

# Blind Deconvolution with Model Discrepancies

Dirisala Sumalatha (M.Tech Scholar)<sup>1</sup> & Sk.Ayesha (Asst Professor)<sup>2</sup>

Electronics and communication Engineering Department <sup>1,2</sup>

AMARA institute of engineering and technology, Guntur, Andhra Pradesh 522549, India<sup>1,2</sup>

[maheshbabu.marineni@gmail.com](mailto:maheshbabu.marineni@gmail.com)<sup>1</sup>, [ayasha4a0@gmail.com](mailto:ayasha4a0@gmail.com)<sup>2</sup>

## 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 blurs 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 Manuscript received 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  $\epsilon$ .

$$g = h * u + \epsilon \quad (1)$$

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 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), 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, 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 parameterization 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  $p(h|g)$ . 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. LITERATURE SURVEY

[1] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, Blind deconvolution is the recovery of a sharp version of a blurred image when the blur kernel is unknown. Recent algorithms have afforded dramatic progress, yet many aspects of the problem remain challenging and hard to understand. The goal of this paper is to analyze and evaluate recent blind deconvolution algorithms both theoretically and experimentally. We explain the previously reported failure of the naive MAP approach by demonstrating that it mostly favors no-blur explanations. We show that, using reasonable image priors, a naive simulation MAP estimation of both latent image and blur kernel is guaranteed to fail even with infinitely large images sampled from the prior. On the other hand, we show that since the kernel size is often smaller than the image size, a MAP estimation of the kernel alone is well constrained and is guaranteed to succeed to recover the true blur. The plethora of recent deconvolution techniques makes an experimental evaluation on ground-truth data important. As a first step toward this experimental evaluation, we have collected blur data with ground truth and compared recent algorithms under equal settings. Additionally, our data demonstrate that the shift-invariant blur assumption made by most algorithms is often violated.

This paper analyzes the major building blocks of recent blind deconvolution algorithms. We illustrate the limitation of the simple MAP<sub>x;k</sub> approach, favoring the no-blur (delta kernel) explanation. One class of solutions involves explicit edge detection. A more principled strategy exploits the dimensionality asymmetry, and estimates MAP<sub>k</sub> while marginalizing over  $x$ . While the computational aspects involved with this marginalization are more challenging, existing approximations are powerful. We have collected motion blur data with ground truth and quantitatively compared existing algorithms. Our comparison suggests that the variational Bayes approximation [5] significantly outperforms all existing alternatives. The conclusions from our analysis are useful for directing future blind deconvolution research. In particular, we note that

modern natural image priors do not overcome the MAP<sub>x;k</sub> limitation (and in our tests did not change the observation in Section 2). While it is possible that blind deconvolution can benefit from future research on natural image statistics, this paper suggests that better estimators for existing priors may have more impact on future blind deconvolution algorithms. Additionally, we observed that the popular spatially uniform blur assumption is usually unrealistic. Thus, it seems that blur models which can relax this assumption have a high potential to improve blind deconvolution results.

[2] J. Pan, Z. Lin, Z. Su, and M. H. Yang, Estimating blur kernels from real world images is a challenging problem as the linear image formation assumption does not hold when significant outliers, such as saturated pixels and non-Gaussian noise, are present. While some existing non-blind deblurring algorithms can deal with outliers to a certain extent, few blind deblurring methods are developed to well estimate the blur kernels from the blurred images with outliers. In this paper, we present an algorithm to address this problem by exploiting reliable edges and removing outliers in the intermediate latent images, thereby estimating blur kernels robustly. We analyze the effects of outliers on kernel estimation and show that most state-of-the-art blind deblurring methods may recover delta kernels when blurred images contain significant outliers. We propose a robust energy function which describes the properties of outliers for the final latent image restoration. Furthermore, we show that the proposed algorithm can be applied to improve existing methods to deblur images with outliers. Extensive experiments on different kinds of challenging blurry images with significant amount of outliers demonstrate the proposed algorithm performs favorably against the state-of-the-art methods.

In this work, we propose a robust kernel estimation algorithm in which effective edges are selected for deblurring images containing significant amount of outliers. We present detailed analysis on the effects of outliers on kernel estimation. Furthermore, we show that the proposed method can be applied to improve the accuracy of existing blind deblurring methods. In the final deconvolution step, we develop

a robust method to restore the latent image under the guidance of the proposed outlier-aware function where the effects of outliers are minimized. Extensive experimental evaluations on real images and benchmark datasets demonstrate the proposed algorithm performs favorably against the state-of-the-art methods for uniform as well as non-uniform deblurring.

