Fast Bayesian blind deconvolution with Huber Super Gaussian priors ☆

Xu Zhou [a,b,∗], Miguel Vega [c], Fugen Zhou [a], Rafael Molina [c], Aggelos K. Katsaggelos [d]

[a] *Image Processing Center, Beihang University, Beijing 100191, China*
[b] *ShanghaiTech University, Shanghai 200031, China*
[c] *Departamento de Ciencias de la Computación e I.A., Universidad de Granada, Granada 18071, Spain*
[d] *Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208-3118, USA*

A R T I C L E   I N F O

A B S T R A C T

Expectation Maximization (EM) based inference has already proven to be a very powerful tool to solve blind image deconvolution (BID) problems. Unfortunately, three important problems still impede the application of EM in BID: the undesirable saddle points and local minima caused by highly nonconvex priors, the instability around zero of some of the most interesting sparsity promoting priors, and the intrinsic high computational cost of the corresponding BID algorithm. In this paper we first show how Super Gaussian priors can be made numerically tractable around zero by introducing the family of Huber Super Gaussian priors and then present a fast EM based blind deconvolution method formulated in the image space. In the proposed computational approach, image and kernel estimation are performed by using the Alternating Direction Method of Multipliers (ADMM), which allows to exploit the advantages of FFT computation. For highly nonconvex priors, we propose a Smooth ADMM (SADMM) approach to avoid poor BID estimates. Extensive experiments demonstrate that the proposed method significantly outperforms state-of-the-art BID methods in terms of quality of the reconstructions and speed.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

In the imaging process, factors like camera shake, defocus and atmospheric turbulence cause blurriness in the observed image. To remove the undesirable blur, one can use nonblind or blind deconvolution methods, depending on whether the blur is known or unknown, respectively. Blind image deconvolution (BID) is an interesting and challenging problem which finds application in areas like astronomical imaging, medical imaging, and computational photography [1].

In general, the degraded image y can be approximately modeled as [2]

$$y = Hx + n \tag{1}$$

where x and n are the original image and noise, respectively, and H is the convolution matrix formed from the blur kernel h.

Given a single degraded image y, the BID goal is to estimate the original image x and blur kernel h. BID is a severely ill-posed inverse problem, since there are infinitely many solutions for x and h. To find the desirable ones from the set of uncountable solutions, additional assumptions on both x and h are indispensable. During the last decade, BID has been intensively studied in the image processing and computer vision communities [3–21]. It could be rightly claimed that the vast majority of state-of-the-art BID methods make use of sparsity promoting image priors. Most existing BID methods can be classified into three categories, that is, VB (variational Bayes), MAP (maximum *a posterior*), and other methods which we now briefly review. Notice that in this classification EM inference is included in the VB category.

Miskin and MacKay [3] assume a mixture of Laplacian priors for cartoon images and propose a VB approach, termed ensemble learning, to estimate the blur. Likas and Galatsanos [4] propose a similar VB approach with Gaussian priors, see also Molina et al. [22] for VB with Gaussian priors for image and blur. Fergus et al. [5] exploit a mixture-of-Gaussians (MOG) prior and extend the VB approach [3] to handle natural images with nonparametric motion blurs. Levin et al. [10] also utilize an MOG prior on the latent image and propose a MAP_h approach, namely marginalizing all latent images and estimating the kernel alone. Making use of the conjugate concave principle, Babacan et al. [13] propose a general

∗ Corresponding author.
E-mail addresses: zhouxu179@163.com (X. Zhou), mvega@ugr.es (M. Vega), zhfugen@buaa.edu.cn (F. Zhou), rms@decsai.ugr.es (R. Molina), aggk@eecs.northwestern.edu (A.K. Katsaggelos).

Bayesian framework, which allows the use of general sparse priors for BID. This framework includes [5] and [10] as particular cases. Recently, following the MAP_h approach [10] and assuming a Gaussian representation model, Zhang and Wipf [20] have proposed a BID method with parameter estimation to remove camera shake. Following the work in [13], Vega et al. [21] suggest introducing an additional parameter on the image prior and show that this parameter can be used to improve the quality of the restoration. Interested readers are referred to [2] and [23] for recent reviews on VB based BID methods.

