PARAMETER ESTIMATION IN BAYESIAN BLIND DECONVOLUTION WITH SUPER GAUSSIAN IMAGE PRIORS

Miguel Vega, Rafael Molina *

Universidad de Granada,
18071 Granada Spain,
e-mails: mvega@ugr.es, rms@decsai.ugr.es

Aggelos K. Katsaggelos

Department of Electrical Engineering
and Computer Science,
Northwestern University,
Evanston, IL 60208 USA,
e-mail: aggk@eecs.northwestern.edu

ABSTRACT

Super Gaussian (SG) distributions have proven to be very powerful prior models to induce sparsity in Bayesian Blind Deconvolution (BD) problems. Their conjugate based representations make them specially attractive when Variational Bayes (VB) inference is used since their variational parameters can be calculated in closed form with the sole knowledge of the energy function of the prior model. In this work we show how the introduction in the SG distribution of a global strength (not necessary scale) parameter can be used to improve the quality of the obtained restorations as well as to introduce additional information on the global weight of the prior. A model to estimate the new unknown parameter within the Bayesian framework is provided. Experimental results, on both synthetic and real images, demonstrate the effectiveness of the proposed approach.

Index Terms— Bayesian methods, image processing, image restoration, Super Gaussian, blind deconvolution.

1. INTRODUCTION

Blind image deconvolution is the problem of restoring an image $\mathbf{x}$ from its blurred and noisy version $\mathbf{y}$ when the blur $\mathbf{H}$ is unknown. Generally, the image $\mathbf{y}$ is modeled as

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

where $\mathbf{n}$ is the noise. Both $\mathbf{y}$, $\mathbf{x}$ and $\mathbf{n}$ are lexicographically arranged $N \times 1$ vectors, and $\mathbf{H}$ is an $N \times N$ matrix. In many cases the blur $\mathbf{H}$ is spatially-varying, but in this paper we assume that $\mathbf{H}$ is a spatially invariant two dimensional convolution operator of unknown nucleus $\mathbf{h}$. Since $\mathbf{h}$, $\mathbf{x}$ and $\mathbf{n}$ are unknown, the problem is highly ill-posed and there are infinitely many solutions for $\mathbf{x}$ and $\mathbf{h}$.

*This work was partially supported by the “Ministerio de Ciencia e Innovacion” under contract TIN2010-15137, and “CEI BioTic” at the University of Granada under contract CEI2014-PTIC11.

Blind image deconvolution is a widely investigated problem in signal/image processing and computer vision [1], and recently attracted much attention mostly geared towards removing camera shake [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. Fergus *et. al.* [2] employed the variational Bayesian approach of Miskin and Mackay [13] with a mixture-of-Gaussians (MoG) image prior for modeling natural image statistics. After the success of this approach, subsequent methods proposed new image and blur modeling schemes and highly efficient inference methods [5, 6, 7, 8].

In [14] a general VB blind deconvolution method, which can be used on a large family of sparsity promoting priors, was proposed. In this paper we extend the model in [14] by the introduction of a global scale parameter which is used to improve the quality of the restoration. The parameter is automatically estimated.

The rest of this paper is organized as follows. Section 2 describes our Bayesian modelling of the blind deconvolution problem. The Bayesian inference and the proposed algorithm are described in Section 3. In Section 4 we present the experimental results. Section 5 concludes the paper.

2. BAYESIAN MODELS

2.1. Degradation model

In this paper the deconvolution problem is formulated in the filter space [5, 8], by applying high-pass filters $\{\mathbf{F}_\gamma\}_{\gamma=1}^{L}$ (such as derivatives, wavelets, curvelets, etc.) to the blurred noisy image $\mathbf{y}$ to obtain L pseudo-observations

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

with $\mathbf{x}_\gamma = \mathbf{F}_\gamma\mathbf{x}$, and $\mathbf{n}_\gamma = \mathbf{F}_\gamma\mathbf{n}$. In (2) it is assumed that $\mathbf{H}$ and $\mathbf{F}_\gamma$ have the same set of e-vectors, so that they commute.