[3] W. Ren, X. Cao, J. Pan, X. Guo, W. Zuo, and M. H. Yang, Low rank matrix approximation has been successfully applied to numerous vision problems in recent years. In this paper, we propose a novel low rank prior for blind image deblurring. Our key observation is that directly applying a simple low rank model to a blurry input image significantly reduces blur even without using any kernel information, while preserving important edge information. The same model can be used to reduce blur in the gradient map of a blurry input. Based on these properties, we introduce an enhanced prior for image deblurring by combining the low rank prior of similar patches from both the blurry image and its gradient map. We employ a weighted nuclear norm minimization method to further enhance the effectiveness of low rank prior for image deblurring, by retaining the dominant edges and eliminating fine texture and slight edges in intermediate images, allowing for better kernel estimation. In addition, we evaluate the proposed enhanced low rank prior for both uniform and non-uniform deblurring. Quantitative and qualitative experimental evaluations demonstrate that the proposed algorithm performs favorably against the state-of-the-art deblurring methods.

In this paper, we present a novel enhanced low rank prior for blind image deblurring. The low rank properties of both intensity and gradient maps from image patches are exploited in the proposed algorithm. We present a weighted nuclear norm minimization approach based low rank properties to effectively recover latent images. Experimental results on benchmark datasets show that the proposed algorithm performs favorably against the state-of-the-art deblurring methods.

### III. PROPOSED METHOD

#### A. Automatic Relevance Determination

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

$$\epsilon_i = g_i - H u_i = g_i - U_i h, \quad i = 1, \dots, N, \quad (2)$$

where  $H$  and  $U$  are convolution matrices performing convolution with the blur and latent image, respectively, 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  $\alpha$ , i.e.

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

Such assumption leads to the common 2 data term  $\sum_i (g_i - U_i h)^2$ . 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 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, we propose using a Gaussian distribution with pixel-dependent factors  $\gamma_i$  modulated by the overall noise precision  $\alpha$ . The error model of is then defined as

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

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  $\alpha$  and  $\gamma_i$  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  $\gamma$ . Let  $G$  denote the standard Gamma distribution, defined as  $G(\xi|a, b) = (1/a) b^a \xi^{a-1} \exp(-b\xi)$ . We define the  $\gamma$  prior as

$$p(\gamma|v) = \prod_i \mathcal{G}(\gamma_i|v, v) \quad (5)$$

Marginalizing  $p(\epsilon|\alpha, \gamma)p(\gamma|v)$  over  $\gamma$  gives us the Student's t-distribution with zero mean, precision  $\alpha$  and degrees of freedom  $2v$ . From the above model it follows that the mean of  $\gamma_i$  is equal to  $a/b = v/v = 1$ . If  $v$  becomes large then  $G(\gamma_i|v, v)$  tends to the delta distribution at 1 and the error model will be just a Gaussian distribution. As  $v$  decreases, tails decay more slowly and  $\gamma_i$  will be allowed to adjust and automatically suppress outliers violating the acquisition model. The conventional ARD model used e.g. in [10] is

$$p^*(\epsilon|\gamma) = \prod_i \mathcal{N}(\epsilon_i|0, \gamma_i^{-1})$$

$$p^*(\gamma_i) = \mathcal{G}(\gamma_i|a_\gamma, b_\gamma) \quad (6)$$

The marginal distribution of this prior over  $\gamma$  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  $a_\gamma, b_\gamma$  as small as possible, yielding Student's t-prior with infinite variance. Estimation of the hyper parameters  $a_\gamma, b_\gamma$  via a numerical MAP method has been proposed in [10].

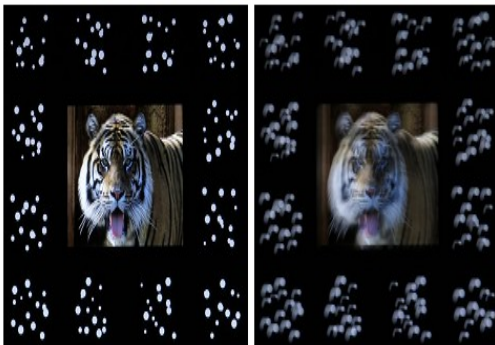


Fig.1. Sharp (left) and intentionally blurred (right) image pair acquired for accurate calculation of the blur PSF from the known patterns surrounding the image.

