

A Logarithmic Image Prior for Blind Deconvolution

Daniele Perrone¹ · Paolo Favaro¹

Received: 2 May 2015 / Accepted: 7 September 2015
© Springer Science+Business Media New York 2015

Abstract Blind Deconvolution consists in the estimation of a sharp image and a blur kernel from an observed blurry image. Because the blur model admits several solutions it is necessary to devise an image prior that favors the true blur kernel and sharp image. Many successful image priors enforce the sparsity of the sharp image gradients. Ideally the L_0 “norm” is the best choice for promoting sparsity, but because it is computationally intractable, some methods have used a logarithmic approximation. In this work we also study a logarithmic image prior. We show empirically how well the prior suits the blind deconvolution problem. Our analysis confirms experimentally the hypothesis that a prior should not necessarily model natural image statistics to correctly estimate the blur kernel. Furthermore, we show that a simple Maximum a Posteriori formulation is enough to achieve state of the art results. To minimize such formulation we devise two iterative minimization algorithms that cope with the non-convexity of the logarithmic prior: one obtained via the primal-dual approach and one via majorization-minimization.

Keywords Blind deconvolution · Majorization–minimization · Primal-dual · Image prior · Total variation · Logarithmic prior

Communicated by Thomas Brox.

✉ Daniele Perrone
daniele85@gmail.com
Paolo Favaro
favaro@iam.unibe.ch

¹ Department of Computer Science and Applied Mathematics, University of Bern, Neubrückstrasse 10, 3012 Bern, Switzerland

1 Introduction

Mobile phones are arguably the most ubiquitous imaging system in use nowadays, and hence have taken on a fundamental role in capturing our favorite memories. Unfortunately, mobile phone cameras come equipped with a small aperture that suffers from high noise levels when imaging low-light scenes. To compensate for low-light, exposure time can be increased. However, longer exposures make the mobile phone more sensitive to motion (of the camera or of the scene) and may result in visibly blurry pictures. Hence, in the past decade, the task of blur removal has become more and more pressing.

In this work we are interested in removing motion blur with the help of computational methods and we consider only a single blurry photograph as input. For this purpose, we use the following blur model

$$f = k * u + n \quad (1)$$

where k depends on the motion of the camera and is called the blur kernel (or point spread function), u is the sharp (or uncorrupted) image and n is the sensor noise. In this model the blur does not change across the image. This assumption does not hold in real scenes with depth variation and/or with general camera motions. Given both the blurry image f and the blur k , the estimation of the sharp image is a (*non blind deconvolution*) problem. When instead only the blurry image f is given, the problem of estimating both the sharp image and the blur is called *blind deconvolution*.

Solving blind deconvolution is a challenging task because it is a non-convex and ill-posed problem that requires the estimation of both u and k . For this reason it is often tackled with the use of regularization by means of prior knowledge on the distribution of sharp images (Babacan et al. 2012; Cho and Lee 2009; Fergus et al. 2006; Krishnan et al. 2011,

2013; Levin et al. 2011a; Shan et al. 2008; Wipf and Zhang 2014; Xu and Jia 2010; Xu et al. 2013). Several methods use also a prior on the blur kernel [(see for instance the recent works (Keuper et al. 2013; Kenig et al. 2010)]. A thorough analysis of convergence and further references can be found in Chaudhuri et al. (2014). In this paper, we consider only priors on the sharp image and use only the positivity and normalization constraints on the blur.

A common approach to choosing the image prior is to look for the one that best matches the statistics of natural images. This led to the use of Total Variation and its modifications, which brought some impressive results (Chan and Wong 1998; Shan et al. 2008). However, recent theoretical results have also shown that such priors are not suitable for the blind deconvolution problem. These priors prefer a blurry image rather than a sharp one (Levin et al. 2011b; Perrone and Favaro 2014).

Recently, Wipf and Zhang (2014) argued that a good image prior may not need to model realistic natural images. Indeed, they claim that an image obtained by removing gradients from a sharp image may be sufficient to estimate the correct blur kernel. Moreover, they show that their gradient sparsity principle is better than most natural image statistics in distinguishing a blurry image from a sharp one. Since one can then retrieve the sharp image given the estimated blur kernel, their conclusion is that an image prior that encourages strong sparsity in the gradients leads to a better performance in blind deconvolution.

Based on this principle, a natural choice is the L_0 “norm”. This is the ideal sparsity-promoting prior, but it also leads to an intractable combinatorial problem. It is therefore common to use some approximation. The logarithm of the gradient norm of an image has been the most successful approximation and has achieved state-of-the-art performance (Babacan et al. 2012; Wipf and Zhang 2014). Unluckily, these implementations employ elaborate formulations to deal with the non-convexity of the logarithmic prior.

In this work we introduce a parametric family of logarithmic priors and show that, despite being far from modeling natural image statistics, it particularly suits blind deconvolution. We present two simple algorithms that effectively minimize the logarithmic prior and achieve state of the art results. Finally, we study the prior behavior and show what makes it successful for blind deconvolution. The main difference with the work of Wipf and Zhang (2014) is that in this work we propose a more thorough empirical analysis of the logarithmic prior and the algorithms that we propose have a simpler form.

2 Prior Work

In the past decade, several high-performing blind deconvolution schemes using Bayesian principles have been proposed

(Babacan et al. 2012; Cho and Lee 2009; Fergus et al. 2006; Krishnan et al. 2011, 2013; Levin et al. 2011a; Shan et al. 2008; Wipf and Zhang 2014; Xu and Jia 2010; Xu et al. 2013). The first step in the Bayesian framework is to devise a statistical distribution for both the gradients of the sharp image and the measurement noise or the model error. This joint distribution is used to pose a *maximum a posteriori* (MAP) problem

$$p(u, k|f) \propto p(f|u, k)p(u)p(k), \quad (2)$$

where $p(f|u, k)$ is a generative model of the noise, $p(u)$ is a prior of the sharp image and $p(k)$ is a prior of the blur. Commonly used sharp image priors approximate the heavy-tail distribution of natural image gradients (Srivastava et al. 2003) via a sparsity-inducing norm of the gradients of u . The L_2 norm of the gradients (isotropic total variation), or the L_1 norm of the derivatives (anisotropic total variation) are classical choices (Chan and Wong 1998). In contrast to other sparsity-inducing norms, total variation (TV) (Rudin et al. 1992) has the desirable property of being convex. However, it also introduces a loss of contrast in the recovered sharp image (Perrone and Favaro 2014; Strong and Chan 2003). Other methods use heuristics to encourage sharp gradients (Cho and Lee 2009; Xu and Jia 2010; Shan et al. 2008), or some reweighing strategy of the norm of the gradients (Krishnan et al. 2011, 2013). The latter methods aim at approximating the L_0 “pseudo-norm” of the gradients, as proposed also in Xu et al. (2013). In this paper we also encourage sparsity in the gradients. However, we use the logarithm of TV at each pixel, which yields a simple energy term while providing a good approximation to the number of nonzero gradients. Indeed, this prior has already demonstrated promising results in blind deconvolution (Babacan et al. 2012; Wipf and Zhang 2013) and denoising (Ochs et al. 2014).

The MAP estimators are usually discredited to be theoretically less convenient than the conditional mean (CM) estimator (Burger and Lucka 2014). In fact, the CM estimator is the Bayes estimator for the mean square error, while the MAP estimator is only asymptotically the Bayes estimator for the uniform cost function. Nonetheless, in Burger and Lucka (2014) they show theoretical and experimental results that rehabilitate the MAP estimator and justify its successful use in different restoration problems.

For the blind deconvolution problem the MAP formulation has received further criticisms. In fact, in Levin et al. (2011b) and Perrone and Favaro (2014) it is shown that a large class of commonly used image priors favors the blurry image instead of the sharp one.

Because of such limitation, in Levin et al. (2011a) it is suggested to marginalize over all possible sharp images u and thus to solve the reduced problem

$$\begin{aligned} \max_k p(k|f) &= \min_k -\log p(k|f) \\ &= \min_k -\log \int_u p(u, f|k) p(k) du. \end{aligned} \tag{3}$$

In general, the integral in problem (3) is intractable. Therefore, typically one looks for an approximate solution. A common approach is to minimize an upper bound of $-\log p(k|f)$ using a variational Bayes strategy (Babacan et al. 2012; Ferguson et al. 2006; Levin et al. 2011a; Wipf and Zhang 2013). This class of methods has achieved performances comparable to the best methods that directly solve the MAP problem (2).

Despite the apparent performance of the variational Bayes strategy, Wipf and Zhang (2013) show that methods that solve problem (3) are equivalent to a MAP strategy as in problem (2). They experimentally show that with an L_p norm with $p \ll 1$, MAP approaches are able to favor the right sharp solution. They also argue that a variational Bayes approach should be preferred because it is more robust when minimizing a highly non-convex function. Their conclusions are however in contrast with several MAP approaches that have demonstrated effective results in various non-convex problems (Stekalovskiy and Cremers 2014; Möllenhoff et al. 2014a, b; Ochs et al. 2014). The conclusions given in Wipf and Zhang (2013) suggest that minimizing a cost functional as in (2) is not limited per se, as long as one finds a minimization strategy that carefully avoids its local minima.

