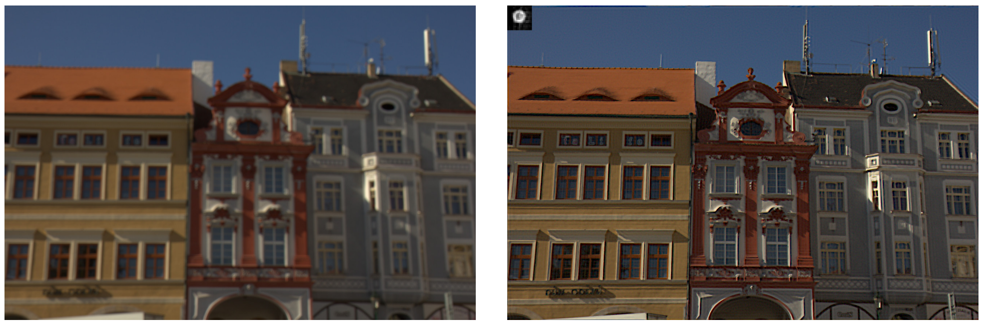
UNDERSTANDING IMAGE PRIORS IN BLIND DECONVOLUTION

Filip Šroubek[⋆]*, Václav Šmídl*[⋆]*, Jan Kotera*[⋆†] [*]

[⋆]UTIA, Academy of Sciences of the Czech Republic, Prague, Czech Republic
[†]Charles University in Prague, Faculty of Mathematics and Physics, Czech Republic



(a) blurred input (b) blind deconvolution

Fig. 1. Deconvolution of a real photo blurred due to incorrect focus: (a) input out-of-focus image; (b) reconstruction using the proposed method with estimated blur kernel in the top left corner ($4\times$ enlarged).

ABSTRACT

Removing blurs from a single degraded image without any knowledge of the blur kernel is an ill-posed blind deconvolution problem. Proper estimators together with correct image priors play a fundamental role in accurate blind deconvolution. We demonstrate a superior performance of the variational Bayesian estimator and discuss suitability of automatic relevance determination distributions as image priors. Restoration of real photos blurred by out-of-focus and motion blur, and comparison with a state-of-the-art method is provided.

Index Terms— blind deconvolution, variational Bayes, marginalization, automatic relevance determination

1. INTRODUCTION

We will assume a classical acquisition model, in which a noisy observed image g is a result of convolution between a latent image u and an unknown kernel h (blur), and corruption with additive noise n,

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

All the variables are 2D random vector fields and they are characterized by corresponding probability distributions denoted as $p(h)$, $p(u)$, and $p(n)$.

If we have only one observation g and we have no knowledge of the blur h, the problem of finding u is extremely ill-posed. One way to tackle this problem is to assume a parametric model of blurs and search in the space of parameters and not in the full space of blurs. Real blurs always differ slightly from their parametric models, which prevents the parametric methods to find an exact solution and we will not consider them any further.

The Bayesian paradigm dictates that inference of u and h from the observed image g is done by modeling a posteriori probability distribution $p(u,h|g) \propto p(g|u,h)p(u)p(h)$. Here we assume that prior distributions $p(u)$ and $p(h)$ are independent, which is always true in practice.

The conditional distribution $p(g|u,h)$ is given by model (1). The prior $p(u)$ forces some type of image smoothness. The prior $p(h)$ can be similar to $p(u)$, but it is rectified to force positivity. Estimating the pair $(\hat{u},\hat{h})$ is equivalent to maximizing the posterior $p(u,h|g)$, which is commonly referred to as the maximum *a posteriori* (MAP) approach. The simplest but also the most common method maximizes the posterior in an alternative manner with respect to u and h. Unfortunately, the posterior has very uneven shape with many local peaks and alternating maximization without any modifications ends in a local maximum, which is rarely a correct solution.

There has been a considerable effort of the image processing community in the last three decades to find a reliable algorithm for blind deconvolution. Older algorithms were not

[*]Thanks to GACR agency project GA13-29225S for funding and J.K. also thanks GA UK 938213 for funding.

sufficiently general and their performance depended on initial estimates of blurs [1–3].