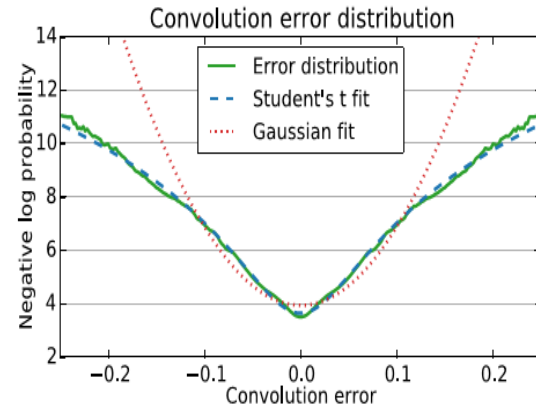


Fig. 2. Convolution error distribution in the case of real motion blurs (solid green).

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,  $2v = 3.5, \alpha = 1.2 \cdot 10^5$ ) 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 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, 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  $p(\cdot)$ . 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]

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

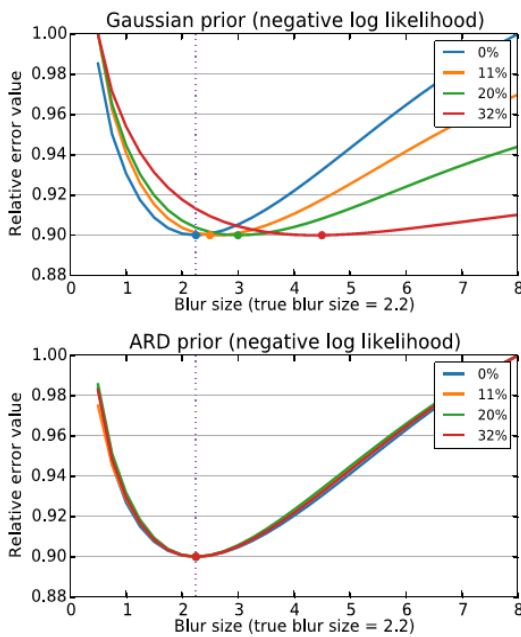


Fig.3. Relative data term value (negative log likelihood) for commonly used Gaussian (top) and ARD (bottom) priors (low value means high probability of the particular PSF) for image blurred with PSF of size 2.2.

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.

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.

### B. Variational Bayesian Inference

There are many examples of the VB inference applied to blind deconvolution in the literature; see e.g. [10]. 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 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 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

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

where  $-1$  is a diagonal covariance matrix having the inverse of the precision vector  $\gamma$  on the main diagonal,  $\Gamma = \text{diag}(\gamma)$ . Let us recall that the precision  $\gamma_i$  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

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

where  $D_i$  is the first order difference at the  $i$ -th pixel and  $\Lambda = \text{diag}(\lambda)$ . 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  $\lambda_i$ 's. It was advocated in [1] 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.

$$\begin{aligned}
 p(h|\beta) &= \mathcal{N}(h|0, B^{-1}) = \prod_i \mathcal{N}(h_i|0, \beta_i^{-1}) \\
 \alpha \prod_i \beta_i^{1/2} \exp\left(-\frac{\beta_i}{2} h_i^2\right) \quad (9)
 \end{aligned}$$

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

$$p(\lambda_i) = \mathcal{G}(\lambda_i|a_\lambda, b_\lambda)$$

$$p(\beta_i) = \mathcal{G}(\beta_i|a_\beta, b_\beta) \quad (10)$$

And for the error precisions according to (4) and (8) we have

$$p(\alpha) = \mathcal{G}(\alpha|a_\alpha, b_\alpha)$$

$$p(\gamma_i|\nu) = \mathcal{G}(\gamma_i|\nu, \nu)$$

$$p(\nu) = \mathcal{G}(\nu|a_\nu, b_\nu) \quad (11)$$

The hyper parameters  $a(\cdot)$  and  $b(\cdot)$  are user-defined constants. Let  $Z = \{u, h, \alpha, \nu, \{\gamma_i\}, \{\lambda_i\}, \{\beta_i\}\}$  denote all the unknown variables and  $Z_k$  its particular member indexed by  $k$ . Using the above defined distributions, the posterior  $p(Z|g)$  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  $p(Z|g)$  by the factorized distribution