In this paper, we extend the initial work in Perrone et al. (2014) where we proposed two MAP strategies to minimize a functional based on a logarithmic non-convex prior by proposing an analysis of the logarithmic prior, by studying the distribution of the parameters and by adding additional experiments that show the properties the logarithmic prior and of the two algorithms.

3 A Logarithmic Image Prior

In this section we introduce our image prior. From a Bayesian perspective, natural images can be described as having a sparse collection of gradients (Srivastava et al. 2003). Hence, one could employ sparsity-inducing priors of the image gradients. However, another point of view is that blurring is the average of shifted and scaled replicas of the same image gradients. The likelihood that such replicas combine to cancel each other is statistically insignificant. Therefore, this averaging is more likely to multiply the number of gradients by the number of nonzero elements in the blur. Thus, a different perspective is that, in the context of deblurring, the role of an image prior is to favor solutions that have as few gradients as possible regardless of their magnitude. Both points of view lead to the same principle, *i.e.*, one should choose as prior

$$\text{Number of non zero elements of } (|\nabla u|) \doteq \|\nabla u\|_0 \tag{4}$$

where $\|\cdot\|_0$ denotes the L_0 “norm” (the Hamming distance to zero) and ∇u is the 2-D gradient of u . Unfortunately, optimization with this prior is very challenging and, typically, smoother alternatives such as L_p norms $\|\nabla u\|_p^p$, with $0 < p < 1$, are used. In this work we also consider a prior with a similar behavior and simple form.

Let us consider the discrete setting. In the 2D discrete case, we have images with $N \times M$ pixels. The (i, j) -th entry of the blurry image u will be denoted by $u_{i,j}$. We consider four possible first order (discrete) derivatives of u according to whether forward or backward differences are used:

$$\nabla_{FF}u \doteq [u_{i+1,j} - u_{i,j} \quad u_{i,j+1} - u_{i,j}]^T \tag{5}$$

$$\nabla_{FB}u \doteq [u_{i+1,j} - u_{i,j} \quad u_{i,j} - u_{i,j-1}]^T \tag{6}$$

$$\nabla_{BF}u \doteq [u_{i,j} - u_{i-1,j} \quad u_{i,j+1} - u_{i,j}]^T \tag{7}$$

$$\nabla_{BB}u \doteq [u_{i,j} - u_{i-1,j} \quad u_{i,j} - u_{i,j-1}]^T. \tag{8}$$

As image prior we propose using the following logarithmic prior¹

$$\begin{aligned} \log \|\nabla u\|_{2,\epsilon}^p &\doteq \sum_{i=1}^N \sum_{j=1}^M \frac{1}{4} \sum_{D \in \mathcal{D}} \log \|\nabla_D u_{i,j}\|_{2,\epsilon}^p \\ &= \frac{p}{2} \sum_{i=1,j=1}^{N,M} \frac{1}{4} \sum_{D \in \mathcal{D}} \log \|\nabla_D u_{i,j}\|_{2,\epsilon}^2, \end{aligned} \tag{9}$$

with $p > 0$, $\mathcal{D} = \{FF, FB, BF, BB\}$, and where

$$\|\nabla u_{i,j}\|_{2,\epsilon}^2 \doteq (u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2 + \epsilon^2 \tag{10}$$

for $\epsilon > 0$ so that the argument of the logarithm is never 0. Since the following analysis and discussion can be applied to each gradient discretization independently, in the remainder of this work we will use the following simplified notation

$$\begin{aligned} \log \|\nabla u\|_{2,\epsilon}^p &\doteq \sum_{i=1}^N \sum_{j=1}^M \log \|\nabla u_{i,j}\|_{2,\epsilon}^p \\ &= \frac{p}{2} \sum_{i=1,j=1}^{N,M} \log \|\nabla u_{i,j}\|_{2,\epsilon}^2. \end{aligned} \tag{11}$$

In Eq. (11) the parameter ϵ leads to a lower bound for this prior equal to $MNp \log \epsilon$. We can formulate our blind

¹ Although we choose an L_2 norm, any L_q norm could be used. However, we have found experimentally that for a wide set of values in q this makes little difference in the final performance.

deconvolution problem using the logarithmic prior in Eq. (11) as

$$\begin{aligned}
 u, k &= \arg \min_{u,k} \lambda \|k * u - f\|_2^2 + \log \|\nabla u\|_{2,\epsilon}^p \\
 \text{s.t. } &k \succcurlyeq 0, \quad \|k\|_1 = 1.
 \end{aligned} \tag{12}$$

Notice how the role of ϵ is fundamental. If $\epsilon = 0$ then the optimal solution will always be $u = 0$ for any λ .

Remark The following limit shows how the log prior approximates the desired L_0 “norm”

$$\lim_{\epsilon \rightarrow 0} \frac{p}{2} + \frac{1}{\log(1/\epsilon^2)} \log \|\nabla u\|_{2,\epsilon^2}^p = \frac{p}{2} \|\nabla u\|_0. \tag{13}$$

Now, assume that $0 < \epsilon \leq 1$ and we substitute λ in problem (12) with $-\lambda p \log \epsilon^2$. Then, in the limit for $\epsilon \rightarrow 0$ we are solving

$$\begin{aligned}
 u, k &= \arg \min_{u,k} \lambda \|k * u - f\|_2^2 + \|\nabla u\|_0 \\
 \text{s.t. } &k \succcurlyeq 0, \quad \|k\|_1 = 1.
 \end{aligned} \tag{14}$$

In the following sections we present two different strategies to minimize problem (12), and we study the prior in Eq. (11) to understand how it relates to other commonly used priors and why it is a good choice for blind deconvolution.

4 Algorithm

To solve problem (12) we use the alternating minimization scheme

<pre> initialize $k^1 = k_1$ iterate $t = 1, \dots, T$ $u^{t+1} = \arg \min_u \lambda \ k^t * u - f\ _2^2 + \log \ \nabla u\ _{2,\epsilon}^p$ $k^{t+1} = \arg \min_k \ k * u^{t+1} - f\ _2^2$ s.t. $k \succcurlyeq 0, \quad \ k\ _1 = 1.$ </pre>	(15)
---	------

While the iteration in the blur k entails solving a convex problem, and we solve it as in Chan and Wong (1998), the minimization in the update of the sharp image u is non convex and requires more attention. To this purpose we introduce two solvers: one based on a primal-dual approach and another on majorization-minimization.

4.1 A Primal-Dual Solver

Recall the deblurring problem (given the blur k^t) in Algorithm (15); here we rewrite it as

$$u = \arg \min_u \sum_{i=1, j=1}^{N,M} ((k^t * u)_{i,j} - f_{i,j})^2 + \frac{1}{\mu} \log \|\nabla u\|_{2,\epsilon}^2, \tag{16}$$

where $\mu = 2\lambda/p$. By using the primal-dual approach of Chambolle and Pock (2011) we obtain the following min-max problem

$$\begin{aligned}
 u &= \arg \min_u \max_{z_1, z_2} \langle k^t * u, z_1 \rangle - F_1^*(z_1) \\
 &\quad + \langle \nabla u, z_2 \rangle - F_2^*(z_2)
 \end{aligned} \tag{17}$$

where F_1^* and F_2^* are conjugate functions of F_1 and F_2 respectively, and we have defined

$$\begin{aligned}
 F_1(x) &\doteq \sum_{i=1, j=1}^{N,M} (x_{i,j} - f_{i,j})^2, \\
 F_2(\xi) &\doteq \frac{1}{\mu} \sum_{i=1, j=1}^{N,M} \log \|\xi_{i,j}\|_{2,\epsilon}^2.
 \end{aligned} \tag{18}$$

The conjugate functions can be computed via the Legendre-Fenchel (LF) transform (Rockafellar 1970) and are convex by construction. Thus problem (17) is an approximation in all variables z_1, z_2 and u of the original problem (16). This formulation has been also proposed for solving the Mumford-Shah problem (Stekalovskiy and Cremers 2014) and for solving problems with L_p norms (with $0 < p < 1$) (Möhlenhoff et al. 2014a). Notice that the convex approximation provided by the primal-dual formulation may not lead to one of the minima of the original non convex cost.