Instead of imposing a prior distribution on the latent image, one can also use shock filters [24] to promote sparsity, see [7,8,16]. The shock filter provides a highly smooth image with predicted sharp step edges, which can be used as a reference image to regularize the latent image. Since the predicted step edges are probably inaccurate, a better approach is to use nondimensional sparsity measures [11,19,17], which model ideal step edges as local minima. Sun et al. [14] propose a parametric step edge model to promote sparsity, where the parameters are learned from synthetic or natural image patches. Recently, Perrone and Favaro [25] have proposed a logarithmic TV (Total Variation) based BID methods using a delayed kernel normalization scheme [26]. The above MAP methods seek a cartoon like image to estimate the blur kernel and hence are robust to noise; however they may fail if the original image lacks step edges or is dominated by textures.

While most of the existing BID methods focus on image priors, priors on the kernel have received much less attention. Cai et al. [9] suggest using curvelet representations for motion blurs, since the curvelet representation takes into account both the continuity and sparsity of motion blur kernels. Goldstein and Fattal [12] propose a spectrum power prior on the blur kernel, based on the assumption that the spectrum power of a natural image drops quadratically as the frequency increases. Zhou et al. [27] propose a variational Dirichlet approach to estimate the blur kernel and show that it helps reduce noise in the estimated kernel.

Although Super Gaussian (SG) [13] distributions have proven to be the right tool to introduce prior information in BID, some of the most widely used such distributions, like ℓ_p and log priors, suffer from lack of differentiability around zero. In this paper we introduce in a sound way a new family of SG distributions fully differentiable and thus well behaved around zero. We call them Huber Super Gaussian (HSG) priors.

As is widely known, VB based BID methods involve two major steps: image and kernel estimation. They correspond to solving large-scale linear equation problems and providing solution to large scale quadratic problems with constraints, respectively. In this work, by incorporating the latest techniques in image deconvolution [13,28], we also present a fast Bayesian blind deconvolution method formulated in the image space for Huber Super Gaussian priors.

Recent advances in BID [13,20,17,23,25,27] show that highly nonconvex priors are very effective in BID problems. However, the local minima and saddle point problem caused by nonconvex priors, e.g., log prior, has not been well addressed. We report that the EM approach, which provides an estimate of the mean of the posterior distribution of the latent image and has been used in Bayesian BID methods [10,13,20], often leads to a stationary point with a large objective function value and an image estimate which is not adequately smooth. We show that, with suitable parameter settings and iterative schemes, ADMM can be used to tackle this problem. The proposed ADMM with a large penalty weight finds a *smooth* approximate solution to the reweighted linear equations, rather than the exact solution as searched by the EM approach. Experiments show that the proposed ADMM, which seeks a smooth image mean rather than the exact image mean pro-

vided by the EM approach, leads to a smoother image with a lower cost value, and the proposed BID method significantly outperforms state-of-the-art BID methods in terms of restoration quality and speed.

In summary, the contributions of this paper are a) the introduction of the Huber Super Gaussian prior family to improve numerical stability and convergence speed, and b) a fast method for nonconvex image deconvolution, which avoids undesirable stationary points and yields comparable or even better results than the EM approach in terms of ringing artifact reduction.

The paper is organized as follows. In section 2 we formulate the BID problem. We introduce the observation model, define the family of Huber Super Gaussian priors and present the inference framework. Section 3 is devoted to the image and blur reconstruction algorithms. Section 4 discusses, in detail, all the parameters involved in the BID algorithm we propose. Section 5 compares our approach with current state-of-the-art BID algorithms, and section 6 concludes the paper.

2. Prior model

Let the sizes of y, x, and h in Eq. (1) be $\sqrt{N} \times \sqrt{N}$, $\sqrt{N} \times \sqrt{N}$ and $\sqrt{K} \times \sqrt{K}$, respectively. We assume that H is block circulant with circulant blocks, and then have $H = \mathcal{F}^{-1} diag(\mathcal{H})\mathcal{F}$, where $\mathcal{F}$ denotes the 2-D DFT and $\mathcal{H} = \mathcal{F}Ph$ is the Fourier transform of h, with P being an $N \times K$ matrix that pads h with zeros to the size $\sqrt{N} \times \sqrt{N}$. The dimensions of H and $diag(\mathcal{H})$ are therefore $N \times N$. For simplicity, we overload the notations x, y, and h with two meanings, a 2-D array or a column vector arranged in lexicographic order; which notation is being used will be clear from the context.