$$p(Z|g) \approx q(Z) \\ = q(u)q(h)q(v)q(\gamma)q(\lambda)q(\beta) \quad (12)$$

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

$$\log q(z_k) \propto \mathbb{E}_{l \neq k} [\log p(Z|g)] \quad (13)$$

where  $\mathbb{E}_{l \neq k}$  denotes expectation with respect to all factors  $q(Z_l)$  except  $q(Z_k)$ . Formula (13) gives implicit solution, because each factor  $q(Z_k)$  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 [10] 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 nonGaussian part of the error modeled by precision  $\gamma$ . 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

$$\bar{\gamma}_i = \frac{1 + 2\bar{v}}{\bar{\alpha} \mathbb{E}_{u,h} [(g_i - H_i u)^2] + 2\bar{v}} \quad (14)$$

where  $(\cdot)$  denotes a mean value. Relating the inference to the classical minimization of energy function  $-\log p(u, h|g)$ , the precision  $\gamma_i$  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  $v/\alpha$ ) and it is updated during iterations, as the image and blurs 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  $\gamma$  depends, apart from  $u$  and  $h$ , only on the mean values  $\alpha$  and  $v$ . Using again the VB inference formula (13), one can deduce that both  $q(\alpha)$  and  $q(v)$  are Gamma distributions with mean values

$$\bar{\alpha} = \frac{N + 2a_\alpha}{\sum_{i=1}^N \bar{\gamma}_i \mathbb{E}_{u,h} [(g_i - H_i u)^2] + 2b_\alpha} \quad (15)$$

$$\bar{v} = \frac{N + 2a_v}{2 \sum_{i=1}^N (\bar{\gamma}_i - \mathbb{E}_{\gamma_i} [\log(\gamma_i)] - 1) + 2b_v} \quad (16)$$

The update equation of  $v$  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  $\alpha$  is expected to be inversely proportional to the level of Gaussian noise in the input image. It is therefore interesting to observe how  $\alpha$  behaves during iterations. After the initialization, when the reconstruction error is high, the weight  $\alpha$  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,  $\alpha$  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/regularize 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



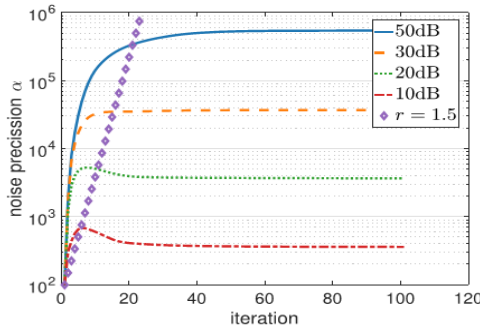


Fig.4. Estimated noise precision as a function of iterations:

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  $\alpha_k = 1.5\alpha_{k-1}$ .

Stop when the correct  $\alpha$  (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).

### C. Image Prior

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

$$\log q(u) = -E_{h,\alpha,\gamma,\lambda} [\alpha(g - Hu)^T \Gamma (g - Hu) + u^T D^T \Lambda D u] \quad (17)$$

where the terms independent of  $u$  are omitted. The distribution  $q(u)$  is a normal distribution. The mean  $\bar{u}$  and covariance  $\text{cov}(u)$  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

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

And for the covariance we get

$$\text{cov } u = (\bar{\alpha} \mathbb{E}_h [H^T \bar{\Gamma} H] + D^T \bar{\Lambda} D)^{-1} \quad (19)$$

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

$$\bar{\lambda}_i = \frac{1 + 2a_\lambda}{\mathbb{E}_u [(D_i u)^2] + 2b_\lambda} \quad (20)$$

The parameter  $b_\lambda$  plays the role of relaxation, as it prevents division by zero in the case  $D_i u = 0$ .

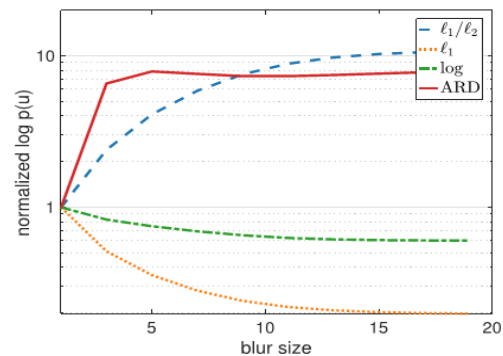


Fig.5. Comparison of priors: The graph shows  $-\log p(u)$  of priors as a function of amount of blurring.

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 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 means values calculated on various images (photos of nature, human faces, and buildings).

