1] that commonly used ℓ^p $(p < 1)$ priors with $|D_iu|^p$ in the exponent, counter-intuitively favor blurred images over the sharp ones and therefore cannot avoid the "no-blur" solution by itself. The sparsity of image derivatives decreases with increasing blur but the variance of image derivatives decreases as well. The second effect is stronger in natural images and therefore the total prior probability increases with blur. We want to analyze if the ARD image prior $p(u|\lambda)$ in (8) behaves better in this respect and how it compares with the unconditional (marginalized) version $p(u)$. Following the analysis of Gaussian scale mixtures in [27] and [29], the unconditional prior $p(u)$ is obtained by marginalizing over λ, which in the limit for $a_\lambda \to 0$ yields

$$p(u) = \int p(u,\lambda)\mathrm{d}\lambda = \int p(u|\lambda)p(\lambda)\mathrm{d}\lambda \propto \prod_i \exp\left(-\frac{1}{2}\log\left((D_iu)^2+b\right)\right). \quad (21)$$

The marginalized ARD image prior is of the exponential form with exponent $\log\left((D_iu)^2+\mathrm{const}\right)$. It is thus equivalent to the log prior proposed in [22]. It is a non-convex prior that aggressively favors sparsity of the natural image statistics. In this sense it resembles the ℓ^p priors with $p \to 0$.

We calculated $-\log p(u)$ on natural images (photos of nature, human faces, buildings) blurred with Gaussian blur of varying size. We also tried motion and uniform blur and the behavior was identical. For each image we have normalized $-\log p(u)$ calculated on differently blurred versions of the image so the original sharp image (no blur) gives 1. The normalized $-\log p(u)$ of different priors as a function of blur size and averaged over all images is plotted in Fig. 5. As expected, the exponent of the ℓ^1 prior decreases as the blur increases, i.e. this prior favors blurred images over sharp ones. The log prior, which is the marginalized ARD prior, decreases less but still favors blurred images. In the VB framework,

however, we do not work with the marginalized ARD prior and instead iteratively estimate the prior precisions λ_i's from the current estimate of u using the update formula (20). Let us assume an ideal situation in which the precisions are estimated from the sharp image, then the ARD prior shows correct behavior similarly to the normalized prior ℓ^1/ℓ^2 [39] that compensates for the effect of decreasing image variance. This ideal case is not achievable in practice, since we do not have a correct estimate of the sharp image u at the beginning, but it can be regarded as an upper bound. As the VB inference makes the approximation of the posterior more accurate with every iteration, we approach this upper bound.

C. Blur Prior

As stated earlier, we use the same ARD model also for the blur prior (9). Analogously to the derivation of the image distribution $q(u)$ in (17), the form of blur factor $q(h)$ is a Gaussian distribution given by

$$\log q(h) = -\mathbb{E}_{u,\alpha,\gamma,\beta}\left[\alpha(g-Uh)^T\Gamma(g-Uh)+h^TBh\right].$$

Then the mean $\overline{h}$ is the solution of the linear system

$$\left(\mathbb{E}_u\left[U^T\overline{\Gamma}U\right]+\overline{\alpha}^{-1}\overline{B}\right)\overline{h}=\overline{U}^T\overline{\Gamma}g \tag{22}$$

and the covariance is

$$\operatorname{cov} h = \left(\overline{\alpha}\mathbb{E}_u\left[U^T\overline{\Gamma}U\right]+\overline{B}\right)^{-1}. \tag{23}$$

The distribution $q(\beta)$ of the blur precision is again a Gamma distribution and for the mean values of β_i we get analogously to (20)

$$\overline{\beta}_i = \frac{1+2a_\beta}{\mathbb{E}_h\left[h_i^2\right]+2b_\beta}. \tag{24}$$

State-of-the-art blind deconvolution methods often estimate h while enforcing positivity and constant energy, i.e. $h_i \geq 0$ and $\sum_i h_i = 1$. Enforcing such constraints in our case means to solve the least squares objective associated with (22) under these constraints. Since the constraints form a convex set, we can use, *e.g.*, the alternating direction method of multipliers (ADMM) [40], that solves convex optimization problems by breaking them into smaller pieces, each of which is then easier to handle. However, applying such constraints would take us outside the VB framework, as $q(h)$ is then no longer a Gaussian distribution and $\operatorname{cov} h$ is intractable. To test the influence of the constraints, we have used the proximal algorithm to solve the constrained (22), albeit violating the VB framework, but we have noticed no improvement.