Assuming that $\mathbf{n}$ is uncorrelated white Gaussian noise of known inverse variance β, we define the distribution of the observed images $\mathbf{y}_\gamma$ as

$$\mathrm{p}(\mathbf{y}_\gamma|\mathbf{x}_\gamma, \mathbf{h}) = \mathcal{N}(\mathbf{y}_\gamma|\mathbf{H}\mathbf{x}_\gamma, \beta_\gamma^{-1}\mathbf{I})\,, \tag{3}$$

Table 1. Different choices for the penalty function

Label	$\rho(s)$	$\rho'(s)/\lvert s\rvert$
ℓ_p , $0 < p \leq 1$	$\frac{1}{p}\lvert s\rvert^p$	$\lvert s\rvert^{p-2}$
log	$\log(\epsilon + \lvert s\rvert)$	$(\epsilon + \lvert s\rvert)^{-1}\lvert s\rvert^{-1}$
bup, $p > 2$	$\frac{1}{2-p}(\epsilon + \lvert s\rvert)^{2-p}$	$(\epsilon + \lvert s\rvert)^{-p}$

where we use the approximation $\frac{1}{\beta}\mathbf{F}_\gamma\mathbf{F}_\gamma^{\mathrm{T}} \approx \frac{1}{\beta_\gamma}\mathbf{I}$.

2.2. Prior models

Let us now proceed to model our prior knowledge on the image. To promote sparsity on $\mathbf{x}_\gamma$ (derivatives, wavelets, curvelets, ...) we could use on each $x_\gamma(i)$ a Super Gaussian distribution,

$$\mathrm{p}\left(x_\gamma(i)|\alpha_\gamma\right) = Z_\gamma(\alpha_\gamma)\exp\left(-\alpha_\gamma\rho\left(x_\gamma(i)\right)\right) \,, \tag{4}$$

for $i = 1, \ldots, N$, where $Z_\gamma(\alpha_\gamma)$ is the partition function, and the parameter $\alpha_\gamma > 0$ globally regulates the prior strength. Table 1 shows some penalty functions, corresponding to SG distributions (see [14]). In Table 1, the acronym bup stands for the, so-called in [14], bottom-up approach, for a given value of p.

Notice that the global parameter α_γ in (4), which is not considered in [14] but introduced in this work, plays a very important role as we will see in the experimental section. Notice also that, it represents much more than a rescaling of $x_\gamma(i)$, see for instance the log prior in Table 1 where the introduction of the α_γ represents powering in the filter space.

For $\mathrm{p}\left(x_\gamma(i)|\alpha_\gamma\right)$ in (4) to be SG, $\rho(\cdot)$ has to be symmetric around zero, and the function $\rho(\sqrt{s})$ has to be increasing and concave for $s \in (0, \infty)$ [15]. This condition is equivalent to $\rho'(s)/s$ being decreasing on $(0, \infty)$. If this condition is satisfied, then ρ can be represented as (see [16, Ch. 12])

$$\rho\left(x_\gamma(i)\right) = \inf_{\eta_\gamma(i)>0} \frac{1}{2}\,\eta_\gamma(i)\,x_\gamma^2(i) \, - \rho^*\left(\frac{1}{2}\eta_\gamma(i)\right) \tag{5}$$

$$\Rightarrow \rho\left(x_\gamma(i)\right) \leq \frac{1}{2}\,\eta_\gamma(i)\,x_\gamma^2(i) - \rho^*\left(\frac{1}{2}\eta_\gamma(i)\right) \tag{6}$$

where inf denotes the infimum, $\rho^*\left(\cdot\right)$ is the concave conjugate of $\rho(\cdot)$ and $\boldsymbol{\eta}_\gamma = \{\eta_\gamma(i)\}_{i=1}^N$ are positive variational parameters. These parameters have an intuitive meaning and extreme importance in the deconvolution performance, as will be shown later. The relationship dual to (5) is given by [16]

$$\rho^*\left(\frac{1}{2}\eta_\gamma(i)\right) = \inf_{x_\gamma(i)} \frac{1}{2}\,\eta_\gamma(i)\,x_\gamma^2(i) - \rho\left(x_\gamma(i)\right) \,. \tag{7}$$