It was demonstrated in [1] that commonly used  $p(p < 1)$  priors with  $|D_i u|^p$  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  $\lambda$ , which in the limit for  $\alpha\lambda \rightarrow 0$  yields.

$$\begin{aligned} & p(u) \\ = & \int p(u, \lambda) d\lambda \\ = & \int p(u|\mu) p(\lambda) d\lambda \alpha \prod_i \exp\left(-\frac{1}{2} \log\left(\frac{D_i u}{\alpha}\right)^2 + b\right) \end{aligned} \quad (21)$$

The marginalized ARD image prior is of the exponential form with exponent  $\log(D_i u)^2 + \text{const}$ . 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  $p$  priors with  $p \rightarrow 0$ . We calculated  $-\log p(u)$  on natural images (photos of nature, human faces, and 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 1 prior decrease 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  $\lambda_i$ '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 [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.

#### D. 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  $\log q(h) = -Eu, \alpha, \gamma, \beta \alpha(g - Uh)^T (g - Uh) + h^T B h$ . Then the mean  $\bar{h}$  is the solution of the linear system

$$(\mathbb{E}[U^T \bar{\Gamma} U] + \bar{\alpha}^{-1} \bar{B}) \bar{h} = \bar{U}^T \bar{\Gamma} g \quad (22)$$

And the covariance is

$$\text{cov } h = (\bar{\alpha} \mathbb{E}_u [U^T \bar{\Gamma} U] + \bar{B})^{-1} \quad (23)$$

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

$$\bar{\beta}_i = \frac{1 + 2\alpha\beta}{\mathbb{E}_h [h_i^2] + 2b_\beta} \quad (24)$$

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  $\text{cov } h$  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  $a \rightarrow 0$ . If the hyper parameters 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  $\text{cov } h$  has an additional positive influence on the behavior of the PSF precision  $\beta$ . The denominator of (24) expands to  $h^2 i + \text{cov } h_i + 2b$ . From (23) it follows that  $\text{cov } h_i$  is inversely proportional to  $\alpha + \beta i$ . We can ignore  $\gamma$ , since it is in average around 1 anyway as it captures only local nonGaussian errors. We have seen in Fig. 4 that  $\alpha$  starts small, which implies larger  $\text{cov } h_i$  and thus small PSF precision  $\beta i$ . Small  $\beta i$  loosely constrains the estimation of the PSF  $h$  during initial iterations. As  $\alpha$  increase later on,  $\text{cov } h_i$  decreases and  $\beta i$  increases, which helps to fix the estimated values of  $h$ .

### E. Algorithm

All equations in the VB inference are relatively easy to solve, except for the calculation of covariance matrices  $\text{cov } u$  in (19) and  $\text{cov } h$  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  $E[\cdot]$ . To tackle this problem, we approximate precision matrices by diagonal ones. This is different from Tzikas's work [10], where  $\text{cov } h$  is approximated by a convolution matrix. The experimental section demonstrates that the diagonal approximation performs better. We show the approximation procedure on  $\text{cov } h$  and calculation of  $u$ . The approximation of  $\text{cov } u$  and calculation of  $h$  is similar. First we approximate the covariance matrix  $\text{cov } h$  by inverting only the main diagonal of the precision matrix, i.e.,

$$(\text{Diag}(\alpha \mathbb{E}u \ U \ T \ U + B))^{-1}.$$

Here we use the syntax of popular numerical computing tools such as MATLAB, Python or R, and assume that the operator  $\text{diag}(\cdot)$  if applied to a

matrix returns its main diagonal. The covariance  $\text{cov } h$  is required in the evaluation of  $Eh \ H \ T \ H$  in (18). After some algebraic manipulation, we conclude that  $Eh \ H \ T \ H = H \ T \ H + Ch$ , where  $Ch$  is a diagonal matrix constructed by convolving  $\gamma$  with  $\text{cov } h$ . We can interpret the main diagonals of  $\text{cov } h$  and  $Ch$  as 2D images and then by slightly abusing the notation write  $Ch = \text{diag } \gamma * (\text{diag}(\alpha \mathbb{E}u \ T \ U + C_u + B))^{-1}$ , where the outer operator  $\text{diag}(\cdot)$  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 covariance's 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 down sampled blurred.

---

#### Algorithm 1

---

**Require:** blurred image  $g$

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

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

---

Algorithm 1 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  $\gamma_i$ 's that lie along image boundaries to zero.