Assuming an i.i.d. Gaussian noise with variance σ^2, we can write

$$p(y|x, h) = \frac{1}{(\sqrt{2\pi}\sigma)^N} \exp(-\frac{\|Hx - y\|_2^2}{2\sigma^2}) \tag{2}$$

Let us now proceed to define the image prior model we use in the paper by introducing first its energy (penalty) function.

Let $\rho(s) : \Re \rightarrow [-\infty, +\infty]$ be symmetric around zero with $\rho(\sqrt{s})$ concave and increasing for $s \in [0, \infty)$. This condition is equivalent to $\rho'(s)/s$ being decreasing on $(0, +\infty)$, that is, for $s_1 \leq s_2$, $\rho'(s_1)/s_1 \geq \rho'(s_2)/s_2$. A probability distribution of the form $p(s) \propto \exp(-\rho(s))$ is called a Super Gaussian (SG) distribution. SG distributions promote sparsity. Their use in blind image restoration problems was first proposed in [13]. Examples of energies associated to SG distributions are shown in Figs. 1(a) and 1(b), in which $\rho(s) = \log(|s|)$, the so-called log prior, was first introduced in [13].

The energy, $\rho(\cdot)$, associated to an SG distribution can be represented as (see [29])

$$\rho(s) = \inf_{\xi > 0} \frac{1}{2}\xi s^2 - \rho^* \left(\frac{1}{2}\xi \right) \tag{3}$$

where $\rho^*(\xi/2)$ is the concave conjugate function

$$\rho^* \left(\frac{1}{2}\xi \right) = \inf_s \frac{1}{2}\xi s^2 - \rho(s) . \tag{4}$$

Furthermore, as shown in [13], the infimum in (3) is achieved at $\xi = \rho'(s)/s$.

Equation (3) provides a quadratic upper bound to the energy of an SG distribution which naturally leads to a Gaussian approximation. Unfortunately, for very useful energies associated to SG distributions, like $\rho(s) = |s|^p/p$ $(0 < p < 1)$ and $\rho(s) = \log(|s|)$, $\rho(s)$ is not differentiable at zero and $\xi = \rho'(s)/|s|$ is unbounded as s approaches zero. We overcome this problem by defining here

$\rho(s)$
$\frac{1}{2}s^2$
$\frac{1}{p}\lvert s \rvert^p$
$\log \lvert s \rvert$
$-\log \sum_j \frac{\pi_j}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{s^2}{2\sigma_j^2}\right)$

(a)