The quadratic bound for ρ in (6) allows us to bound the prior in (4) with a Gaussian form, specifically we can write

$$\begin{aligned}
&\mathrm{p}\left(x_\gamma(i)|\alpha_\gamma\right) \\
&\geq Z_\gamma(\alpha_\gamma)\exp[-\alpha_\gamma(\frac{1}{2}\eta_\gamma(i)x_\gamma^2(i) - \rho^*\left(\frac{\eta_\gamma(i))}{2}\right))] \\
&= Z_\gamma(\alpha_\gamma)\exp[-\alpha_\gamma L(x_\gamma(i), \eta_\gamma(i))]. \;\; \forall\eta_\gamma(i) > 0
\end{aligned} \tag{8}$$

Combining Eqs. (3) & (8) we obtain the global variational lower bound

$$\begin{aligned}
\mathrm{p}(\boldsymbol{\Theta}, \mathbf{Y}) &\geq \mathrm{p}(\mathbf{h})\prod_{\gamma=1}^{L}\left(\mathrm{p}(\alpha_\gamma)\mathrm{p}(\mathbf{y}_\gamma|\mathbf{x}_\gamma,\mathbf{h})\right. \\
&\quad \times \prod_i Z_\gamma(\alpha_\gamma)\exp[-\alpha_\gamma L(x_\gamma(i), \eta_\gamma(i))]) \\
&= \mathrm{P}(\boldsymbol{\Theta}, \mathbf{Y}, \boldsymbol{\eta}) \;\; \boldsymbol{\eta} > \mathbf{0}
\end{aligned} \tag{9}$$

where $\boldsymbol{\Theta} = \{\mathbf{h}, \alpha_1, \mathbf{x}_1, \ldots, \alpha_L, \mathbf{x}_L\}$, $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_L\}$ and $\boldsymbol{\eta} = \{\boldsymbol{\eta}_{\gamma_1}, \ldots, \boldsymbol{\eta}_{\gamma_L}\}$.

3. BAYESIAN INFERENCE

The inference scheme is based on the posterior distribution $\mathrm{p}(\boldsymbol{\Theta}|\mathbf{Y}) = \frac{\mathrm{p}(\boldsymbol{\Theta},\mathbf{Y})}{\mathrm{p}(\mathbf{Y})}$. Since the posterior distribution cannot be obtained in closed form, we base our estimation on VB inference [17], where the posterior $\mathrm{p}(\boldsymbol{\Theta}|\mathbf{Y})$ is approximated by the distribution $\mathrm{q}(\boldsymbol{\Theta}) = \prod_{\theta\in\boldsymbol{\Theta}}\mathrm{q}(\theta)$, with

$$\log\mathrm{q}(\theta) = \mathrm{E}\left[\log\mathrm{p}(\boldsymbol{\Theta}, \mathbf{Y})\right]_{\mathrm{q}(\boldsymbol{\Theta}\setminus\{\theta\})} + \mathrm{const}\,, \tag{10}$$

obtained by taking the expectation of the joint distribution with respect to all unknowns except the one of interest, $\boldsymbol{\Theta}\setminus\{\theta\}$. We derive next the estimates of the different variables.