One explanation is that a non-negative solution is a local extreme of VB approximation which attracts the optimization when the initial estimate is also non-negative. Let us assume that the PSF is initialized with non-negative values, which is always true in practice as PSFs are typically initialized with delta functions. If during VB iterations, any h_i approaches zero then the corresponding precision calculated in (24) grows, reaching $1/(2b)$ if $a \to 0$. If the hyperparameters are sufficiently small (which is our case), this correspond to a very tight distribution $q(h_i)$ that traps h_i at zero and prevents further changes.

The covariance $\operatorname{cov} h$ has an additional positive influence on the behavior of the PSF precision β. The denominator of (24) expands to $\overline{h}_i^2 + \operatorname{cov} h_i + 2b$. From (23) it follows that $\operatorname{cov} h_i$ is inversely proportional to $\alpha + \beta_i$. We can ignore γ, since it is in average around 1 anyway as it captures only local non-Gaussian errors. We have seen in Fig. 4 that α starts small, which implies larger $\operatorname{cov} h_i$ and thus small PSF precision β_i. Small β_i loosely constrains the estimation of the PSF h during initial iterations. As α increases later on, $\operatorname{cov} h_i$ decreases and β_i increases, which helps to fix the estimated values of h.

D. Algorithm

All equations in the VB inference are relatively easy to solve, except for the calculation of covariance matrices $\operatorname{cov} u$ in (19) and $\operatorname{cov} h$ in (23), which involves inverting precision (concentration) matrices. Both matrices are large and their inversion is not tractable since they are a combination of convolution and diagonal matrices. The covariance is important in the evaluation of expectation terms $\mathbb{E}[\cdot]$. To tackle this problem, we approximate precision matrices by diagonal ones. This is different from Tzikas's work [10], where $\operatorname{cov} h$ is approximated by a convolution matrix. The experimental section demonstrates that the diagonal approximation performs better.

We show the approximation procedure on $\operatorname{cov} h$ and calculation of u. The approximation of $\operatorname{cov} u$ and calculation of h is similar. First we approximate the covariance matrix $\operatorname{cov} h$ by inverting only the main diagonal of the precision matrix, i.e., $(\operatorname{diag}(\overline{\alpha}\mathbb{E}_u\left[U^T\overline{\Gamma}U\right]+\overline{B}))^{-1}$. Here we use the syntax of popular numerical computing tools such as MATLAB, Python or R, and assume that the operator $\operatorname{diag}(\cdot)$ if applied to a matrix returns its main diagonal. The covariance $\operatorname{cov} h$ is required in the evaluation of $\mathbb{E}_h\left[H^T\overline{\Gamma}H\right]$ in (18). After some algebraic manipulation, we conclude that $\mathbb{E}_h\left[H^T\overline{\Gamma}H\right] = \overline{H}^T\overline{\Gamma}\overline{H}+C^h$, where C^h is a diagonal matrix constructed by convolving γ with $\operatorname{cov} h$. We can interpret the main diagonals of $\operatorname{cov} h$ and Γ as 2D images and then by slightly abusing the notation write $C^h = \operatorname{diag}\left(\gamma * (\operatorname{diag}(\alpha U^T\Gamma U + C^u + B))^{-1}\right)$, where the outer operator $\operatorname{diag}(\cdot)$ returns a diagonal matrix with pixels of the convolution result arranged on the main diagonal.

The blind deconvolution algorithm is summarized in Algorithm 1.

The most time consuming steps are 3 and 5, which are large linear systems. Fast inversion is not possible because the matrices are composed of convolution and diagonal ones. We thus use conjugate gradients to solve these systems. Steps 10, 11 and 12 update precisions and they are calculated pixel-wise. Since the covariances of u and h are approximated by diagonal matrices, the expectation terms in these update steps are easy to evaluate and likewise in steps 7 and 8.

There are two important implementation details. For the method to be applicable to large blurs (20 pixels wide or more), we must use a pyramid scheme. The above algorithm first runs on a largely downsampled blurred image g.