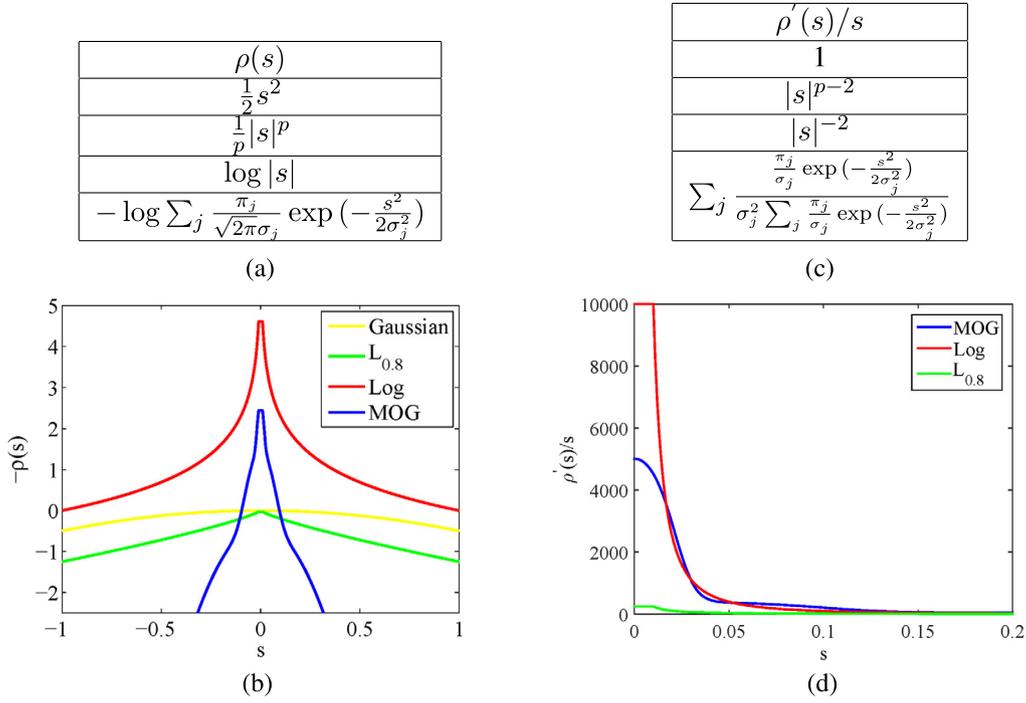| $\rho'(s)/s$ |
|---|
| 1 |
| $\lvert s \rvert^{p-2}$ |
| $\lvert s \rvert^{-2}$ |
| $\sum_j \dfrac{\frac{\pi_j}{\sigma_j} \exp\left(-\frac{s^2}{2\sigma_j^2}\right)}{\sigma_j^2 \sum_j \frac{\pi_j}{\sigma_j} \exp\left(-\frac{s^2}{2\sigma_j^2}\right)}$ |

(c)

(b)

(d)

Fig. 1. (a) and (b): examples of penalty functions $\rho(s)$, where the MOG is obtained from Levin et al. [10]. (c): their corresponding $\rho'(s)/s$. (d) Plots of $\rho'_\epsilon(s)/s$, where $\epsilon = 0.01$ for Log and $L_{0.8}$.

for $\epsilon > 0$ the robust symmetric around zero function $\rho_\epsilon(s)$ associated with $\rho(s)$ as follows

$$\rho_\epsilon(s) = \begin{cases} \rho(s) & \text{for } s \geq \epsilon \\ \frac{\rho'(\epsilon)}{2\epsilon}s^2 - \left(\frac{\rho'(\epsilon)}{2}\epsilon - \rho(\epsilon)\right) & \text{for } 0 \leq s \leq \epsilon \end{cases} \quad (5)$$

Furthermore, since $\rho'(s)/s$ is decreasing on $(0, \infty)$,

$$\frac{\rho'_\epsilon(s)}{s} = \min(\rho'(s)/s, \rho'(\epsilon)/\epsilon) = \begin{cases} \frac{\rho'(s)}{s} & \text{for } s \geq \epsilon \\ \frac{\rho'(\epsilon)}{\epsilon} & \text{for } 0 \leq s \leq \epsilon \end{cases} \quad (6)$$

and so $\rho_\epsilon(s)$ is also the energy of an SG distribution.

For the log and ℓ_p priors mentioned above we have $\rho'_\epsilon(s)/s = \min(s^{-2}, \epsilon^{-2})$ and $\rho'_\epsilon(s)/s = \min(s^{p-2}, \epsilon^{p-2})$, respectively, see Fig. 1(d). The use of $\rho_\epsilon(s)$ makes $\rho(s)$ differentiable at zero and the prior tractable. Following the terminology in [30], we call these priors Huber Super Gaussian (HSG) priors. We defer the analysis of this prior to section 4.

Having defined the robust energy associated to an HSG distribution, we now proceed to introduce the image model we use. The prior $p(x)$ on the latent image has the form

$$p(x) = Z \exp\left(-\sum_{\gamma=1}^{L} \sum_i \rho_\epsilon(x_\gamma(i))\right), \quad (7)$$

where $x_\gamma = f_\gamma \otimes x$ with $\otimes$ denoting a circular convolution operator and $\{f_\gamma\}$ being a set of L high pass filters. We will discuss the filters used in this work in the experimental section.