Over the last few years, blind deconvolution experiences a renaissance. The key idea of 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. The activity started with the work of Fergus *et al.* [4] and Molina *et al.* [5], who applied Variational Bayes (VB) to approximate the posterior $p(u, h|g)$ by a simpler distribution $q(u, h) = q(u)q(h)$. Other authors [6–9] stick to the "good old" alternating MAP approach, but by using ad hoc steps, which often lack rigorous explanation, they converge to a correct solution. Levin *et al.* in [10] proved that a proper estimator matters more than the shape of priors. They showed that marginalizing the posterior with respect to the latent image u leads to the correct solution of h. The marginalized probability $p(h|g)$ can be expressed in a closed form only for simple priors that are, e.g., Gaussian. Otherwise approximation methods such as VB [11] or the Laplace approximation [12] must be used.

In this paper, we use VB approximation to solve blind deconvolution and automatically determine all parameters in image and blur priors including noise variance. We analyze in detail the category of so-called Automatic Relevance Determination (ARD) priors [13], compare them with other commonly used priors, and explain their superiority. Finally, we compare ourselves with, to our knowledge, the best working blind deconvolution algorithm in [9].

2. PROBLEM DEFINITION

Since we are in the discrete domain, convolution is equivalent to vector-matrix multiplication and we rewrite the model for every image pixel i as

$$g_i = H_i u + n_i = U_i h + n_i\,, \tag{2}$$

where H and U are convolution matrices performing convolution with the blur and latent image, respectively, and h and u are now column vectors containing lexicographically ordered elements of the 2D random vector fields. Subscript i denotes the i-th element (row) of a vector (matrix).

The noise n is assumed to be i.i.d. with a zero mean and unknown variance $1/\gamma, p(n_i) \equiv N(n_i|0, \gamma^{-1})$. Plugging into (2) we get the likelihood function

$$\begin{aligned} p(g|u, h, \gamma) &= \prod_i N(g_i|H_i u, \gamma^{-1}) \\ &\propto \prod_i \gamma^{1/2} \exp\left\{-\frac{\gamma}{2}(g_i - H_i u)^2\right\}\,. \end{aligned} \tag{3}$$

In blind deconvolution, the image prior plays an essential role, since it provides necessary constraints that make the problem better posed. Intensity values have very random distributions that depend on image content, which is an undesirable property. Image features, such as derivatives, are sparse and their histograms fit well into various heavy-tailed distributions (Laplacian or Gaussian Mixture) irrespective of image content. Therefore various image priors based on image gradient has been proposed in the literature [4, 14]. It has been shown that Gaussian priors with variable precision over pixels (ARD priors) accurately model heavy-tailed distributions [15]. We will use image derivatives (D) as features and the Gaussian image prior with variable precision becomes

$$\begin{aligned} p(u|\lambda) = N(Du|0, \Lambda) &= \prod_i N(D_i u|0, \lambda_i^{-1}) \\ &\propto \prod_i \lambda_i^{1/2} \exp\left\{-\frac{\lambda_i}{2}(D_i u)^2\right\}\,, \end{aligned} \tag{4}$$

where D_i is the first order difference in the i-th pixel[1]. The image prior looks very simple as it is a product of normal distributions of image features (in our case image derivatives). Its ability to capture sparse features is due to different precision λ_i in every pixel, which is unknown and must be determined from data.

Blur priors play far less important role. A general blind deconvolution method must work for any type of blur kernel. We thus avoid any sparsity constraints, which were commonly applied in blind deconvolution methods in past. The only constraint we force is positivity. In order to have closed-form solutions in the VB framework and keep the problem tractable, we force positivity not directly on h but on blur prior mean values. The blur prior is simply

$$\begin{aligned} p(h|v, \beta) &= \prod_i N(h_i|v_i, \beta^{-1}) \\ &\propto \prod_i \beta^{1/2} \exp\left\{-\frac{\beta}{2}(h_i - v_i)^2\right\} \end{aligned} \tag{5}$$

and the distribution of mean values v_i is improper distribution

$$p(v_i) \propto \exp\{-\psi(v_i)/2\},\ \text{where}\ \psi(v_i) = \begin{cases} \infty & v_i < 0, \\ 0 & v_i \geq 0. \end{cases} \tag{6}$$

The above distributions are governed by a set of unknown parameters γ, $\{\lambda_i\}$, β. Each parameter is a random variable with a conjugate prior. In our case all the parameters are precision and the conjugate priors are gamma distributions:

$$\begin{aligned} p(\gamma) &= \mathrm{Gam}(\gamma|a, b)\,, \\ p(\lambda_i) &= \mathrm{Gam}(\lambda_i|a, b)\,, \\ p(\beta) &= \mathrm{Gam}(\beta|a, b)\,, \end{aligned} \tag{7}$$

where a and b are parameters of the gamma distributions. We set a close to zero as it forces the corresponding factor to be influenced only by the data and not by its prior. The gamma distribution has a finite integral only for $a > 0$ and thus we write $a \to 0$.