Algorithm 1

Require: blurred image g

initialize variables:
$h \leftarrow$ delta function,
$\alpha \leftarrow 10$; $\gamma, \nu, \lambda \leftarrow 1$; $\beta, C^h, C^u \leftarrow 0$.
repeat
  $u \leftarrow$ solve $(H^T\Gamma H + C^h + \frac{1}{\alpha}D^T\Lambda D)u = H^T\Gamma g$
  $C^u \leftarrow \text{diag}\left(\gamma * (\text{diag}(\alpha H^T\Gamma H + C^h + D^T\Lambda D))^{-1}\right)$
  $h \leftarrow$ solve $(U^T\Gamma U + C^u + \frac{1}{\alpha}B)h = U^T\Gamma g$
  $C^h \leftarrow \text{diag}\left(\gamma * (\text{diag}(\alpha U^T\Gamma U + C^u + B))^{-1}\right)$
  $\alpha \leftarrow (N + 2a)/\left(\sum_i \gamma_i \mathbb{E}_{u,h}\left[(g_i - H_i u)^2\right] + 2b\right)$ (5)
  $\nu \leftarrow (N+2a)/\left(2\sum_{i=1}^{N}(\gamma_i - \mathbb{E}_{\gamma_i}\left[\log(\gamma_i)\right] - 1) + 2b\right)$
  for all i pixel indices **do**
    $\gamma_i \leftarrow (1 + 2\nu)/\left(\alpha\mathbb{E}_{u,h}\left[(g_i - H_i u)^2\right] + 2\nu\right)$
    $\lambda_i \leftarrow (1 + 2a)/\left(\mathbb{E}_u\left[(D_i u)^2\right] + 2b\right)$
    $\beta_i \leftarrow (1 + 2a)/\left(\mathbb{E}_h\left[(h_i)^2\right] + 2b\right)$
  end for
until stopping criterion is satisfied
return sharp image u and PSF h

The pseudo-code of the proposed blind deconvolution algorithm. The mean $\overline{(\cdot)}$ is omitted for brevity.

The estimated blur h is then upsampled and used as an initialization in the next run of the algorithm with the corresponding scale of g. This is repeated until the original scale of g is reached. We tested various configurations and concluded that 5 scales with a scale factor of 1.5 is sufficient, which was then used in all our experiments. Passing other variables (*e.g.* precisions) between scales except h proved superfluous. The second point is to handle convolution boundary conditions, which is necessary in the case of real images. We solve it naturally by forcing precisions γ_i's that lie along image boundaries to zero. This way the algorithm assumes maximal model discrepancy and completely ignores the boundary regions.

The hyperparameters $a_{(\cdot)}$ and $b_{(\cdot)}$ in (10) and (11) are the only user-defined parameters. Checking update equations (15) and (16) reveals that (a_α, b_α) and (a_ν, b_ν) have a negligible effect since both α and ν are scalars and data terms in the update equations are dominant. In all our experiments we set them thus to zero. On the other hand, (a_λ, b_λ) and (a_β, b_β) are important since both λ and β are calculated for every pixel and their update equations (20) and (24) are influenced by the hyperparameters. We have searched for the best parameters and determined that both a_λ and a_β can be set to zero but b_λ and b_β must be in the interval $(10^{-9}, 10^{-6})$, otherwise the algorithm is unstable. In the case of the noise model with conventional ARD, α and ν (and corresponding hyperparameters a_α, b_α, a_ν, b_ν) are not present. Instead, we have new hyperparameters a_γ, b_γ, which we set in all our experiments to 0 and 10^{-4}, respectively. This corresponds to fixing the number of degrees of freedom to zero.

V. Experiments

We evaluated our method on both synthetically blurred images as well as real images with unknown blur and compared its performance with other state-of-the-art methods (listed in Tab. I) based on similar principle or targeting the same scenario. The presented experiments test the ability to blindly identify regions violating the assumed convolution model in input images and to exclude such regions from blur estimation without affecting its accuracy. Our method is represented in two flavors differing in the error model: Ours-$\alpha\gamma$ with common precision α and estimated degrees of freedom ν in (4), and Ours-γ with conventional ARD and fixed degrees of freedom a_γ and b_γ in (6). The reason for this is to investigate if the more complex model with estimated degrees of freedom is beneficial in practice and in which scenarios. All methods marked in Tab. I as "our implementation" were run with the same pyramid scheme and parameter settings to make a fairer comparison. The method of Pan16 resembles objective wise our method, however uses an entirely different principle not based on VB. Xu10 is a generally robust deconvolution method without extra handling of model discrepancies. Babacan12 is similarly robust deconvolution method but based on VB. Zhong13 is a method which specifically handles Gaussian noise in the input.

In the first of the synthetic experiments we take three images and eight PSFs (see Fig. 6), and blur them to get 24 inputs in

TABLE I
List of Methods Included in the Experimental Evaluation

Ours-$\alpha\gamma$	Our method, full ARD noise model (4).
Ours-γ	Our method, conventional ARD noise model (6).
Tzikas09	Method of [10], our implementation.
Pan16	Method of [2], authors' implementation.
Xu10	Method of [6], authors' implementation.
Zhong13	Method of [5], authors' implementation.
Babacan12	Method of [22], our implementation.