For the blur we do not assume any prior knowledge apart from the fact that it must be nonnegative and its coefficients should add to one. That is, $p(h) \propto \text{const}$, subject to: $h(i) \geq 0$, $\sum_i h(i) = 1$.

3. Bayesian inference

The process to estimate the blur and image will be carried out as follows. First, we estimate the blur using the EM approach to solve

$$\hat{h} = \arg\max_h \int p(x, h, y) dx \quad (8)$$

$$s.t.\, h(i) \geq 0, \sum_i h(i) = 1, \quad (9)$$

and then once the blur has been estimated we use it to estimate the original image utilizing a nonblind image deconvolution algorithm. Notice that, in principle, we should use $p(x|\hat{h}, y)$ to estimate the original image. However, our sparsity promoting image prior based on $\rho_\epsilon(\cdot)$ is designed to obtain a good blur estimate when the joint distribution is marginalized on the image at the cost of removing fine details in the image estimate.

Directly applying EM inference on $p(x, h, y)$ to estimate the blur is infeasible due to the form of $p(x)$. Since $\rho_\epsilon(s)$ is the energy associated to an SG distribution, we can write

$$p(x) \geq Z \prod_{\gamma=1}^{L} \prod_i \exp\left(-\left(\frac{\xi_\gamma(i)}{2}x_\gamma^2(i) - \rho_\epsilon^*\left(\frac{1}{2}\xi_\gamma(i)\right)\right)\right), \forall \xi_\gamma(i) > 0. \quad (10)$$

This Gaussian-like lower bound allows for the expectation of the joint distribution to be calculated analytically. We have

$$p(x, h, y) \geq p(y|x, h)p(h)Ze^{-\sum_{\gamma=1}^{L}\sum_i\left(\frac{\xi_\gamma(i)}{2}x_\gamma^2(i) - \rho_\epsilon^*\left(\frac{1}{2}\xi_\gamma(i)\right)\right)}$$

$$\equiv Q(y, x, h, \boldsymbol{\xi}), \quad (11)$$

where $\boldsymbol{\xi} = \{\xi_\gamma(i), \gamma = 1, \dots, L, i = 1, \dots, N\}$ with all components positive.

We finally obtain

$$\hat{h} = \arg\max_h E[\log(Q(y, x, h, \hat{\boldsymbol{\xi}}))]_{\hat{q}(x)}, \quad (12)$$

$$\hat{q}(x) \propto Q(y, x, \hat{h}, \hat{\boldsymbol{\xi}}), \quad (13)$$

$$\hat{\boldsymbol{\xi}} = \arg\max_{\boldsymbol{\xi}>0} E[\log(Q(y, x, \hat{h}, \boldsymbol{\xi}))]_{\hat{q}(x)}. \quad (14)$$

Notice that $\hat{q}(x)$ is an approximation of the posterior distribution $p(x|\hat{h}, y)$.

3.1. Estimation of image and variational parameter

For the latent image, we obtain from (13),

$$\log \hat{q}(x) = -\frac{1}{2\sigma^2} \|\hat{H}x - y\|_2^2 - \frac{1}{2} \sum_{\gamma=1}^{L} x_\gamma^T diag(\hat{\xi}_\gamma) x_\gamma, \quad (15)$$

which is a multivariate Gaussian with precision matrix

$$C_x^{-1} = \frac{1}{\sigma^2} \hat{H}^T \hat{H} + \sum_{\gamma=1}^{L} F_\gamma^T diag(\hat{\xi}_\gamma) F_\gamma, \quad (16)$$

where F_γ is an $N \times N$ circulant convolution matrix formed by the kernel f_γ. The mean $\hat{x}$ is used as the estimate for x, which is obtained by solving the following linear system of equations

$$C_x^{-1} \hat{x} = \frac{1}{\sigma^2} \hat{H}^T y. \quad (17)$$

This linear system can be solved by ADMM [28] very efficiently, as we will show in the following.