Our general primal-dual algorithm to solve problem (17) is

<pre> $z_1^{n+1} = \text{prox}_{\sigma F_1^*}(z_1^n + \sigma k^t * \bar{u}^n)$ $z_2^{n+1} = \text{prox}_{\sigma F_2^*}(z_2^n + \sigma \nabla \bar{u}^n)$ $u^{n+1} = u^n - \tau (k_-^t * z_1^{n+1} + \nabla \cdot z_2^{n+1})$ $\bar{u}^{n+1} = u^{n+1} + \theta(u^{n+1} - u^n)$ </pre>	(19)
---	------

where k_-^t denotes the mirrored blur kernel k^t (along both axes), n is the iteration index, $\theta \in (0, 1]$ and $\tau\sigma\|K\|^2 < 1$, with $\tau, \sigma > 0$, where K is the matrix operator implementing both the blur k and the finite difference operator ∇ . Two of the 4 iterations in the above algorithm are defined based on the proximity operator. The proximity operator $\text{prox}_{\sigma F_1^*}$ is computed via

$$\begin{aligned} \text{prox}_{\sigma F_1^*}(z) &= z - \sigma \text{prox}_{F_1/\sigma}(z/\sigma) \\ &= z - \sigma \arg \min_x \frac{1}{2} \left\| \frac{z}{\sigma} - x \right\|_2^2 + \sigma F_1(x) \quad (20) \\ &= \frac{1}{\sigma + 1} (z - \sigma f). \end{aligned}$$

The proximity operator $\text{prox}_{\sigma F_2^*}$ is instead computed via

$$\text{prox}_{\sigma F_2^*}(z) = z - \sigma \arg \min_x \frac{1}{2} \left\| \frac{z}{\sigma} - x \right\|_2^2 + \sigma F_2(x). \quad (21)$$

In Eq. (21) we use Moreau’s Identity (Rockafellar 1970) to express the proximity operator of F_2^* in terms of the proximity operator of F_2 . This strategy has been used successfully also for other priors and problems (Möllenhoff et al. 2014a; Strekalovskiy and Cremers 2014). While this algorithm was originally studied for convex problems (Chambolle and Pock 2011), recently convergence has been studied also for non-convex problems (Möllenhoff et al. 2014b).

Since the minimization problem is separable, let us consider the solution obtained for only one element $x_{i,j}$ and $z_{i,j}$ of the variables x and z respectively. With an abuse of notation, instead of the element-wise cumbersome notation $x_{i,j}$ and $z_{i,j}$ we simply refer to x and z in the next equations. We use the representation $x \doteq \rho w$, where $\rho \geq 0$ and $\|w\|_2 = \|z\|_2/\sigma$. Then, let $\xi = z/\sigma$ and we have

$$\begin{aligned} \arg \min_x \frac{1}{2} \|\xi - x\|_2^2 + \sigma F_2(x) \\ = \arg \min_{\rho, w} \frac{\rho^2}{2} \left\| \frac{\xi}{\rho} - w \right\|_2^2 + \frac{\sigma}{\mu} \log \left(\rho^2 \frac{\|z\|_2^2}{\sigma^2} + \epsilon^2 \right). \end{aligned} \quad (22)$$

Notice that the logarithmic term now depends only on ρ . Hence, we can first solve the minimization problem with respect to w . By simplifying the least squares term we obtain

$$\begin{aligned} \arg \min_{w, \|w\| = \frac{\|z\|}{\sigma}} \frac{\rho^2}{2} \left\| \frac{\xi}{\rho} - w \right\|_2^2 \\ = \arg \min_{w, \|w\| = \frac{\|z\|}{\sigma}} \left\| \frac{\xi}{\rho} \right\|_2^2 + \|w\|_2^2 - 2\langle \xi/\rho, w \rangle \\ = \arg \min_{w, \|w\| = \frac{\|z\|}{\sigma}} \frac{\|z\|_2^2}{\sigma^2} - 2\langle \xi/\rho, w \rangle \quad (23) \\ = \arg \max_{w, \|w\| = \frac{\|z\|}{\sigma}} \langle \xi, w \rangle \end{aligned}$$

which immediately yields $w = \frac{\|z\|_2}{\sigma \|\xi\|_2} \xi = z/\sigma$. By substituting the expression of w back into Eq. (22) and by using $\xi = z/\sigma$ we finally have

Table 1 The proposed primal-dual algorithm

initialize
$h^1 = h_1$
iterate $t = 1, \dots, T$
iterate $n = 1, \dots, N_0$
$z_1^{n+1} = \frac{1}{\sigma+1} (z_1^n + \sigma (k^t * \tilde{u}^n - f))$
$z_2^{n+1} = \left(1 - \mathcal{H} \left(\frac{z_2^n + \sigma \nabla \tilde{u}^n}{\sigma}, \mu, \epsilon, \sigma \right) \right) (z_2^n + \sigma \nabla \tilde{u}^n)$
$\tilde{u}^{n+1} = \tilde{u}^n - \tau \left(k^t * z_1^{n+1} + \nabla \cdot z_2^{n+1} \right)$
$\tilde{u}^{n+1} = \tilde{u}^{n+1} + \theta (\tilde{u}^{n+1} - \tilde{u}^n)$
end iterate n
$u^{t+1} = \tilde{u}^{N_0+1}$
$h^{t+1} = \arg \min_k \ k * u^{t+1} - f\ _2^2$
s.t. $k \succeq 0, \ k\ _1 = 1$
end iterate t

$$\begin{aligned} \arg \min_x \frac{1}{2} \|\xi - x\|_2^2 + \sigma F_2(x) \\ = \xi \cdot \arg \min_{\rho} \frac{1}{2} \left\| \xi - \rho \frac{z}{\sigma} \right\|_2^2 + \frac{\sigma}{\mu} \log \left(\rho^2 \frac{\|z\|_2^2}{\sigma^2} + \epsilon^2 \right) \\ = \xi \cdot \arg \min_{\rho} \frac{\mu}{2\sigma} (1 - \rho)^2 \|\xi\|_2^2 + \log \left(\rho^2 \|\xi\|_2^2 + \epsilon^2 \right). \end{aligned} \quad (24)$$

We can define \mathcal{H} as the solution of the 1D problem

$$\begin{aligned} \mathcal{H}(\xi, \mu, \epsilon, \sigma) = \arg \min_{\rho} \frac{\mu}{2\sigma} (\rho - 1)^2 \|\xi\|_2^2 \\ + \log(\rho^2 \|\xi\|_2^2 + \epsilon^2) \end{aligned} \quad (25)$$

and build it into a lookup table.² The proximity operator $\text{prox}_{\sigma F_2^*}$ can then be written as

$$\text{prox}_{\sigma F_2^*}(z) = \left(1 - \mathcal{H} \left(\frac{z}{\sigma}, \mu, \epsilon, \sigma \right) \right) z. \quad (26)$$

The final algorithm is summarized in Table 1. A similar approach was proposed for the minimization of L_p norms with $0 < p < 1$ by Möllenhoff et al. (2014a). Notice how several operations are parallelizable, thus leading to a very efficient implementation.

4.2 A Majorization–Minimization Approach

As a more accurate alternative to the primal-dual algorithm, one could use a majorization–minimization (MM) approach (Hunter and Lange 2004), in a similar manner as proposed by Candes et al. (2008). In the MM approach one defines an

² Notice that the 1D problem leads to a third order polynomial equation for which closed-form solutions are known.

upper bound functional $\psi(u|u^t)$ given the current estimate u^t at time t . This upper bound must satisfy the following properties

$$\begin{aligned} \psi(u|u^t) &\geq \sum_{i=1}^N \sum_{j=1}^M \log \|\nabla u_{i,j}\|_{2,\epsilon}^p \\ \psi(u^t|u^t) &= \sum_{i=1}^N \sum_{j=1}^M \log \|\nabla u_{i,j}^t\|_{2,\epsilon}^p. \end{aligned} \tag{27}$$

Then, one can apply the following iterative scheme

$$u^{t+1} = \arg \min_u \sum_{i=1, j=1}^{N, M} \lambda \left((k * u)_{i,j} - f_{i,j} \right)^2 + \psi(u|u^t) \tag{28}$$

and provably reach a local minimum of the original function. To define the upper bound, we consider using the Taylor expansion of the logarithm around the t -th estimate of $\|\nabla u\|_{2,\epsilon}^p$ up to the first term

$$\begin{aligned} \psi(u|u^t) &= \sum_{i=1, j=1}^{N, M} \log \|\nabla u_{i,j}^t\|_{2,\epsilon}^p \\ &\quad + \frac{\|\nabla u_{i,j}\|_{2,\epsilon}^p - \|\nabla u_{i,j}^t\|_{2,\epsilon}^p}{\|\nabla u_{i,j}^t\|_{2,\epsilon}^p}. \end{aligned} \tag{29}$$

The properties (27) hold because of the concavity of the logarithm function. Finally, by plugging ψ in Eq. (28) we obtain the following update

$$\begin{aligned} u^{t+1} &= \arg \min_u \sum_{i=1, j=1}^{N, M} \lambda \left((k * u)_{i,j} - f_{i,j} \right)^2 \\ &\quad + \log \|\nabla u_{i,j}\|_{2,\epsilon}^p + \frac{\|\nabla u_{i,j}\|_{2,\epsilon}^p - \|\nabla u_{i,j}^t\|_{2,\epsilon}^p}{\|\nabla u_{i,j}^t\|_{2,\epsilon}^p} \\ &= \arg \min_u \sum_{i=1, j=1}^{N, M} \lambda \left((k * u)_{i,j} - f_{i,j} \right)^2 + \frac{\|\nabla u_{i,j}\|_{2,\epsilon}^p}{\|\nabla u_{i,j}^t\|_{2,\epsilon}^p}. \end{aligned} \tag{30}$$

so that the majorization–minimization algorithm can be summarized in Table 2. Notice the similarity with reweighed least squares algorithms when $p = 2$.