Fig. 6. Original images and PSFs used in our synthetic experiments.

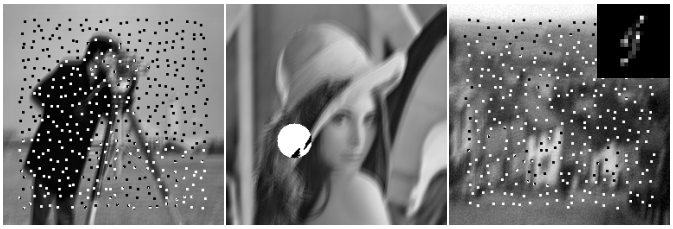

Fig. 7. Sample of input images in our synthetic tests. Left: many small regions with non-Gaussian error (7%), middle: one large region with non-Gaussian error (2.2%), right: combination of non-Gaussian error (5%) and Gaussian noise (SNR=25dB). The ground-truth PSF is in the top-right corner. Results of PSF estimation for these inputs are in Fig. 9 and the corresponding precision γ_is is in Fig. 10.

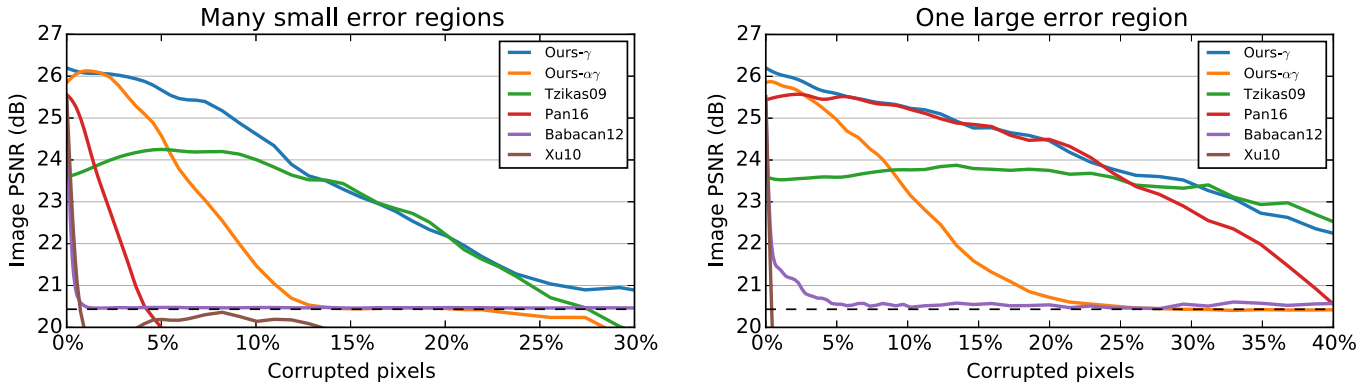

Fig. 8. Restored image PSNR as a function of degree of model violation (higher values mean better performance). Left: increasing number of small model-violating regions, right: one increasingly large region. Dashed line shows blurred image PSNR for reference.

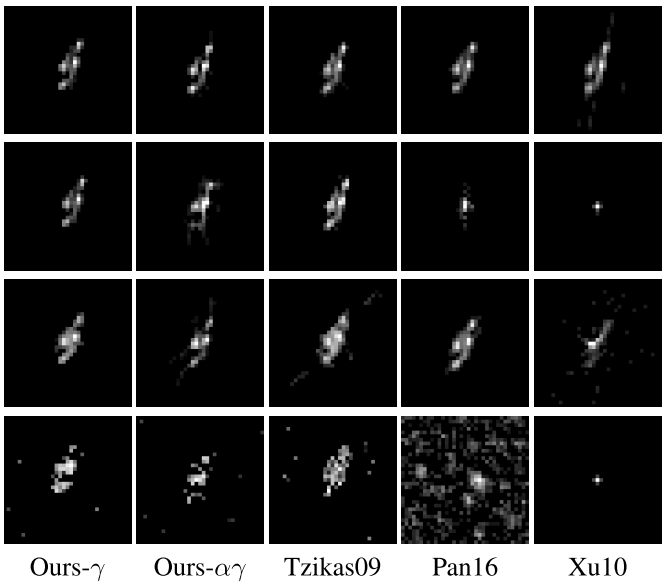

Fig. 9. Example of estimated PSFs in our synthetic experiment, the ground-truth is in Fig.7 top-right. Top row: no intentional input error, second row: many small error regions as in Fig. 7-left, third row: one large error region as in Fig. 7-middle, bottom row: mixture of non-Gaussian and Gaussian noise as in Fig. 7-right.