Instead of using $\mathrm{p}(\boldsymbol{\Theta}, \mathbf{Y})$ we utilize its lower bound, which includes the variational parameter, and solve iteratively

$$\begin{aligned}
\log\mathrm{q}(\theta) &= \mathrm{E}\left[\log\mathrm{P}(\boldsymbol{\Theta}, \mathbf{Y}, \hat{\boldsymbol{\eta}})\right]_{\Theta\setminus\{\theta\}} + \mathrm{const}\,, \\
\hat{\boldsymbol{\eta}} &= \arg\max_{\boldsymbol{\eta}}\mathrm{E}\left[\log\mathrm{P}(\boldsymbol{\Theta}, \mathbf{Y}, \boldsymbol{\eta})\right]_{\mathrm{q}(\boldsymbol{\Theta})}
\end{aligned} \tag{11}$$

3.1. Estimation of Image and Blur

Assuming a flat prior $\mathrm{p}(\mathbf{h})$ we obtain for the blur

$$\log\mathrm{q}(\mathbf{h}) = \sum_\gamma \mathrm{E}\left[\log\mathrm{p}(\mathbf{y}_\gamma|\mathbf{x}_\gamma, \mathbf{h})\right]_{\mathrm{q}(\mathbf{x}_\gamma)} + \mathrm{const}\,. \tag{12}$$

Since the size of the blur is normally very small in comparison with the number of observations we treat the blur as a deterministic parameter which is only constrained to satisfy $\mathbf{h} \geq 0$, $\sum_i h(i) = 1$. Then from (12) we obtain

$$\begin{aligned}
\hat{\mathbf{h}} &= \arg\min_{\mathbf{h}}\sum_\gamma \mathrm{E}\left[\|\mathbf{y}_\gamma - \mathbf{H}\mathbf{x}_\gamma\|_2^2\right]_{\mathrm{q}(\mathbf{x}_\gamma)} \\
&= \arg\min_{\mathbf{h}}\; \mathbf{h}^T\mathbf{C}_{\mathbf{h}}^{-1}\mathbf{h} - 2\,\mathbf{h}^T\,\mathbf{b}_{\mathbf{h}}
\end{aligned} \tag{13}$$

subject to $\mathbf{h} \geq 0$, $\sum_i h(i) = 1$. For a kernel size of $M \times M$, $\mathbf{b_h}$ is $M^2 \times 1$, the size of $\mathbf{C_h}^{-1}$ is $M^2 \times M^2$ and we have

$$C_{\mathbf{h}}^{-1}(m,n) = \sum_{\gamma} \sum_{j=1}^{N} \mathrm{E}\left[x_\gamma(m+j)\right] \mathrm{E}\left[x_\gamma(n+j)\right] + C_{\mathbf{x}_\gamma}(m+j, n+j) \tag{14}$$

$$b_{\mathbf{h}}(m) = \sum_{\gamma} \sum_{j=1}^{N} \mathrm{E}\left[x_\gamma(m+j)\right]\, y_\gamma(j)\,. \tag{15}$$

The estimation of the blur in (13) is thus a simple quadratic problem which can be solved very efficiently.

For the filtered images we obtain

$$\log \mathrm{q}(\mathbf{x}_\gamma) = \log \mathrm{p}(\mathbf{y}_\gamma|\mathbf{x}_\gamma, \mathbf{h}) - \frac{1}{2}\alpha_\gamma \mathbf{x}_\gamma^T \mathrm{diag}\left(\boldsymbol{\eta}_\gamma\right) \mathbf{x}_\gamma + \mathrm{const} \tag{16}$$

which is a multivariate Gaussian with precision matrix

$$\mathbf{C}_{\mathbf{x}_\gamma}^{-1} = \beta_\gamma \mathbf{H}^T\mathbf{H} + \alpha_\gamma \mathrm{diag}\left(\boldsymbol{\eta}_\gamma\right), \tag{17}$$

and whose mean $\hat{\mathbf{x}}_\gamma$, which we take as the estimate of $\mathbf{x}_\gamma$, is obtained by solving the following linear system

$$\mathbf{C}_{\mathbf{x}_\gamma}^{-1}\hat{\mathbf{x}}_\gamma = \beta_\gamma \mathbf{H}^T \mathbf{y}_\gamma\,. \tag{18}$$

This system can be solved efficiently using *Conjugate Gradient* (CG) without finding $\mathbf{C}_{\mathbf{x}_\gamma}$ explicitly. However, this covariance matrix is needed in (14). Since its computation is extremely expensive, we apply in (14) the Jacobi approximation, and invert only the diagonal of $\mathbf{C}_{\mathbf{x}_\gamma}$.