For the variational parameter $\hat{\xi}$, we obtain from (14)

$$\hat{\xi}_\gamma(i) = \frac{\rho'_\epsilon(v_\gamma(i))}{v_\gamma(i)} = \min(\rho'(v_\gamma(i)))/v_\gamma(i)), \rho'(\epsilon)/\epsilon), \quad (18)$$

where $v_\gamma(i) = \sqrt{E[x_\gamma^2(i)]}$, $1 \le i \le N$, and the expected value is easily calculated using the distribution $\hat{q}(x)$ and the approximation similar to [10,13]

$$C_x \approx (diag(C_x^{-1}))^{-1}. \quad (19)$$

Given a kernel estimate, we need to alternatively solve the linear system (17) and update the weights (18). This "joint" estimation of the image and weights is understood as a standard nonblind image deconvolution procedure. It proceeds as an IRLS (Iterative Reweighted Least Squares) frequently used in sparse signal recovery problems [28]. Previous methods [10,13,20] use the Conjugate Gradient (CG) algorithm to solve (17), which is very time consuming. Instead of using CG, we use ADMM to solve (17). The core idea is to treat the solution of a linear system as the minimizer of a quadratic function (see [28]). We transform the unconstrained problem into a constrained one by introducing the auxiliary variables, $v_\gamma = f_\gamma \otimes x$ [31]. Consequently, for a given blur H, $\hat{x}$ can be rewritten as the minimizer of

$$f(x, v_\gamma) = \frac{1}{2} \|Hx - y\|_2^2 + \sigma^2 \sum_{\gamma=1}^{L} v_\gamma^T diag(\hat{\xi}_\gamma) v_\gamma \quad (20)$$
$$s.t. \ v_\gamma = F_\gamma x,$$

which can be solved efficiently using ADMM. To be precise, we form the augmented Lagrangian

$$\mathcal{L}_{\beta_v}(x, v_\gamma, dv_\gamma) = \frac{1}{2} \|Hx - y\|_2^2 + \sigma^2 \sum_{\gamma=1}^{L} v_\gamma^T diag(\hat{\xi}_\gamma) v_\gamma$$
$$+ \frac{\beta_v}{2} \sum_{\gamma=1}^{L} \|v_\gamma - F_\gamma x + dv_\gamma\|_2^2 \quad (21)$$

where β_v is the penalty weight which enforces the constraints in (20) and controls the convergence rate, dv_γ are the (scaled) dual variables (a.k.a., Lagrangian multipliers). ADMM consists of the iterations

---

Algorithm 1 Fast image estimation algorithm.

Require: y, h, initial x, $\rho_\epsilon(s)$, β_v, σ,
1: Precompute $\mathcal{H}$, $\overline{\mathcal{H}}$, $\overline{\mathcal{H}}\mathcal{F}y$, Λ_h, Λ_γ
2: $dv_\gamma = 0$, $C_x = 0$.
3: Initialize ξ_γ with a large constant or (18).
4: **repeat**
5:   Use ADMM (see (22), (23), and (24)) to update the solution x to (17)
6:   Update ξ_γ using (18)
7:   Approximate $C_x(i, i)$ with $1/C_x^{-1}(i, i)$
8: **until** The relative variation in x is very small
9: **return** x

$$x = \mathcal{F}^{-1}\left\{ \frac{\overline{\mathcal{H}}\mathcal{F}y + \beta_v \mathcal{F}[\sum_{\gamma=1}^{L} F_\gamma^T(v_\gamma + dv_\gamma)]}{\Lambda_h + \beta_v \sum_{\gamma=1}^{L} \Lambda_\gamma} \right\}, \quad (22)$$

$$v_\gamma = \frac{\beta_v}{\beta_v + \sigma^2 \hat{\xi}_\gamma}(F_\gamma x - dv_\gamma), \quad (23)$$

$$dv_\gamma = dv_\gamma + v_\gamma - F_\gamma x, \quad (24)$$

where $\overline{\mathcal{H}}$ is the complex conjugate of $\mathcal{H}$, and Λ_h and Λ_γ are the eigenvalues of $H^T H$ and $F_\gamma^T F_\gamma$, respectively.