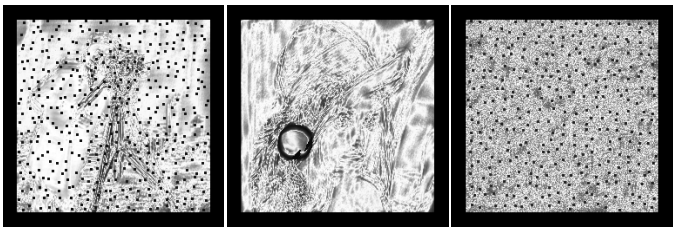

Fig. 10. Estimated γ (precision of the non-Gaussian part of reconstruction error) for the inputs in Fig. 7. Dark (light) pixels mean low (high) precision, i.e. model fidelity.

total. Then we intentionally set random parts of the blurred images to 0 or 1, which represents a non-Gaussian error simulating under-exposed or over-exposed regions (remark: affected pixel i in image g was altered according to $g(i) = \text{round}(1 - g(i))$ to achieve an error of at least .5 and avoid scenarios in which we saturate pixels that are already white or vice versa). To test the influence of erroneous region geometry we corrupted the input in two different ways. In one case we altered an increasing number of small 3x3 pixel squares randomly scattered in the image (Fig 7-left), simulating several small saturated regions in a scene, and in the other case we altered one large blob of pixels (Fig. 7-middle), simulating e.g. an object occluding otherwise a uniformly blurred scene. Then we proceeded with the blur estimation using all the tested methods but applied the same non-blind deconvolution (ℓ^1-TV with boundary handling) to make the results directly comparable. Deconvolution was performed on the blurred image without the intentional corruption (otherwise the input error would constitute most of the restoration error). We repeated this process for many degrees of corruption (percentage of altered pixels), several realizations of random placement of altered pixels and averaged the image estimation accuracy (PSNR) as a function of percentage of damaged pixels. The results of this experiment are summarized in Fig. 8.

In the ideal case, if the model-violating regions are identified exactly, the restored image accuracy should remain constant as long as the rest of the image constitutes enough data for accurate blur estimation. After this point, the accuracy can be expected to drop rapidly. Ours-$\alpha\gamma$ shows certain robustness to input error but its accuracy quickly drops. Ours-γ exhibits the best performance – it retains high PSNR even if a significant portion of the input is corrupted. A good characteristic is that the outlier rejection process works in both scenarios, scattered and compact error regions, as opposed to Pan16. When the error is scattered in the whole area of the input (Fig. 7-middle and Fig. 8-left), Pan16 performance is significantly worse than most of the contenders, but when the error is concentrated (Fig. 7-middle and Fig. 8-right), it performs almost on par with Ours-γ. Tzikas09 shows similar robustness as Ours-γ but performs overall worse, except for too severe input error when all methods start to fail. Babacan12 and Xu10, with no extra outlier handling other than being generally robust, fail as soon as the non-Gaussian error becomes non-negligible.

A qualitative demonstration of the PSF estimation is in Fig. 9, which depicts results for the particular inputs in Fig. 7. The top row corresponds to unaltered ideal input,

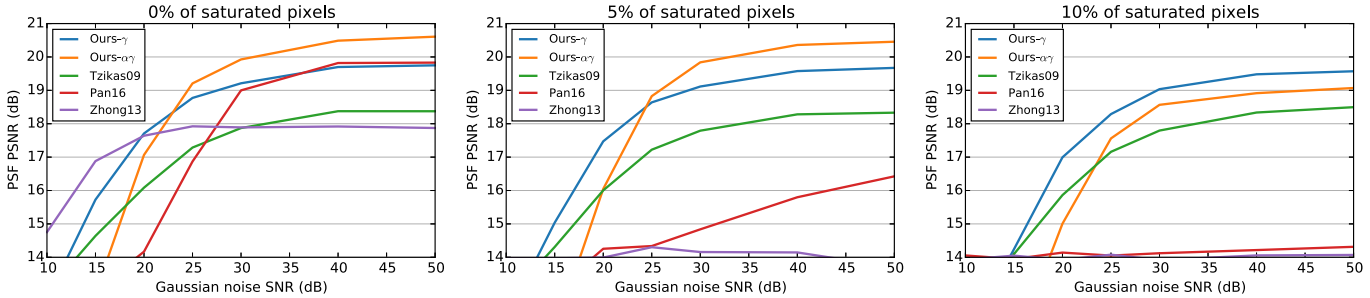

Fig. 11. Estimated blur PSNR as a function of Gaussian noise in input images (SNR on x-axis) for three different levels of non-Gaussian noise (0%, 5% and 10% of pixels randomly saturated).