5 Analysis of the Logarithmic Prior

From a Bayesian perspective the logarithm prior in Eq. (11) with $p = \frac{2}{\pi\epsilon}$ is equivalent to the assumption that the magnitude of the gradient of u is independent and identically distributed according to a Cauchy distribution. In fact, if we define the prior as

Table 2 The proposed majorization–minimization algorithm

initialize
$h^1 = h_1$
iterate $t = 1, \dots, T$
$u^{t+1} = \arg \min_u \sum_{i=1, j=1}^{N, M} \lambda \left((k^t * u)_{i,j} - f_{i,j} \right)^2 + \frac{\ \nabla u_{i,j}\ _{2,\epsilon}^p}{\ \nabla u_{i,j}^t\ _{2,\epsilon}^p}$
$h^{t+1} = \arg \min_k \ k * u^{t+1} - f\ _2^2$
s.t. $k \succcurlyeq 0, \ k\ _1 = 1$
end iterate t

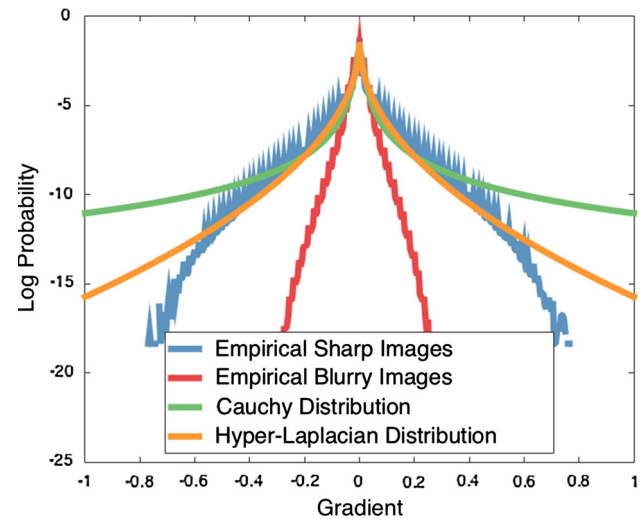


Fig. 1 Log probability of natural and blurry images compared to Cauchy and Hyper-Laplacian distributions (Color figure online)

$$\frac{1}{\pi\epsilon} \prod_{i=1, j=1}^{N, M} \left(\frac{\epsilon^2}{\|\nabla u_{i,j}\|_2^2 + \epsilon^2} \right), \tag{31}$$

the minimization in Eq. (12) is equivalent to a maximum a posteriori (MAP) estimation in which Eq. (31) is used as image prior. The Cauchy distribution does not fit particularly well natural image statistics, as we show in Fig. 1. Why should it be a good choice for blind deconvolution?

5.1 Image Statistics or Blur “Reconstructability”?

Following the arguments made in Wipf and Zhang (2014) we argue that in blind deconvolution an image prior does not have to necessarily model natural image statistics. Rather, it could model another family of images as long as it allows the estimation of the correct blur kernel k . In support to this thesis we show that blur estimation is sufficiently accurate for a wide range of “unnatural” images, and hence we can use them in practice. Our analysis focuses on the image family modeled by the logarithmic prior, but similar results are obtained by using Total Variation, which is another typical

cartooning prior successfully used in blind/nonblind deconvolution.

To substantiate this claim, we perform an empirical evaluation of the logarithmic prior. The family of “unnatural” images is generated by solving the denoising problem

$$u_\lambda = \arg \min_x \frac{\lambda}{2} \|x - u\|_2^2 + \log \|\nabla x\|_{2,\epsilon}^p, \tag{32}$$

where u is a ground truth sharp image and the image u_λ is its approximation via the logarithmic prior (for a given λ). We consider several degrees of regularization by varying λ . As sharp images we use the dataset introduced in Sun et al. (2013) composed by 80 images.

After computing the image u_λ , which is essentially a “cartooned” version of the original image u , we estimate the blur k_λ by solving the problem

$$\begin{aligned} \min_k \quad & \|k * u_\lambda - f\|_2^2 \\ \text{s.t.} \quad & k \succeq 0, \quad \|k\|_1 = 1 \end{aligned} \tag{33}$$

where the blurry image f is generated by convolving one of the 8 different blurs in Sun et al. (2013) with the ground truth data u . Thus, the model $k * u_\lambda$, $\lambda < \infty$, can never match f exactly. Because of the dependency on λ we denote the optimal blur k_λ . Finally, to evaluate the accuracy of the estimated blur, we compute the error between k_λ and the ground truth blur k by using the L_2 norm of their difference. When λ is very high, the cartooning disappears and u_λ approaches u . In this case the recovered blur k_λ also approaches the true blur k . It is more interesting to observe k_λ when λ becomes small, as this corresponds to working with a family of less “natural” images u_λ .

In Fig. 2 we show the average blur error for different values of λ . When λ is decreased u_λ becomes more “cartooned”, until, at $\lambda = 0.1$, it becomes almost constant (see Fig. 5). This translates in an increasing error for the blur k_λ as λ becomes smaller. In the second column of Fig. 4 one can see how even a cartooned version of the sharp image can yield a blur with a relatively small error compared to the ground truth blur. The other columns in Figs. 4 and 5 show how images that look different still yield similar blur estimations.

As typically done in the literature we also perform a final non-blind deconvolution step using the estimated blurs. This experiment allows establishing a connection between the error in the blur reconstruction and the error in the final sharp image. For this evaluation we use the SSD ratio proposed in Levin et al. (2011b). The ratio can be computed by

$$r = \frac{\sum_{i=1, j=1}^{N, M} (u_{i,j}^{k_\lambda} - u_{i,j}^g)^2}{\sum_{i=1, j=1}^{N, M} (u_{i,j}^{k_g} - u_{i,j}^g)^2} \tag{34}$$

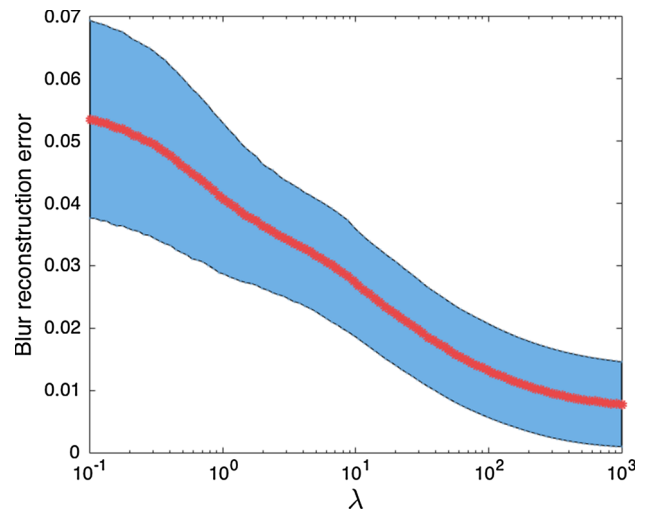


Fig. 2 Plot of the average error $\|k_\lambda - k\|_2^2$ obtained over the dataset from Sun et al. (2013) for different values of λ . The blue region denotes the area around the average bounded by one standard deviation (Color figure online)

where u^g is the ground truth sharp image, u^{k_g} is the image obtained by solving a non-blind deconvolution problem with the ground truth blur, and u^{k_λ} is the image obtained by solving a non-blind deconvolution problem with the blur k_λ . This metric is commonly used to evaluate blind deconvolution algorithms, because it takes into account the intrinsic difficulty of each blur. In our experiment we evaluate the blurs estimated with images obtained with different values of λ using the non-blind deconvolution proposed in Zoran and Weiss (2011).

In Fig. 3 we show the percentage of images with $r \leq 5$ for each value of λ . We choose to consider error ratios smaller than 5 because it has been shown that reconstructed images with such error are still visually pleasant (Michaelli and Irani 2014). Notice that for values of $\lambda > 60$, 100% of the images have error ratio smaller than 5 (Fig. 4). This range includes also values for which a cartooned version of the sharp image is generated. Indeed, from Fig. 5 one can see that u_λ is evidently cartooned for $\|k_\lambda - k\| \leq 0.01$. In Fig. 2 such blur error corresponds to $\lambda < 200$, which largely overlaps with the range $\lambda > 60$ obtained from Fig. 3.