3.2. Estimation of the Variational Parameter $\boldsymbol{\eta}_\gamma$

Using (10), we obtain for the variational parameters

$$\hat{\eta}_\gamma(i) = \arg\min_{\eta_\gamma(i)} \frac{1}{2}\eta_\gamma(i)\nu_\gamma^2(i) - \rho^*\left(\frac{\eta_\gamma(i)}{2}\right) \tag{19}$$

where $\nu_\gamma(i) = \sqrt{\mathrm{E}\left[x_\gamma^2(i)\right]}$.

Since

$$\rho^*(\frac{\hat{\eta}_\gamma(i)}{2}) = \min_x \frac{1}{2}\hat{\eta}_\gamma(i)x^2 - \rho(x) \tag{20}$$

whose minimum is achieved at $x = \nu_\gamma(i)$, we have, differentiating the right hand side of the above equation with respect to x,

$$\hat{\eta}_\gamma(i) = \rho'(\nu_\gamma(i))/\nu_\gamma(i) \tag{21}$$

3.3. Estimation of α_γ

From (11) we obtain the following distribution for α_γ

$$\log \mathrm{q}(\alpha_\gamma) = \mathrm{const} + \log \mathrm{p}(\alpha_\gamma) + \sum_{i=1}^{N} \log Z_\gamma(\alpha_\gamma)\exp[-\alpha_\gamma \rho(\nu_\gamma(i))] \tag{22}$$

We utilize the mode of this distribution as the α_γ estimate. A flat hyperprior ($\mathrm{p}(\alpha_\gamma) \propto \mathrm{const}$) results in the following equation for α_γ

$$\frac{\partial}{\partial \alpha_\gamma} \log Z_\gamma(\alpha_\gamma) = \frac{1}{N}\sum_{i=1}^{N} \rho(\nu_\gamma(i))\,. \tag{23}$$

The ℓ_p and bup penalty functions shown in Table 1 produce proper priors, for which the partition function can be evaluated. For the ℓ_p function $Z_\gamma(\alpha_\gamma) = \frac{1}{2\Gamma(1/p)}\alpha_\gamma^{\frac{1}{p}} p^{1-\frac{1}{p}}$, and for the bup one $Z_\gamma(\alpha_\gamma) = \frac{1}{2\Gamma(1/(2-p))}\alpha^{\frac{1}{2-p}}(2-p)^{\frac{1-p}{2-p}}$. However, the log penalty function produces an improper prior. We tackle this problem examining, for $\alpha_\gamma \neq 1$, the behavior of

$$Z_\gamma(\alpha_\gamma, K)^{-1} = \int_{-K}^{K} e^{-\alpha_\gamma \rho(s)} ds\,. \tag{24}$$

and keeping in $\partial \log Z_\gamma(\alpha_\gamma, K)/\partial \alpha_\gamma$ the term that depends on α_γ. This produces the estimate

$$\frac{1}{\alpha_\gamma - 1} = \frac{1}{N}\sum_{i=1}^{N} \rho(\nu_\gamma(i))\,. \tag{25}$$

3.4. The proposed Algorithm

Algorithm 1 Blind Deconvolution using Sparse Image Priors

Require: Degraded image $\mathbf{y}$, noise parameter β, choice for ρ function, and filters $\mathbf{F}_\gamma$.
repeat
  for $\gamma = 1$ **to** L **do**
    Initialization: Set $\hat{\mathbf{x}}_\gamma = \mathbf{y}_\gamma$, $\mathbf{C}_{\mathbf{x}_\gamma} = 0$.
    1. Compute $\nu_\gamma(i) = \sqrt{\mathrm{E}\left[x_\gamma^2(i)\right]}$, $\boldsymbol{\eta}_\gamma$ using (21), and α_γ solving (23).
    2. Estimate filtered image $\mathbf{x}_\gamma$ by solving (18).
  end for
  3. Estimate the blur kernel $\mathbf{h}$ using (13) .
  4. Approximate $C_{\mathbf{x}_\gamma}(i,i)$ with $1/\mathbf{C}_{\mathbf{x}_\gamma}^{-1}(i,i)$.
until convergence
5. Compute the final image estimate $\hat{\mathbf{x}}$ by solving