Fig. 12. Real motion blur, each row represents a new experiment (different photo with different blur). Estimated PSFs are superimposed in the top-right corner. Best viewed on computer screen.

second through fourth row correspond to the three input images in Fig. 7, respectively. All methods perform reasonably well on the error-free input (top row), but with added error some of the methods get significantly worse or fail completely. Pan16 is not sufficiently stable and fails when the erroneous pixels are scattered. The method of Xu10 does not actively seek outliers and as a result fails when the error is non-negligible. Our methods (Ours-γ and Ours-$\alpha\gamma$), and moderately also Tzikas09, produce results comparable to the ideal case, unless the overall input error is too severe (bottom row). The good performance is due to the automatic estimation of the noise precision γ as shown in Fig. 10 for the same input images. The dot pattern on the left image is recognized almost flawlessly. The large disk in the middle image has zeros mainly around its edges. This is because the homogeneous inside of the disk can be explained by any PSF, including the current estimate, and the method thus considers this area as "in accordance with the convolution model". The black frame are enforced zeros (meaning absolute uncertainty) to suppress convolution boundary artifacts.

The difference in performance between Ours-γ and Ours-$\alpha\gamma$ is worth attention. Results in Figs. 8 and 9 imply that the extra complexity provided by Ours-$\alpha\gamma$ is not worth it when the non-Gaussian error is too strong. We conjecture that due to the strong outliers the estimation of the degrees of freedom does not provide meaningful results. The noise distribution with degrees of freedom fixed to zero (albeit improper) is more suitable since it does not penalize strong outliers.

In the second synthetic experiment, we investigated the effect of Gaussian noise in the input on the accuracy of blur estimation. Fig. 11 shows the results of PSF estimation accuracy as a function of the Gaussian noise level for increasing levels of non-Gaussian error (plots left to right). For low levels of non-Gaussian error, Ours-$\alpha\gamma$ method performs best and is superseded by Ours-γ only when the SNR gets too low. This can be explained by the estimation of the degrees of freedom, value of which is relatively high (around 2). The prior model thus approaches a Gaussian distribution, which is the correct model for this experiment. When the non-Gaussian corruption gets too severe, Ours-γ takes the lead. The estimated value of the degrees of freedom decreases, however, not sufficiently (to 0.6), the tails are not heavy enough to accommodate for the severe outliers and the model with fixed zero degrees of freedom performs better. Tzikas09 is similarly stable as Ours-γ but with the performance falling by 1dB in average. Pan16 is slightly less robust to increasing Gaussian noise with zero non-Gaussian noise, and as in the previous experiment, when non-Gaussian noise grows, it starts to fail quickly. Zhong13 shows stable (though somewhat subpar) performance for zero non-Gaussian error (left plot) and even outperforms other methods for exceedingly noisy images. However, it breaks instantly with any presence of non-Gaussian error.

Besides synthetic experiments, we also tested our method on a number of images captured in real world conditions with intentional blur caused either by camera motion or incorrect focus. In the synthetic tests, the convolutional model is violated either completely or not at all, which is fairly well distinguishable even in the blind scenario and the identified outliers can be rejected. In real cases, however, the convolutional model never really holds completely, instead the error changes

Fig. 13. Real out-of-focus blur, each row represents a new experiment (different photo with different blur). Estimated PSFs superimposed in the top-right corner. Best viewed on computer screen.

more or less smoothly from pixel to pixel, so there is no clear distinction between in- and outliers. We demonstrate the results in Fig. 12 and 13 for camera motion and out-of-focus blur, respectively. We first applied all compared blind methods to each image and kept the estimated PSFs. To make the results of the sharp image estimation comparable, we then used the same non-blind method, which was our VB-ARD with the estimated PSFs (i.e. PSF estimation steps 5 and 6 in the algorithm were omitted); all other variables, such as α and γ, were inferred from the data to account for noise, saturations, or other model violations. For these experiments we present only qualitative comparison between our method and several others. We can see that especially Ours-γ method performs well in both motion blur and incorrect focus scenario – the ringing is less prominent while the level of detail and the quality of the PSF estimation is better than in the case of other methods.

More results are provided in the supplementary material. The MATLAB code of the proposed blind deconvolution algorithm is available on our website.