[1] In 2D, D consideres both derivatives in x and y independently.

Let $\mathcal{Z} \equiv \{u, h, \gamma, \{\lambda_i\}, v, \beta\}$ denote unknown random variables to be estimated. Having all the necessary probability distributions, the joint distribution is given by

$$p(g, \mathcal{Z}) = p(g|u, h, \gamma)p(\gamma)p(u|\lambda)p(\lambda)p(h|v, \beta)p(v)p(\beta)\,. \tag{8}$$

The VB framework provides an elegant way to approximate the joint distribution by a variational distribution $q(\mathcal{Z})$ which factorizes between the variables

$$q(\mathcal{Z}) \equiv q(u)q(h)q(\gamma)q(\lambda)q(v)q(\beta)\,. \tag{9}$$

The equation for estimating the factors is as simple as

$$\ln q(\mathcal{Z}_l) \propto \mathbb{E}_{k\neq l}[\ln p(g, \mathcal{Z})]\,, \tag{10}$$

where the expectation $\mathbb{E}_{k\neq l}$ is with respect to all factors $q(\mathcal{Z}_k)$ except $q(\mathcal{Z}_l)$. Since the factors $q(\cdot)$ depend on statistics of other variables in $\mathcal{Z}$, an iterative VB algorithm is used, which cycles through the factors and updates each probability distribution using (10).

3. ARD IMAGE PRIOR

To better understand the ARD prior $p(u|\lambda)$ in (4), we compare its effect with other priors. In order to do so, we need to express $p(u)$ not conditioned on λ. This is accomplished by marginalization,

$$p(u) = \int p(u, \lambda)\mathrm{d}\lambda = \int p(u|\lambda)p(\lambda)\mathrm{d}\lambda\,. \tag{11}$$

After substituting from (4) and (7), a closed-form solution exists and for our case of $a \to 0$ it is of the form

$$p(u) \propto \prod_i \exp\left\{-\frac{1}{2}\ln((D_i u)^2 + b)\right\}\,. \tag{12}$$

Note that this prior on the image features resembles Stundet's-t distribution.

Remark 1. Let us compare the exponent of the prior $-\ln p(u) \propto \ln\sum_i((D_i u)^2 + b)$ with other commonly used priors, such as the Laplacian prior or standard Gaussian prior. The exponent of the Laplacian prior is the $L1$-norm of image differences, which is also called Total Variation (TV), $\sum_i |D_i u|$. The exponent of the simple Gaussian prior is the $L2$-norm of image differences, $\sum_i (D_i u)^2$. Fig. 2 compares all three priors. One can see that the ARD prior (12) has a non-convex exponent. The same non-convex function was e.g. mentioned in [16] as a regularization term for image segmentation. Many state-of-the-art blind deconvolution methods [17] started to use non-convex functions, such as Lp-norms with $p < 1$, in the prior exponent to improve performance. However, it was shown in [10] that these priors favor (give higher probability to) blurred images over the sharp ones, which goes against intuition. On contrary, the ARD prior favors sharp images as demonstrated in Fig. 3. Our prior belongs to the family of non-convex functions and in addition favors sharp images, which are the reasons for its superiority.

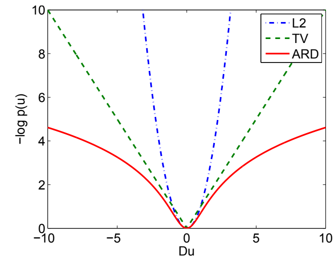


Fig. 2. Comparison of priors: The graph shows prior's exponent, $-\ln p(u)$, as a function of image derivatives. The Gaussian prior with variable precision (ARD - solid red line) is non-convex compare to the priors with L2 or L1 (TV) norms in the exponent.

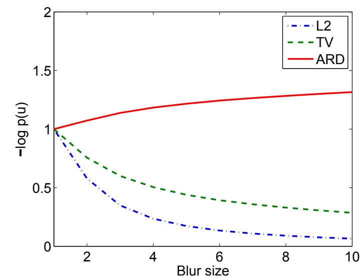


Fig. 3. Comparison of priors: The graph shows prior's exponent, $-\ln p(u)$, as a function of amount of blurring. The Gaussian prior with variable precision (ARD - solid red line) increases with increasing blur whereas the L2 or L1-norm priors decrease. The value of priors is normalized to give 1 on sharp images (0 blur size). The curves show mean values calculated on various images (nature, faces, buildings).