The previous set of experiments shows that a large set of reconstructed images allows the correct estimation of the PSF. It also shows how this set includes the ground truth image and images where details (which do not affect the blur reconstruction) are removed. In conclusion, our experiments demonstrate that it is limiting to look only for images that resemble as much as possible the sharp ground truth. Finally, we show how the logarithmic prior is able to generate different images that allow a correct reconstruction of the blur.

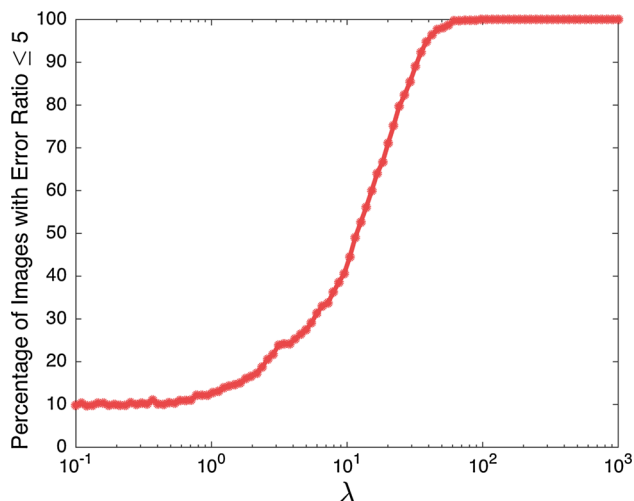


Fig. 3 Percentage of images with error ratio smaller or equal than 5 for different values of λ

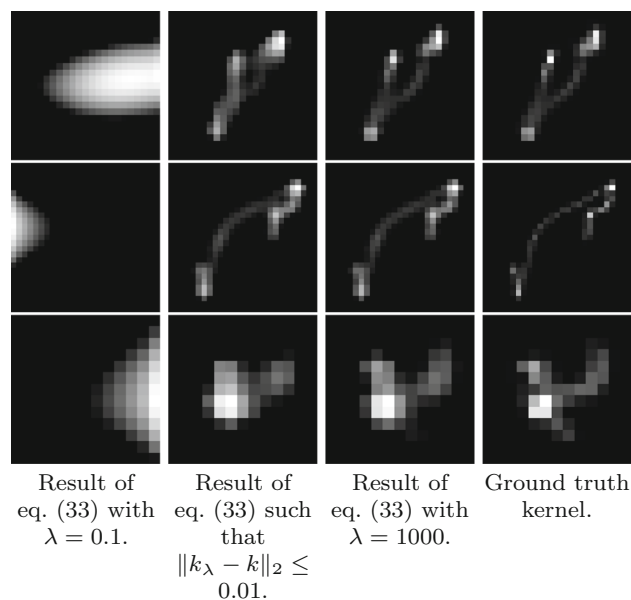


Fig. 4 Example of kernels obtained by solving Eq. (33) using different u_λ

5.2 Favoring Cartooned Images

Another fundamental aspect of the image prior is whether it favors the blurry input or some “cartooned” version of the true sharp image. To understand how the logarithmic prior behaves, we compare it with the total variation (TV) prior and the Hyper-Laplacian prior. We consider the difference $d(\lambda) = \psi(f) - \psi(\tilde{u}_\lambda)$. If $d(\lambda)$ is positive the prior favors u_λ over f . If it is negative, the prior favors f instead of u_λ . In Fig. 6 we show the average value of $d(\lambda)$ computed on the dataset in Sun et al. (2013) for the three priors and different values of λ .

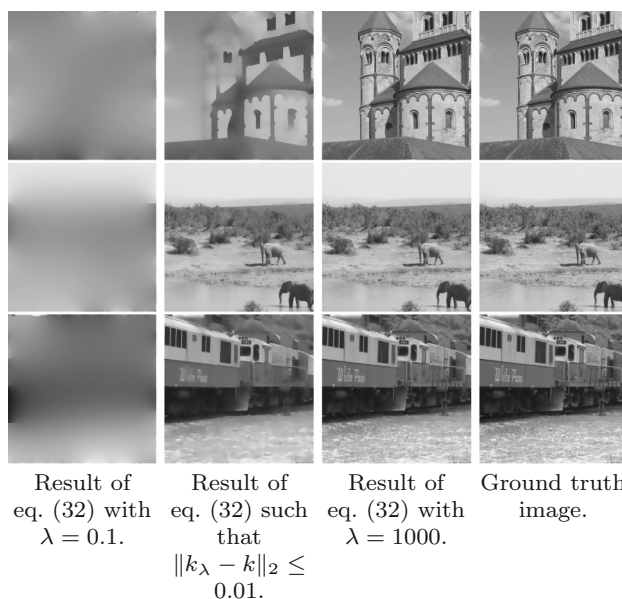


Fig. 5 Enlarged regions of images obtained by solving Eq. (32) and corresponding to the kernels in Fig. 4

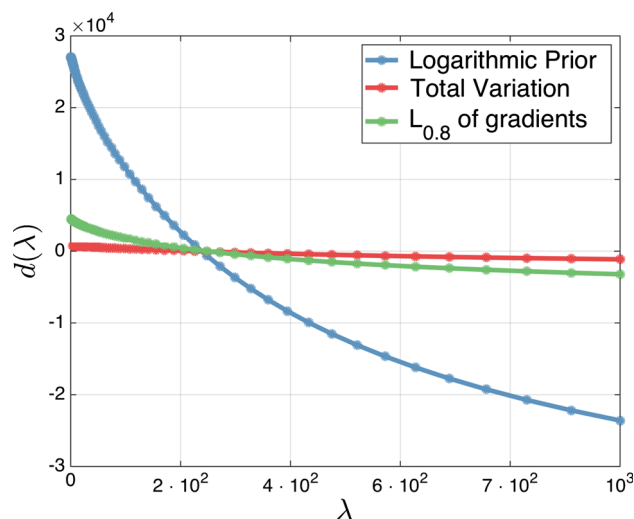


Fig. 6 Difference between the prior computed on the image u_λ and the blurry image f (Color figure online)

Since each prior has a different sensitivity to the parameter λ we normalize the plot such that each curve intersects the zero (in d) at the same point.

A first interesting result is that all three priors have a range for which the image \tilde{u} is favored. This aspect was highlighted also in Wipf and Zhang (2014). However, the magnitude of $d(\lambda)$ also matters. The more negative $d(\lambda)$ is, the more the prior discriminates between the sharp explanation and the blurry one. From this perspective, the logarithmic prior seems to have a stronger preference towards the sharp image compared to the other priors.

6 The Role of ϵ

In this section we study the behavior of the logarithmic prior introduced in Eq. (11) according to the choice of the parameter ϵ .

One relevant aspect to consider is how to avoid the degenerate constant solution. In this case we can compare two cases: one when $u = \text{constant}$ and one when $u = f$ and $k = \delta$. The idea is to make sure that the cost function favors the no-blur solution over the constant one. We can therefore plug in the two cases in the cost of problem (12) and obtain the following inequality

$$\log \|\nabla f\|_{2,\epsilon^2}^p < \lambda \|\bar{f} - f\|_2^2 + \frac{p}{2} MN \log \epsilon^2 \tag{35}$$

or, alternatively,

$$\log \left\| \frac{1}{\epsilon^2} \nabla f \right\|_{2,1}^p < \lambda \|\bar{f} - f\|_2^2, \tag{36}$$

where \bar{f} is the average value of f . Then, we use Jensen's inequality and the fact that the logarithm is a concave function to obtain an upper bound of the left hand side of Eq. (36)

$$\begin{aligned} & p/2 \sum_{i=1,j=1}^{N,M} \log \left[\left\| \frac{1}{\epsilon} \nabla f_{i,j} \right\|_{2,1}^2 \right] \\ & \leq \frac{pMN}{2} \log \left[\frac{1}{MN} \sum_{i=1,j=1}^{N,M} \left\| \frac{1}{\epsilon} \nabla f_{i,j} \right\|_{2,1}^2 \right]. \end{aligned} \tag{37}$$

Then, if we choose ϵ such that

$$\epsilon > \sqrt{\frac{\frac{1}{MN} \sum_{i=1,j=1}^{N,M} \|\nabla f_{i,j}\|_2^2}{e^{\frac{2\lambda}{pMN} \|f - \bar{f}\|_2^2} - 1}}} \tag{38}$$

the degenerate constant solution will be avoided. Also, notice that $\frac{2\lambda}{pMN} \|f - \bar{f}\|_2^2 > 0$ and $\frac{1}{MN} \sum_{i=1,j=1}^{N,M} \|\nabla f_{i,j}\|_2^2 > 0$ unless f is constant (in this case u constant is a plausible solution and it should not be avoided). This means that an ϵ that satisfies Eq. (38) always exists.