The final algorithm for image estimation is presented in Algorithm 1, which achieves state-of-the-art performance in terms of quality of restorations as well as speed. It should be noted that the image estimation problem is not trivial when $\rho(s)$ is nonconvex. For instance, if we choose the log prior, it is shown in [28] that Algorithm 1 will converge to a stationary point of the cost function, which might be a saddle point or a local minimum. Therefore, the optimization scheme makes a considerable difference to the final reconstruction.

Before concluding this section we would like to mention that if $v_\gamma(i) = \sqrt{E^2[x_\gamma(i)]}$ ($1 \le i \le N$, $C_x = 0$), then our procedure would not take into account the uncertainty of the image estimate and we would have a MAP procedure.

3.2. Estimation of blur

To estimate the blur we have

$$\hat{h} = \arg\min_h \|H\hat{x} - y\|_2^2 + h^T D_x h \quad (25)$$

$$s.t. \ h(i) \ge 0, \sum_i h(i) = 1, \quad (26)$$

where D_x is the $K \times K$ matrix given by

$$D_x(m, n) = \sum_{j=1}^{N} C_x(m + j, n + j). \quad (27)$$

Given an image estimate, we use ADMM to efficiently solve the quadratic optimization problem (25) with nonnegativity and normalization constraints (26). By introducing the constraint $\mathcal{H} = \mathcal{F}Ph$ (see [18] for a different variable splitting), we find $\hat{h}$ in (25) as the minimizer of

$$g(h, \mathcal{H}) = \|\mathcal{H} \circ \mathcal{X} - \mathcal{F}y\|_2^2 + h^T D_x h \quad (28)$$

$$s.t. \ h(i) \ge 0, \sum_i h(i) = 1,$$
$$\mathcal{H} = \mathcal{F}Ph,$$

where $\mathcal{X} = \mathcal{F}x$ denotes the Fourier transform of x and $\circ$ element-wise multiplication. The ADMM algorithm for this constrained problem is presented in Algorithm 2, in which the updates for $\mathcal{H}$ and h are given by

Algorithm 2 Fast kernel estimation algorithm.

Require: y, x, initial h, β_H,
1: Precompute D_x using (27), $dH = 0$.
2: **while** not converged **do**
3: 　　Update $\mathcal{H}$ using (29)
4: 　　Update h by finding the KKT point of (30)
5: 　　$dH = dH + \mathcal{H} - \mathcal{F}Ph$
6: **end while**
7: **return** h

$$\mathcal{H} = \frac{\overline{\mathcal{X}} \circ (\mathcal{F}y) + \beta_H(\mathcal{F}Ph - dH)}{\overline{\mathcal{X}} \circ \mathcal{X} + \beta_H} \tag{29}$$

$$h = \arg\min_h \|h - P^T \mathcal{F}^{-1}(\mathcal{H} + dH)\|_2^2 + h^T D_x h \tag{30}$$

$$s.t. \ h(i) \geq 0, \ \sum_i h(i) = 1$$

where the L_2-*norm* term in (30) is obtained by observing that $\forall h \in \Re^K$, the following equalities hold $\|\mathcal{F}Ph\|_2^2 = \|h\|_2^2 = \|P^T\mathcal{F}^{-1}\mathcal{H}\|_2^2$.

The convex problem (30) can be solved by the Karush–Kuhn–Tucker (KKT) conditions, which refers to solving a system of nonlinear equations. Specifically, we use a fixed point iteration method to find the KKT point. It should be noted that the computation complexity of solving problem (30) is only $O(K)$ rather than $O(K^2)$, since D_x is approximated by a diagonal matrix.

The proposed ADMM splits the constrained large scale quadratic program into two simpler subproblems: one which can be solved directly in Fourier domain and another which is a small scale quadratic program that can be solved iteratively. Moreover, with 10 ADMM iterations, a very good approximate solution can be obtained. In most cases, we observed that only 5 ADMM iterations are needed for convergence.