VI. Conclusions

We have presented a blind deconvolution algorithm using the Variational Bayesian approximation with the Automatic Relevance Determination model on likelihood and image and blur priors. The derived coherent algorithm consists of two linear systems of equations that can be efficiently solved with the Conjugate Gradients method, and three simple pixel-wise update equations for noise, image and blur precisions. We have shown that the Automatic Relevance Determination model correctly favors sharp images over the blurred ones, enforces PSF non-negativity and most importantly adjusts for convolution model discrepancies. The experimental section has demonstrated that allowing variable data precision is essential for dealing with outliers such as saturated regions, occlusions or convolution boundary effects. Estimation of the degrees of freedom of the noise prior is beneficial only for Gaussian noise. For non-Gaussian noise distributions, it is more effective to fix the number of degrees of freedom to zero.

Acknowledgment

Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), is greatly appreciated.

References

[1] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding blind deconvolution algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2354–2367, Dec. 2011.

[2] J. Pan, Z. Lin, Z. Su, and M. H. Yang, "Robust kernel estimation with outliers handling for image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2800–2808.

[3] W. Ren, X. Cao, J. Pan, X. Guo, W. Zuo, and M. H. Yang, "Image deblurring via enhanced low-rank prior," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3426–3437, Jul. 2016.

[4] L. Xu, S. Zheng, and J. Jia, "Unnatural L0 sparse representation for natural image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 1107–1114. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6618991

[5] L. Zhong, S. Cho, D. Metaxas, S. Paris, and J. Wang, "Handling noise in single image deblurring using directional filters," in *Proc. IEEE CVPR*, Jun. 2013, pp. 612–619.

[6] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," in *Proc. 11th Eur. Conf. Comput. Vis. (ECCV)*, Berlin, Germany, 2010, pp. 157–170. [Online]. Available: http://portal.acm.org/citation.cfm?id=1886063.1886077

[7] M. S. C. Almeida and L. B. Almeida, "Blind and semi-blind deblurring of natural images," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 36–52, Jan. 2010.

[8] S. Cho and S. Lee, "Fast motion deblurring," *ACM Trans. Graph.*, vol. 28, no. 5, 2009, Art. no. 145.

[9] R. M. Neal, *Bayesian Learning for Neural Networks*. New York, NY, USA: Springer, 1996.

[10] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "Variational Bayesian sparse kernel-based blind image deconvolution with student's-t priors," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 753–764, Apr. 2009.

[11] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. 28, no. 11, pp. 1867–1875, Nov. 1980.

[12] G. Ayers and J. C. Dainty, "Iterative blind deconvolution method and its application," *Opt. Lett.*, vol. 13, no. 7, pp. 547–549, Jul. 1988.

[13] T. F. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 370–375, Mar. 1998.

[14] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," in *Proc. SIGGRAPH*, New York, NY, USA, 2006, pp. 787–794.

[15] R. Molina, J. Mateos, and A. K. Katsaggelos, "Blind deconvolution using a variational approach to parameter, image, and blur estimation," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3715–3727, Dec. 2006.

[16] J. Jia, "Single image motion deblurring using transparency," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4270054

[17] N. Joshi, R. Szeliski, and D. Kriegman, "PSF estimation using sharp edge prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.