$$(\beta \mathbf{H}^T\mathbf{H} + \sum_\gamma \alpha_\gamma \mathbf{F}_\gamma^T \mathrm{diag}\left(\boldsymbol{\eta}_\gamma\right)\mathbf{F}_\gamma)\,\hat{\mathbf{x}} = \beta \mathbf{H}^T\,\mathbf{y}\,. \tag{26}$$

The proposed Algorithm 1 has two parts. First, Algorithm 1 iteratively alternates between the estimates of the filtered images $\mathbf{x}_\gamma$ and the blur $\mathbf{h}$. At each iteration, the method also estimates the parameters $\boldsymbol{\nu}_\gamma$, α_γ, and $\boldsymbol{\eta}_\gamma$, and approximates $C_{\mathbf{x}_\gamma}(i,i)$. In the estimation of $\mathbf{h}$, the pyramid coarse-to-fine approach suggested in [9] has been applied. In its second part, Algorithm 1 has to construct the image $\mathbf{x}$ from the $\mathbf{x}_\gamma$ filtered images. As this requires a non trivial integration of all $\mathbf{x}_\gamma$, we propose instead the estimate shown in (26), which still enforces sparsity in the filter domain through the use of $\boldsymbol{\eta}_\gamma$, and requires only one more CG application.

4. EXPERIMENTAL RESULTS

The effectiveness of the proposed method, for the ρ penalty functions ℓ_p, log, and bup, in Table 1, has been quantitatively assessed over the synthetic dataset in [8]. Results for the ℓ_p penalty function with $p = 0.8$, called $\ell_{0.8}$, and for the bup penalty with $p = 3$, called bu3, are reported. We use the Sum Square Distances (SSD) between the deconvolved and original images as figure of merit.

For the images in the dataset in [8], Figure 1 shows the cumulative SSD histograms of the blindly deconvolved images obtained when using the penalty functions in Table 1 and Levin *et al.*'s MAP_k approach [9], a MoG prior for kernel estimation. This method was shown in [9] to outperform the MAP based approaches of [5] and [6] in this dataset. Figure 1 also shows the SSD histograms for non-blind deconvolutions, and for the ratios between blind and non-blind SSD values. These SSD values are better than the reported in [14]. in Figure 1 it can be observed that the log prior gives the best performance, and that both ℓ_p and bup outperform MoG. The figures of merits for the ℓ_p prior for non-blind deconvolution are very goods. However, their blind SSD values, and their ratios are poorer. The results obtained with the bup prior are better for blind than for non-blind deconvolution, and this prior gives the best performance in ratio terms.

Figure 2 shows the results obtained with the different priors for one of the images in the dataset, and Figure 3 compares the proposed method with the methods in [6], and in [10], using their implementations, and parameter settings, on a real image.

The introduction of the α_γ parameter in the general sparse prior in (4) improves the performance of the method proposed in [14]. The improvement is very minor the log prior, for which the estimated parameter values are very close to one. For the ℓ_p penalty $\alpha_\gamma \approx 10^2$ has been obtained, and $\alpha_\gamma \approx 10^{-4}$ for bup.

5. CONCLUSIONS

In this paper we have introduced an additional parameter in the Super Gaussian (SG) prior modeling of the image in the filter space. We have shown how this parameter can be estimated using variational inference. Its effectiveness has been demonstrated experimentally on real and synthetic images.

REFERENCES

[1] T.E. Bishop, S.D. Babacan, B. Amizic, T. Chan, R. Molina, and A.K. Katsaggelos, "Blind image deconvolution: problem formulation and existing approaches," in *Blind image deconvolution: theory and applications*. CRC press, 2007.