To understand more in detail how the choice of ϵ changes the behavior of the logarithmic prior we performed some empirical evaluation on the blind and non-blind problems. In this case we extracted patches of size 201×201 from the dataset proposed in Sun et al. (2013) and we synthetically blurred them with blurs of size 11×11 . We then solved the non-blind deconvolution problem (using the ground truth blur) and the blind deconvolution problem with different values of λ and ϵ . We finally computed the Sum of Squared Differences (SSD) error between the reconstructed images

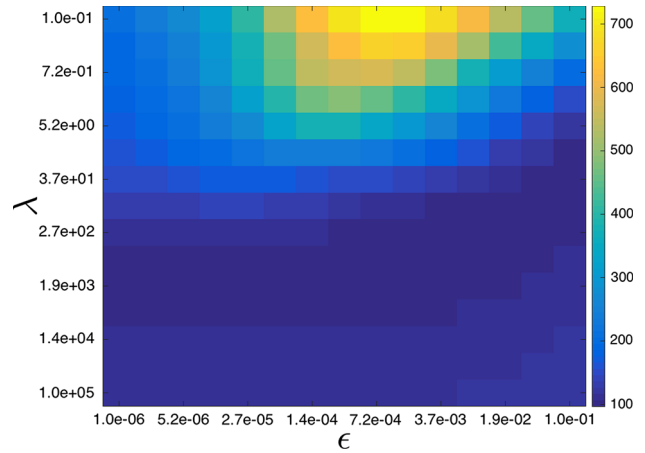


Fig. 7 SSD reconstruction error for the non-blind deconvolution problem with different values of λ and ϵ (Color figure online)

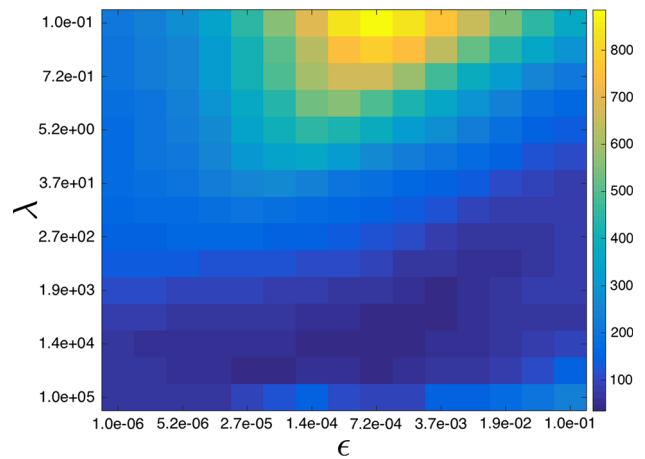


Fig. 8 SSD reconstruction error for the blind deconvolution problem (12) with different values of λ and ϵ (Color figure online)

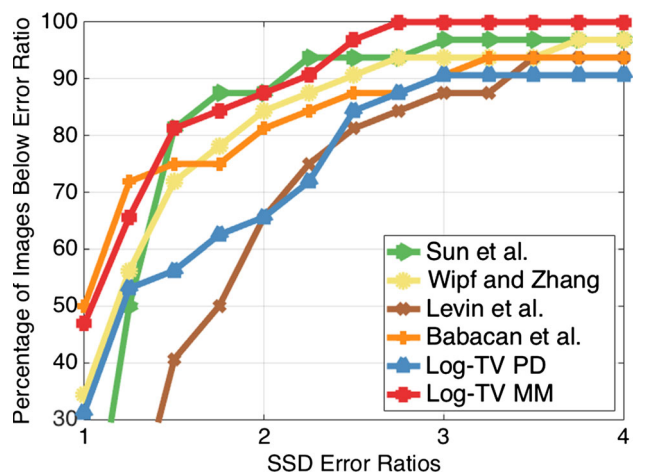


Fig. 9 Cumulative histogram of SSD ratio results on the dataset (Levin et al. 2011b) (Color figure online)

and the ground truth sharp image. In Fig. 7 we show a visualization of the SSD error for the non-blind deconvolution problem. The blue color denotes smaller errors while the

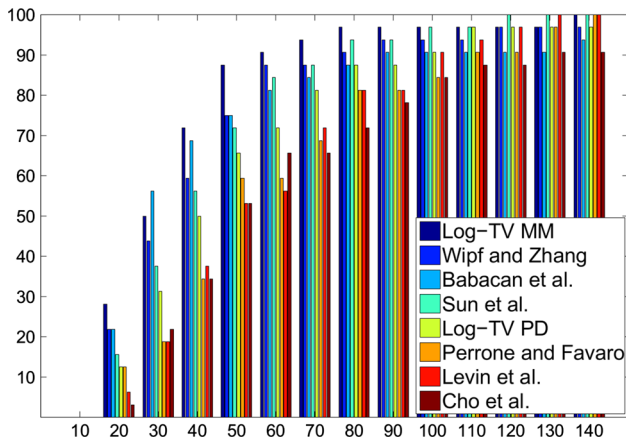


Fig. 10 Cumulative histogram of SSD results per image of the dataset (Levin et al. 2011b) (Color figure online)

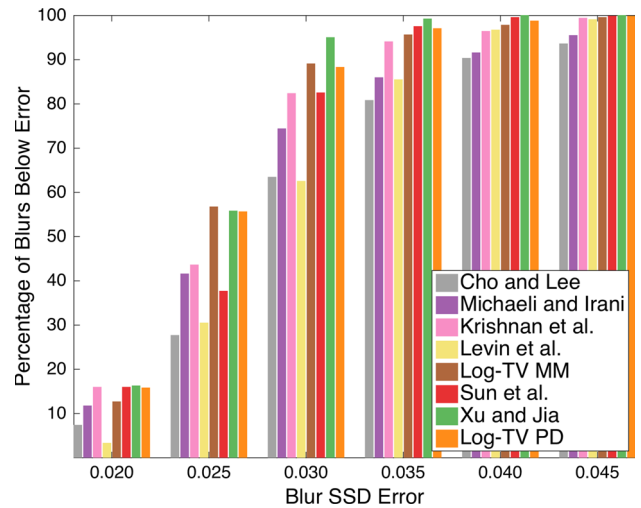


Fig. 13 Cumulative histogram of SSD kernel error on the dataset (Sun et al. 2013) (Color figure online)

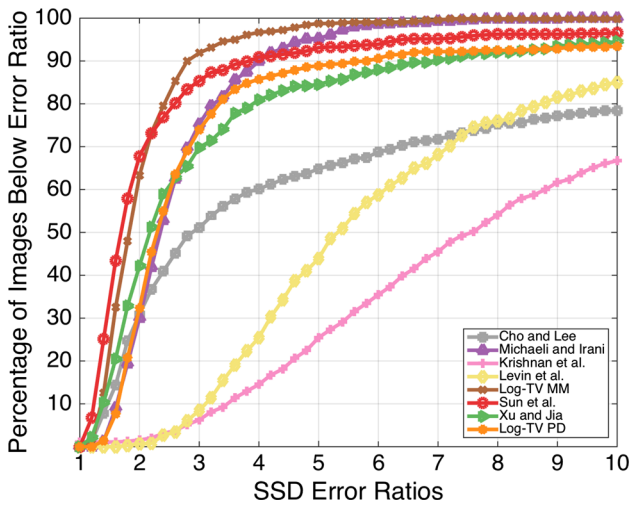


Fig. 11 Cumulative histogram of SSD ratio results on the dataset (Sun et al. 2013) (Color figure online)

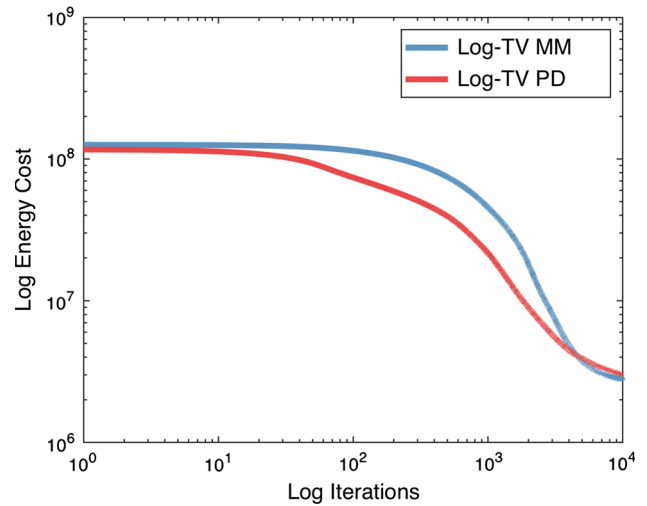


Fig. 14 Average cost per iteration for the Log-TV MM and Log-TV PD algorithms (Color figure online)

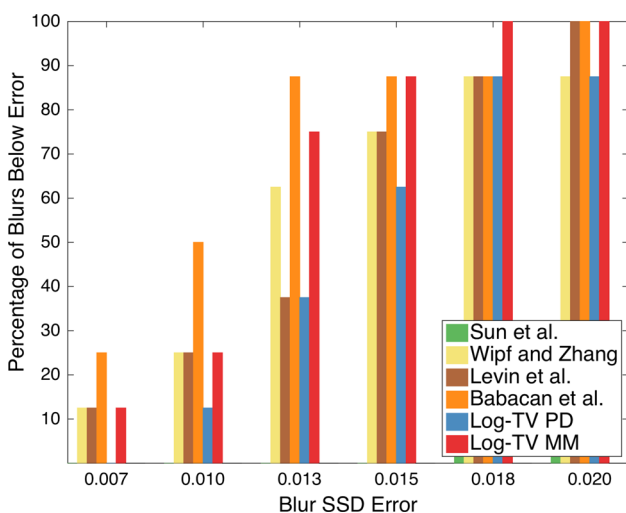


Fig. 12 Cumulative histogram of SSD kernel error on the dataset (Levin et al. 2011b) (Color figure online)

yellow color denotes larger errors. A small value of λ gives large errors because it removes details of the image, while a large value slightly increases the error because it is not able to remove the noise in the image. In the region surrounding the best value of λ , the value of ϵ appears to be less relevant. Nonetheless, large values of ϵ increase the error.