Implementing the marginalized image prior (12) in the VB algorithm is not straightforward [14]. Instead we consider the original conditioned prior (4) and find the solution of the image factors $q(u)$ and $q(\lambda)$ according to (10). For the image factor we get

$$\begin{aligned} q(u) &= N(u|\bar{u}, \mathrm{cov}(u)) \\ &\propto \exp\left\{-\mathbb{E}_{h,\gamma,\lambda}\left[\gamma\|g - Hu\|^2 + u^T D^T \Lambda D u\right]\right\}\,, \end{aligned} \tag{13}$$

where terms independent of u are omitted. We see that the image factor $q(u)$ is again a normal probability distribution as the image prior $p(u|\lambda)$. One can get update equations for the mean $\bar{u}$ and covariance $\mathrm{cov}(u)$ by taking the first and second derivatives of $\ln q(u)$. For the sake of brevity, we show only the update equation for the mean, which is the solution of the linear system

$$[\mathbb{E}_{h,\gamma}[\gamma H^T H] + D^T \bar{\Lambda} D]\bar{u} = \bar{\gamma}\overline{H}^T g\,, \tag{14}$$

where $\overline{(\cdot)}$ denotes the mean value. Note that to evaluate the expectation $\mathbb{E}_{h,\gamma}[\gamma H^T H]$ we also need the covariance h, which is determined from $q(h)$. The mean values of other variables are determined after solving their corresponding factors q. Here we are interested in $\bar{\lambda}_i$'s that form the diagonal matrix $\bar{\Lambda}$. The solution of the image prior precision factor $q(\lambda)$ is of the gamma form

$$q(\lambda) = \prod_i \mathrm{Gam}(\lambda_i | a_i, b_i) \tag{15}$$

and thus for $a \to 0$

$$\bar{\lambda}_i = \frac{a_i}{b_i} = \frac{1}{\mathbb{E}_u[(D_i u)^2] + 2b}\,. \tag{16}$$

The gamma parameter b plays the role of a relaxation parameter, since it prevents the discontinuity of derivatives at $D_i u = 0$.

Remark 2. Let us draw a parallel with another well-established method for solving complicated constrained least-squares deconvolution problems. The equivalent optimization problem that we partially solve in (8) has the form

$$\min_u \gamma (g_i - H_i u)^2 + \frac{1}{2}\ln((D_i u)^2 + b)\,. \tag{17}$$

The standard method for minimizing such energy functions is the so-called half-quadratic (HQ) algorithm [18]. If we ignore expectation in update equations (14) and (16) then these two equations are exactly the update equations of HQ. Contrary to HQ, the VB framework estimates also covariances and the expectation in (16) produces an extra term with $\mathrm{cov}(u)$. The prior precision $\bar{\lambda}_i$ updated in (16) is inversely proportional to the currently estimated image derivatives, i.e. λ's align with edges. Fig. 4 compares the effect of the covariance term, which clearly helps to emphasize edges.

(a) image (b) without $\mathrm{cov}(u)$ (c) with $\mathrm{cov}(u)$

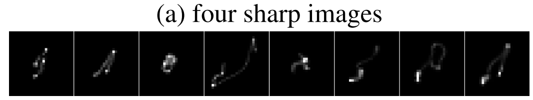
Fig. 4. Visualization of the prior precision λ: (a) sharp image; (b) prior precisions as estimated in every pixel using VB without covariance and (c) with covariance.

4. EXPERIMENTS

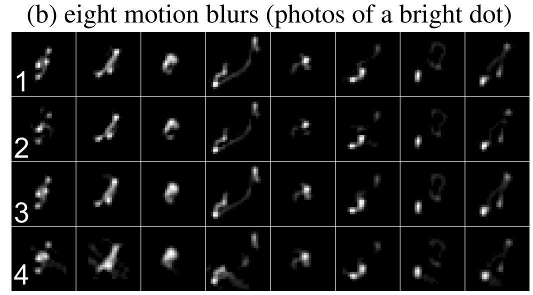
Fig. 1 shows an example of deconvolved real photo (1 Mpixel) taken by a DSLR camera with wrong focus using the VB algorithm with the ARD image prior. Note that the estimated blur kernel well matches the shape of the aperture hole.