[2] R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, and W.T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, pp. 787–794, July 2006.

[3] N. Joshi, C.L. Zitnick, R. Szeliski, and D.J. Kriegman, "Image deblurring and denoising using color priors," in *CVPR*, june 2009.

[4] J. Jia, "Single image motion deblurring using transparency," in *CVPR*, 2007.

[5] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graph. (SIGGRAPH)*, 2008.

[6] S. Cho and S. Lee, "Fast motion deblurring," *ACM Trans. Graph. (SIGGRAPH ASIA)*, vol. 28, no. 5, 2009.

[7] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," in *ECCV*, 2010.

[8] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *CVPR*, 2009, pp. 1964–1971.

[9] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," in *CVPR*, 2011, pp. 2657–2664.

[10] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *CVPR*, 2011.

[11] X. Zhou, F. Zhou, and X. Bai, "Blind deconvolution using a nondimensional gaussianity measure," in *ICIP*, 2013.

[12] C. Wang, Y. Yue, F. Dong, Y. Tao, X. Ma, G. Clapworthy, H. Lin, and X. Ye, "Nonedge-specific adaptive scheme for highly robust blind motion deblurring of natural imagess," *IEEE Trans. Image Proc.*, vol. 22, pp. 884–897, Mar. 2013.

[13] J. W. Miskin and D. J. C. MacKay, "Ensemble learning for blind image separation and deconvolution," in *Adv. in Independent Component Analysis*. 2000, Springer.

[14] S. D. Babacan, R. Molina, M. N. Do, and A. K. Katsaggelos, "Blind deconvolution with general sparse image priors," in *ECCV*, 2012.

[15] J. Palmer, K. Kreutz-Delgado, and S. Makeig, "Strong sub- and super-Gaussianity," in *LVA/ICA*, 2010, vol. 6365, pp. 303–310.

[16] R.T. Rockafellar, *Convex analysis*, Princeton University Press, 1996.

[17] A. C. Likas and N. P. Galatsanos, "A variational approach for Bayesian blind image deconvolution," *IEEE Trans. on Signal Proc.*, vol. 52, pp. 2222–2233, 2004.

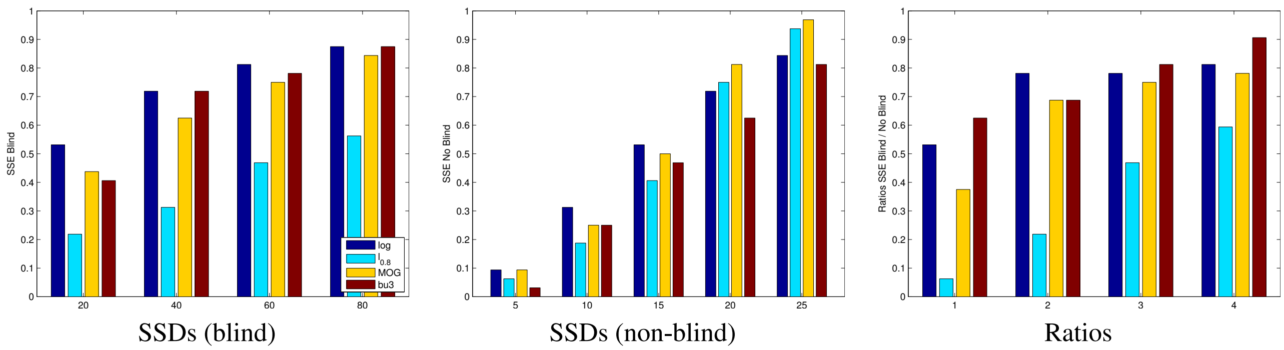

Fig. 1. Cumulative histograms of SSDs on the dataset of [8] for blind and non-blind deconvolution, and their ratios (right).