In Fig. 8 we show a visualization of the SSD error for the blind deconvolution problem. In this case the estimation seems more sensitive to the values of λ and ϵ and the set of values that gives the best errors is narrower.

7 Experiments

We evaluate the proposed algorithms on the datasets proposed in Levin et al. (2011b) and Sun et al. (2013). The first

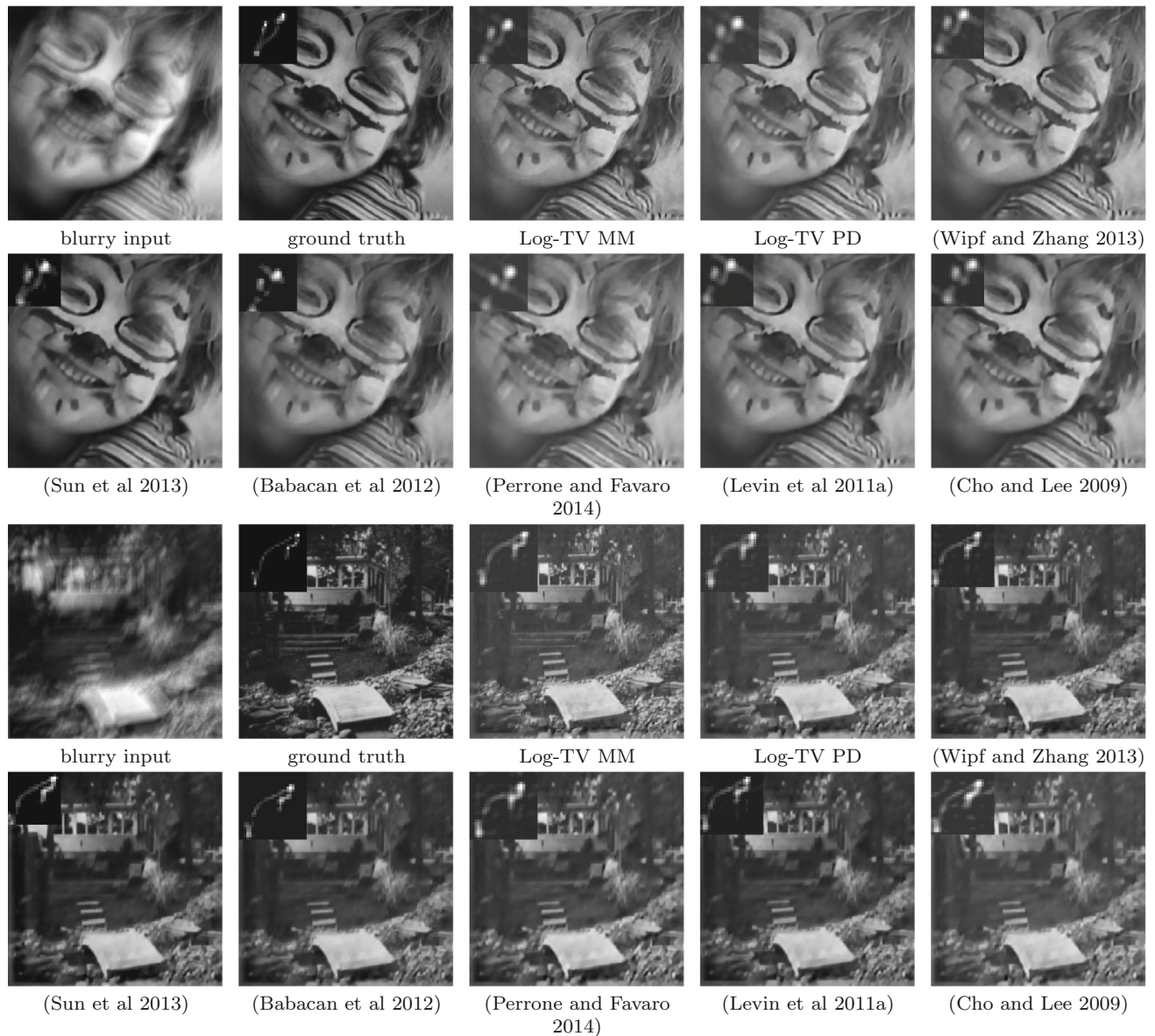


Fig. 15 Examples of deblurred images from dataset in [Levin et al. \(2011b\)](#)

dataset is made of 4 images of size 255×255 pixels blurred with 8 different blurs, and it is provided with ground truth sharp images and blurs. We use the metric defined in Eq. (34) as a performance measure as done also in [Levin et al. \(2011b\)](#).

For each method the same parameters are used for all the 32 blurry images of the dataset. For all the tests we used the non-blind deconvolution algorithm from [Levin et al. \(2007\)](#), where for each method we carefully selected the regularization parameter in order to have the best SSD ratio.

In Fig. 9 we show the cumulative histogram of the SSD ratios for several methods in the literatures and for our proposed algorithms (Log-TV MM and Log-TV PD). The MM algorithm achieves an error ratio equal to 1 for more than 50% of the images, clearly outperforming the methods from

[Wipf and Zhang \(2013\)](#) and [Babacan et al. \(2012\)](#), and, for most error ratios, the method of [Sun et al. \(2013\)](#). Our primal-dual method is on par with high performing variational Bayesian algorithms such as the one from [Levin et al. \(2011a\)](#). In Fig. 10 we also show the cumulative histogram of the SSD errors, while in Fig. 15 we show some of the sharp images obtained on this dataset.³ In Fig. 12 we show the SSD errors of the blur kernels compared with the ground truth kernels. The Log-TV MM algorithm achieves a very good blur reconstruction.

³ A list of all the experiments is available at www.cvg.unibe.ch/dperrone/logtv/.



Fig. 16 Deblurred Images with worst SSD ratio from dataset in [Sun et al. \(2013\)](#)

For our methods we used the same regularization parameter $\lambda = 30,000$, $\epsilon = 0.001$, $p = 1$ and 3500 iterations for each pyramid level. For the primal-dual algorithm we set $N_0 = 1$, $\tau = 0.005$ and $\sigma = \frac{1}{32\tau}$. The parameter values have been found experimentally. We used a pyramid scheme where the input image and the blur are down sampled at each level by $\sqrt{2}$, and the parameter λ is divided by the number 2.1. The number of levels of the pyramid is computed such that at the top level the blur kernel has a support of 3 pixels. For the other methods we used the estimates provided by the authors, or we ran their algorithm using the tuning that gives the best results.

In Fig. 11 we show an evaluation on the dataset from [Sun et al. \(2013\)](#). In this case we have 80 images of average size 1024×768 synthetically blurred with the blurs from [Levin et al. \(2009\)](#). The evaluation is made in a similar manner as in Fig. 9, but using the non-blind deconvolution from [Zoran and Weiss \(2011\)](#) as proposed in [Sun et al. \(2013\)](#). [Michaeli and Irani \(2014\)](#) have highlighted that images in this dataset with an error ratio smaller than 5 are visually pleasant (Fig. 12). Also in this case the MM algorithm outperforms the other methods except for the error ratio smaller than 2, where the algorithm in [Sun et al. \(2013\)](#) performs slightly better. The

primal-dual method performs on par with the algorithm proposed in [Michaeli and Irani \(2014\)](#) until error ratio smaller than 3.5, but then has a drop in performance. This suggests a larger instability of the algorithm for more difficult blurs. For this dataset we set the parameters to $\lambda = 12,500$ and $\epsilon = 0.0005$ for both algorithm. In Fig. 16 we show the images that have the worst error ratio in the dataset from [Sun et al. \(2013\)](#). In Fig. 13 we show the SSD errors of the blur kernels compared with the ground truth kernels. In this case the Log-TV MM algorithm is still a top performer, but not a clear winner like for the other experiments. Nonetheless, the reconstructed blur leads to a better reconstruction than with the other methods. This means that the metric used to assess the blur reconstruction (the Euclidean norm) can only roughly be used to predict the image reconstruction accuracy.