To evaluate the performance of the proposed VB algorithm, we calculated mean squared error (MSE) of estimated blurs on the data set in [19] and compared it with arguably the best working blind deconvolution method in [9]. The results are summarized in Fig. 5. In total 32 blurred photos were used; 4 images in (a) photographed 8 times. Blurs in (c) were estimated by the proposed VB algorithm and they match the ideal blurs in (b) that were obtained as pictures of bright dots. The graph in (d) shows that the proposed algorithm outperforms in terms of MSE the method of Xu *et al.* [9] in most of the cases. The ideal blurs in (b) were taken as ground truth for the MSE calculation.

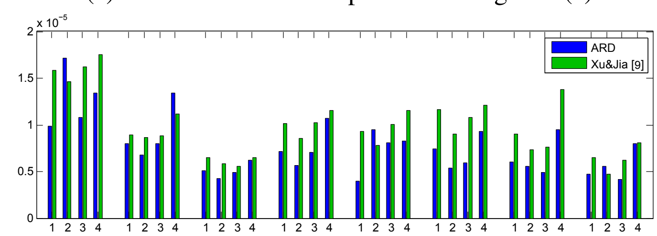
(a) four sharp images



(b) eight motion blurs (photos of a bright dot)



(c) blurs estimated from photos of images in (a)



(d) MSE of estim. blurs; ARD-proposed method vs. [9]

Fig. 5. Blur estimation of the proposed VB algorithm (ARD prior) on 32 photos with motion blur.

5. CONCLUSIONS

Our goal was to provide better insight to image priors in blind deconvolution and understanding of superior performance of the family of ARD priors. We also show that the VB udpate procedure for the ARD prior is equivalent to the half-quadratic algorithm with the additional covariance term, which further emphasizes edges.

6. REFERENCES

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

[2] G.R. Ayers and J.C.Dainty, "Iterative blind deconvolution method and its application," *Optical Letters*, vol. 13, no. 7, pp. 547–549, July 1988.

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

[4] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T. Roweis, and William T. Freeman, "Removing camera shake from a single photograph," in *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, New York, NY, USA, 2006, pp. 787–794, ACM.

[5] R. Molina, J. Mateos, and A. K. Katsaggelos, "Blind deconvolution using a variational approach to parameter, image, and blur estimation," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 37153727, Dec. 2006.

[6] Jiaya Jia, "Single image motion deblurring using transparency," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR '07*, 17–22 June 2007, pp. 1–8.

[7] N. Joshi, R. Szeliski, and D. J. Kriegman, "PSF estimation using sharp edge prediction," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR 2008*, 23–28 June 2008, pp. 1–8.

[8] Qi Shan, Jiaya Jia, and Aseem Agarwala, "High-quality motion deblurring from a single image," in *SIGGRAPH '08: ACM SIGGRAPH 2008 papers*, New York, NY, USA, 2008, p. 110, ACM.

[9] Li Xu and Jiaya Jia, "Two-phase kernel estimation for robust motion deblurring," in *Proceedings of the 11th European conference on Computer vision: Part I*, Berlin, Heidelberg, 2010, ECCV'10, p. 157170, Springer-Verlag.

[10] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding blind deconvolution algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2354–2367, 2011.

[11] James Miskin and David J.C. MacKay, "Ensemble learning for blind image separation and deconvolution.," in *Advances in Independent Component Analysis*, M. Girolani, Ed., pp. 123–142. Springer-Verlag, 2000.

[12] N. P. Galatsanos, V. Z. Mesarovic, R. Molina, and A. K. Katsaggelos, "Hierarchical bayesian image restoration from partially known blurs," *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1784–1797, 2000.

[13] David Wipf and Srikantan Nagarajan, "A new view of automatic relevance determination," in *In NIPS 20*, 2008.

[14] S. Derin Babacan, Rafael Molina, Minh N. Do, and Aggelos K. Katsaggelos, "Bayesian blind deconvolution with general sparse image priors," in *ECCV*, Andrew W. Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, Eds. 2012, vol. 7577 of *Lecture Notes in Computer Science*, p. 341355, Springer.

[15] Dimitris Tzikas, Aristidis Likas, and Nikolas P. Galatsanos, "Variational bayesian sparse kernel-based blind image deconvolution with student's-t priors.," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 753–764, 2009.

[16] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing*, Springer Verlag, New York, 2002.

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

[18] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.

[19] A. Levin, Y. Weiss, F. Durand, and W.T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR '09*, 2009, pp. 1964–1971.