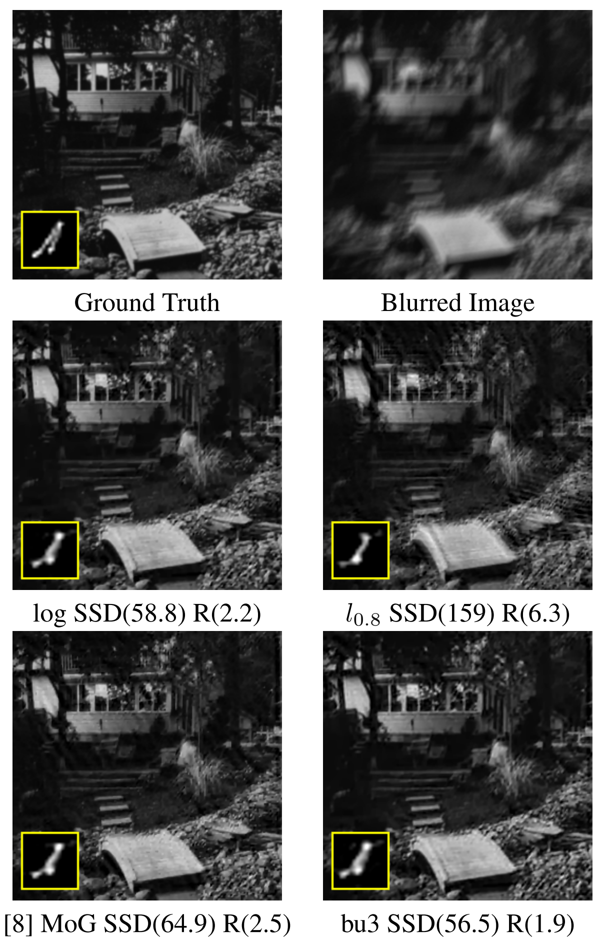

Fig. 2. Ground truth, blurred image and results obtained with our method using the different penalty functions and the method in [8], for image no. 2, and blur no. 2 of the synthetic image set in [8]. Estimated kernels are shown in insets.

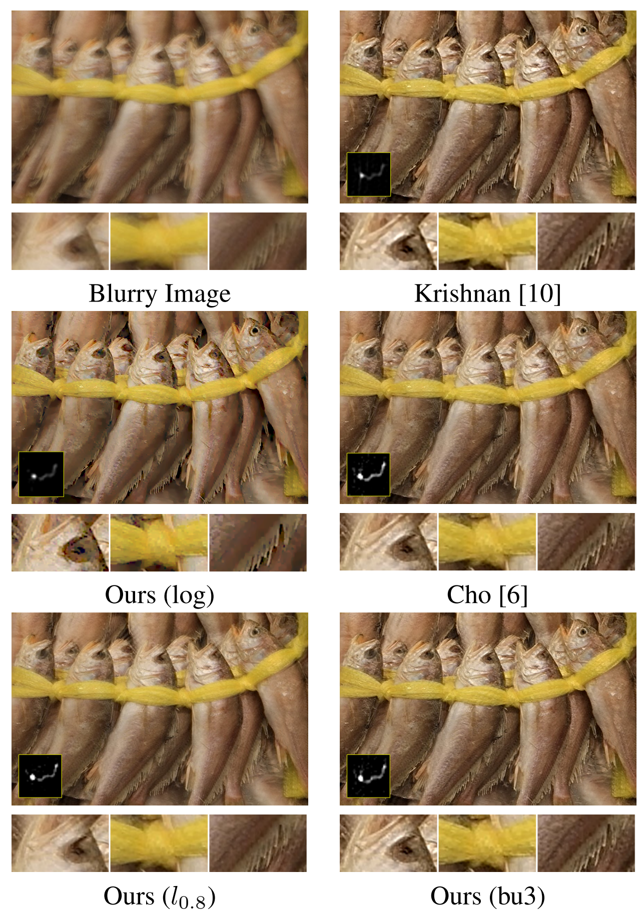

Fig. 3. Blind Deconvolution results on a real image. Estimated kernels are shown in insets.