This way the algorithm assumes maximal model discrepancy and completely ignores the boundary regions. The hyper parameters  $a(\cdot)$  and  $b(\cdot)$  in (10) and (11) are the only user-defined parameters. Checking update equations (15) and (16) reveals that  $(a\alpha, b\alpha)$  and  $(a\nu, b\nu)$  have a negligible effect since both  $\alpha$  and  $\nu$  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,  $(a\lambda, b\lambda)$  and  $(a\beta, b\beta)$  are important since both  $\lambda$  and  $\beta$  are calculated for every pixel and their update equations (20) and (24) are influenced by the hyper parameters.

We have searched for the best parameters and determined that both  $a\lambda$  and  $a\beta$  can be set to zero but  $b\lambda$  and  $b\beta$  must be in the interval  $(10^{-9}, 10^{-6})$ , otherwise the algorithm is unstable. In the case of the noise model with conventional ARD,  $\alpha$  and  $\nu$  (and corresponding hyper parameters  $a\alpha, b\alpha, a\nu, b\nu$ ) are not present. Instead, we have new hyper parameters  $a\gamma, b\gamma$ , 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.

#### IV. EXPERIMENTAL RESULTS

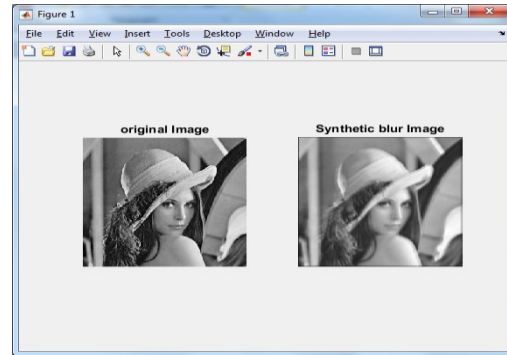


Fig1: (a) Original Image (b) Synthetic Blur Image

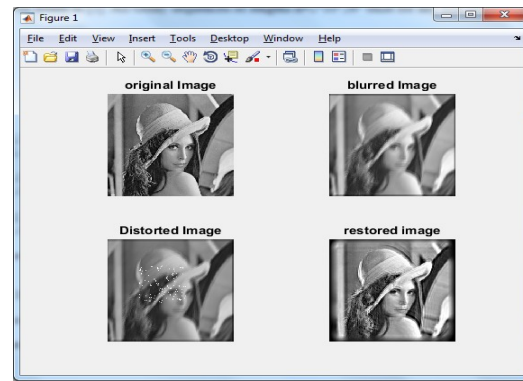


Fig2: (a) Original Image (b) Blurred Image (c) Distorted Image (d) Restored Image

We considered distortion by 1) small regions 2) large regions 3) Gaussian large regions

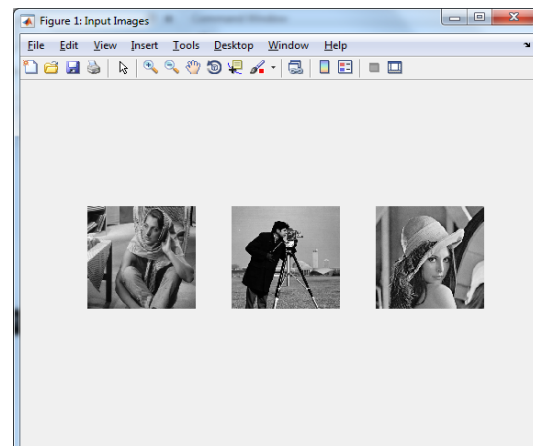


Fig3: Original images available in database

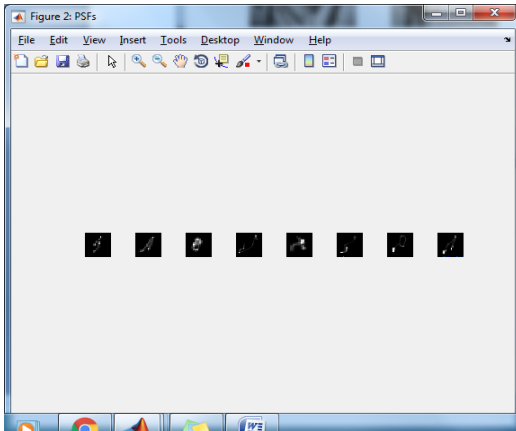


Fig4: Original images and PSFs used in our synthetic experiments.

## V.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.

## 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].

[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.

[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. Tzidakis, 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.