We also measured the performance of the two algorithms on a MacBook Pro with a 2.6 GHz Intel Core i7 processor and 16 GB of RAM with bandwidth of 1600 MHz. We consider the average time required to process an image of size 1024×800 for four different cases: a MATLAB unoptimized implementation of the MM algorithm (MatMM); a MATLAB unoptimized implementation of the PD algorithm (MatPD); a MATLAB implementation of the MM algorithm where the

prior gradient is computed by a C routine (MatCMM); a MATLAB implementation of the PD algorithm where the time of the proximity operators is computed by considering an ideal parallelization where the computation time of the parallelizable components is divided by the number of pixels (MatPPD). In our experiments the MatMM implementation took on average 194.56 min, the MatPD implementation 46.90 min, the MatCMM 46.13 min and the MatPPD would take, in the ideal case, 37.41 min.

To understand the difference in convergence of the MatMM and MatPD algorithms, we show the cost functional (12) at each iteration in Fig. 14. As the cost evolution changes depending on the image and blur, we compute and average of all costs on the dataset of Levin et al. (2011b). Also, because both algorithms use a pyramid scheme, we show the cost evolution of only the last level and we use the same initialization. We use a logarithmic scale for both axes so that two facts are emphasized: (1) The MatPD lowers the cost more quickly than the MatMM algorithm; (2) The MatMM algorithm achieves a smaller cost. Notice that the logarithmic prior in the cost functional can make the cost negative; thus, we add a constant before converting the cost to the logarithmic scale (Figs. 15, 16).

From our experiments it can be concluded that the primal-dual method can result in a faster implementation, but at the cost of being too coarse (due to the convex approximation of the logarithmic prior) to achieve the same accuracy of the MM algorithm.

8 Conclusions

In this paper we presented solutions to blind deconvolution based on a logarithmic image prior. The chosen prior is as effective as L_p norms with $p < 1$ on the image gradients, while at the same time leading to simple optimization schemes despite its non convexity. We show empirical experiments that support the choice of a logarithmic prior for blind deconvolution. To solve blind deconvolution with this image prior we propose a computationally efficient scheme via a primal-dual approach and a high-accuracy scheme via the majorization-minimization approach. Both approaches perform well and converge very robustly.

References

- Babacan, S. D., Molina, R., Do, M. N., & Katsaggelos, A. K. (2012). Bayesian blind deconvolution with general sparse image priors. In *ECCV*. Firenze: Springer.
- Burger, M., & Lucka, F. (2014). Maximum-a-posteriori estimates in linear inverse problems with log-concave priors are proper bayes estimators. *Inverse Problems*, *30*, 114004.
- Candes, E. J., Wakin, M. B., & Boyd, S. (2008). Enhancing sparsity by reweighted l_1 minimization. *Journal of Fourier Analysis and Applications*, *14*(5), 877–905.
- Chambolle, A., & Pock, T. (2011). A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, *40*(1), 120–145.
- Chan, T., & Wong, C. K. (1998). Total variation blind deconvolution. *IEEE Transactions on Image Processing*, *7*(3), 370–375.
- Chaudhuri, S., Velmurugan, R., & Rameshan, R. M. (2014). *Blind image deconvolution*. Cham: Springer.
- Cho, S., & Lee, S. (2009). Fast motion deblurring. *ACM Transactions on Graphics*, *28*(5), 1–8.
- Fergus, R., Singh, B., Hertzmann, A., Roweis, S. T., & Freeman, W. T. (2006). Removing camera shake from a single photograph. *ACM Transactions on Graphics*, *25*(3), 787–794.
- Hunter, D., & Lange, K. (2004). A tutorial on mm algorithms. *The American Statistician*, *58*, 30–37.
- Kenig, T., Kam, Z., & Feuer, A. (2010). Blind image deconvolution using machine learning for three-dimensional microscopy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(12), 2191–2204. doi:10.1109/TPAMI.2010.45.
- Keuper, M., Schmidt, T., Temerinac-Ott, M., Padeken, J., Heun, P., Ronneberger, O., & Brox, T. (2013). Blind deconvolution of wide-field fluorescence microscopic data by regularization of the optical transfer function (otf). In *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2179–2186).
- Krishnan, D., Tay, T., & Fergus, R. (2011). Blind deconvolution using a normalized sparsity measure. In *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 233–240).
- Krishnan, D., Bruna, J., & Fergus, R. (2013). Blind deconvolution with re-weighted sparsity promotion. CoRR arXiv:1311.4029.
- Levin, A., Fergus, R., Durand, F., & Freeman, W. (2007). Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics*, *26*, 70.
- Levin, A., Weiss, Y., Durand, F., & Freeman, W. T. (2009). Understanding and evaluating blind deconvolution algorithms. In *CVPR* (pp. 1964–1971). IEEE.
- Levin, A., Weiss, Y., Durand, F., & Freeman, W. (2011a). Efficient marginal likelihood optimization in blind deconvolution. In *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2657–2664). doi:10.1109/CVPR.2011.5995308.
- Levin, A., Weiss, Y., Durand, F., & Freeman, W. T. (2011b). Understanding blind deconvolution algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(12), 2354–2367.
- Michaeli, T., & Irani, M. (2014). Blind deblurring using internal patch recurrence. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.) *Computer Vision—ECCV 2014*. Lecture Notes in Computer Science, vol. 8691 (pp. 783–798). Cham: Springer International Publishing. doi:10.1007/978-3-319-10578-9_51.
- Möllenhoff, T., Strelakovsky, E., Möller, M., & Cremers, D. (2014a). Low rank priors for color image regularization. In *Proceedings of 10th International Conference, Energy Minimization Methods in Computer Vision and Pattern Recognition, EMMCVPR 2015* (pp. 126–140). Hong Kong, China, January 13–16, 2015.
- Möllenhoff, T., Strelakovsky, E., Möller, M., & Cremers, D. (2014b). The primal-dual hybrid gradient method for semiconvex splittings. CoRR arXiv:1407.1723.
- Ochs, P., Chen, Y., Brox, T., & Pock, T. (2014). ipiano: Inertial proximal algorithm for nonconvex optimization. *SIAM Journal on Imaging Sciences*, *7*(2), 1388–1419.
- Perrone, D., & Favaro, P. (2014). Total variation blind deconvolution: The devil is in the details. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Perrone, D., Diethelm, R., & Favaro, P. (2014). Blind deconvolution via lower-bounded logarithmic image priors. In *Proceedings of 10th International Conference, Energy Minimization Methods in*

- Computer Vision and Pattern Recognition, EMMCVPR 2015* (pp. 112–125). Hong Kong, China, January 13–16, 2015.
- Rockafellar, R. (1970). *Convex analysis*. Princeton, NJ: Princeton University Press.
- Rudin, L. I., Osher, S., & Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1–4), 259–268. doi:10.1016/0167-2789(92)90242-F.
- Shan, Q., Jia, J., & Agarwala, A. (2008). High-quality motion deblurring from a single image. *ACM Transactions on Graphics*, 27, 73.
- Srivastava, A., Lee, A., Simoncelli, E. P., & Zhu, S. C. (2003). On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision*, 18, 17–33.
- Stekalovskiy, E., & Cremers, D. (2014). Real-time minimization of the piecewise smooth Mumford-Shah functional. *Proceedings of Computer Vision 13th European Conference ECCV 2014, Zurich, Switzerland, September 6–12, 2014 Part II* (pp. 127–141).
- Strong, D., & Chan, T. (2003). Edge-preserving and scale-dependent properties of total variation regularization. *Inverse Problems*, 19(6), S165.
- Sun, L., Cho, S., Wang, J., & Hays, J. (2013). Edge-based blur kernel estimation using patch priors. In *2013 IEEE International Conference on Computational Photography (ICCP)* (pp. 1–8).
- Wipf, D., & Zhang, H. (2013). Analysis of Bayesian blind deconvolution. *Energy minimization methods in computer vision and pattern recognition* (pp. 40–53). Berlin: Springer.
- Wipf, D., & Zhang, H. (2014). Revisiting Bayesian blind deconvolution. *Journal of Machine Learning Research*, 15, 3595–3634.
- Xu, L., & Jia, J. (2010). Two-phase kernel estimation for robust motion deblurring. In *Proceedings of the 11th European Conference on Computer Vision: Part I ECCV'10* (pp. 157–170). Berlin: Springer.
- Xu, L., Zheng, S., & Jia, J. (2013). Unnatural l0 sparse representation for natural image deblurring. In *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1107–1114).
- Zoran, D., & Weiss, Y. (2011). From learning models of natural image patches to whole image restoration. In *Proceedings of the 2011 International Conference on Computer Vision, ICCV '11* (pp. 479–486). IEEE Computer Society Washington, DC, USA.