[18] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," in *Proc. SIGGRAPH ACM SIGGRAPH* , New York, NY, USA, 2008, pp. 1–10.
[19] D. Perrone and P. Favaro, "A clearer picture of total variation blind deconvolution," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 38, no. 6, pp. 1041–1055, Jun. 2016.
[20] J. Miskin and D. J. MacKay, "Ensemble learning for blind image separation and deconvolution," in *Advances in Independent Component Analysis*, M. Girolami, Ed. London, U.K.: Springer-Verlag, 2000, pp. 123–142.
[21] N. P. Galatsanos, V. Z. Mesarovic, R. Molina, and A. K. Katsaggelos, "Hierarchical Bayesian image restoration from partially-known blurs," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1784–1797, Oct. 2000.
[22] S. D. Babacan, R. Molina, M. N. Do, and A. K. Katsaggelos, "Bayesian blind deconvolution with general sparse image priors," in *Computer Vision—ECCV*. Berlin, Germany: Springer, 2012, pp. 341–355.
[23] F. Sroubek, V. Smídl, and J. Kotera, "Understanding image priors in blind deconvolution," in *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 4492–4496.
[24] J. A. Palmer, K. Kreutz-Delgado, and S. Makeig, "Strong sub- and super-gaussianity," *Latent Variable Anal. Signal Separat.*, vol. 6365, pp. 303–310, Jan. 2010.
[25] D. Wipf and H. Zhang, "Revisiting Bayesian blind deconvolution," *J. Mach. Learn. Res.*, vol. 15, pp. 3775–3814, Nov. 2014. [Online]. Available: http://research.microsoft.com/apps/pubs/ default.aspx?id=192590
[26] P. Ruiz, X. Zhou, J. Mateos, R. Molina, and A. K. Katsaggelos, "Variational Bayesian blind image deconvolution: A review," *Digit. Signal Process.*, vol. 47, pp. 116–127, Dec. 2015. [Online]. Available: http://dx.doi.org/10.1016/j.dsp.2015.04.012
[27] D. F. Andrews and C. L. Mallows, "Scale mixtures of normal distributions," *J. Roy. Statist. Soc. B, Methodol.*, vol. 36, no. 1, pp. 99–102, 1974. [Online]. Available: http://www.jstor.org/stable/2984774
[28] G. Chantas, N. Galatsanos, A. Likas, and M. Saunders, "Variational Bayesian image restoration based on a product of t-distributions image prior," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1795–1805, Oct. 2008.
[29] G. Chantas, N. P. Galatsanos, R. Molina, and A. K. Katsaggelos, "Variational Bayesian image restoration with a product of spatially weighted total variation image priors," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 351–362, Feb. 2010.
[30] H. Zhang, D. Wipf, and Y. Zhang, "Multi-observation blind deconvolution with an adaptive sparse prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1628–1643, Aug. 2014.
[31] W. Dong, H. Feng, Z. Xu, and Q. Li, "Blind image deconvolution using the fields of experts prior," *Opt. Commun.*, vol. 285, no. 24, pp. 5051–5061, 2012.
[32] J. Christmas and R. Everson, "Robust autoregression: Student-t innovations using variational Bayes," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 48–57, Jan. 2011.
[33] T. Kohler, X. Huang, F. Schebesch, A. Aichert, A. Maier, and J. Hornegger, "Robust multiframe super-resolution employing iteratively re-weighted minimization," *IEEE Trans. Comput. Imag.*, vol. 2, no. 1, pp. 42–58, Mar. 2016.
[34] A. Matakos, S. Ramani, and J. A. Fessler, "Accelerated edge-preserving image restoration without boundary artifacts," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 2019–2029, May 2013.
[35] M. S. C. Almeida and M. A. T. Figueiredo, "Deconvolving images with unknown boundaries using the alternating direction method of multipliers," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3074–3086, Aug. 2013.
[36] S. Cho, J. Wang, and S. Lee, "Handling outliers in non-blind image deconvolution," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 495–502.
[37] R. Tezaur, T. Kamata, L. Hong, and S. S. Slonaker, "A system for estimating optics blur psfs from test chart images," *Proc. SPIE*, vol. 9404, pp. 94040D-1–94040D-10, Feb. 2015. [Online]. Available: http://dx.doi.org/10.1117/12.2081458
[38] C. M. Bishop, *Pattern Recognition and Machine Learning*. New york, NY, USA: Springer-Verlag, 2006.
[39] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Proc. CVPR*, 2011, pp. 233–240.
[40] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

Jan Kotera received the master's and Ph.D. degrees in mathematical modeling and computer science from Charles University in Prague, Czech Republic, in 2011 and 2014, respectively, where he is currently pursuing the Ph.D. degree in computer science. Since 2011, he has been with the Institute of Information Theory and Automation, Czech Academy of Sciences. He has authored several conference papers on blind deconvolution and related topics. His current research interests include all aspects of digital image processing and pattern recognition, particularly blind deconvolution and image restoration in general.

Václav Šmídl (M'05) received the Ph.D. degree in electrical engineering from the Trinity College Dublin, Ireland, in 2004. Since 2004, he has been a Researcher with the Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czech Republic. He authored one Springer monograph *Variational Bayes method in Signal Processing*. His research interests are advanced Bayesian methods for signal and image processing.

Filip Šroubek (M'08) received the M.S. degree in computer science from Czech Technical University, Prague, Czech Republic, in 1998, and the Ph.D. degree in computer science from Charles University, Prague, Czech Republic, in 2003. From 2004 to 2006, he was on a post-doctoral position with the Instituto de Optica, CSIC, Madrid, Spain. In 2010 and 2011, he was the Fulbright Visiting Scholar with the University of California at Santa Cruz, Santa Cruz. He is currently with the Institute of Information Theory and Automation, Czech Academy of Sciences, and teaches at Charles University. He has authored eight book chapters and over 70 journal and conference papers on image fusion, blind deconvolution, superresolution, and related topics.