3.3. Multiscale blind deconvolution

As with previous BID methods [5,10,13], we also adopt here a multiscale approach to estimate the kernel. Fergus et al. [5] point out that single scale BID may suffer from local minima, especially when the support of the kernel is large. The multiscale approach, however, is very useful for large support kernel estimation. The rational behind this approach is that at the coarsest level, the blur is alleviated significantly, so that it is easier to estimate a kernel from the downsampled image. At the next finer level, this kernel estimate is upsampled and can be used as a good starting point for the single scale BID. Repeating this process until the finest level, we obtain a better kernel estimate.

3.4. Final image reconstruction

The estimated image $\hat{x}$ obtained above is appropriate for estimating the blur but, as we have already explained and also as we will show in the experimental section, it is over-smooth. This is because the sparsity promoting weight $\sigma^2 \xi_\gamma$ is too large to recover the fine details, especially in the case of the log prior. To recover more details, we reconstruct the final sharp image x^* by solving the following problem [32,33],

$$x^* = \arg\min_x \frac{1}{2}\|Hx - y\|_2^2 + \frac{\lambda_x}{p}\sum_\gamma \|f_\gamma \otimes x\|_p, \tag{31}$$

where $\{f_\gamma\}$ is a set of derivative filters to be defined in the experimental section, p is typically in the range [0.6, 0.8] [32] and the ℓ_p regularizer should be understood in the HSG form.

3.5. Summary

So far we have presented a fast EM approach for blind image deconvolution formulated in the image space, using a lower bound on the HSG image priors.

The benefit of working in the image space rather than in the filter space [5,10,13] is threefold. First, utilizing the image space can save computations by nearly $100 \times \frac{L-1}{L}\%$, compared to the filter space. Second, the image space is less sensitive to noise, since using the filter space amplifies the noise in the observations. Third, the filter space solution ignores the important integrability constraint [10], whereas the image space solution does not suffer from this problem.

The proposed BID method, like many existing BID methods [5,8, 11,10,13,19,17], just to name a few, alternatively iterates between two major steps, image estimation with the current kernel estimate and kernel estimation with the current image estimate. The image estimation step requires to solve the large scale linear system (17) and update the weights ξ (18). By estimating the image and variational parameters jointly, the image estimation step is a standard nonblind deconvolution procedure.

The weights ξ (variational parameters) play an important role in sparse recovery problems. In fact, large values in ξ determine which pixels should be smoothed, and therefore help remove the blur, provided that the location is correct (consistent with the groundtruth).

4. Parameter analysis of Algorithm 1

In this section, we analyze the key parameters of Algorithm 1, such as the number of ADMM iterations, β_v, ϵ, and σ^2. Throughout this section and what follows, EM stands for the procedure described to estimate image, weights, and blur. Notice that we use ADMM to solve (17).

Regarding the number of ADMM iterations, for the Step 5 of Algorithm 1, we observe that 4 ADMM iterations with a small penalty weight ($\beta_v = 0.001$) is sufficient to find a good approximate solution to (17). However, it turns out that finding such an accurate solution to (17) often leads to poor local minima or saddle points in the sense of the objective function value. Instead of solving (17) fully, we solve (17) partially and use a large penalty weight β_v to find a *smooth approximate* solution to (17). Unfortunately, large β_v reduces the convergence speed [31,34]. We therefore use only one ADMM iteration and update the weights more frequently for acceleration [28]. In the performed experiments, we have observed that appropriate images to estimate the blur are those with many step edges and tiny sharp details while not suitable images to perform the same task are those which are not smooth and also contain ringing artifacts. Based on this observation, we use a large penalty weight β_v to enforce strong image smoothness during the optimization procedure, so that ringing or blur is less likely to emerge. In fact, the denominator of Eq. (22) indicates that large β_v makes the high frequency components of Λ_h well regularized, so that ringing is suppressed even if Λ_h is not accurate. We call this approach Smooth ADMM (SADMM) because it does not provide the exact EM solution to equation (17) but a smoother image estimate that can be used to better estimate the blur.

We should mention that one ADMM iteration for Step 5 cannot guarantee that the cost function drops monotonically. For this to happen, more ADMM iterations are required for Step 5. Interested readers are referred to Fig. 2(e) in [28], for a challenging nonconvex image reconstruction for which the cost function does not drop monotonically but still converges after thousands of iterations. Notice that we can not prove the convergence of Algorithm 1 with
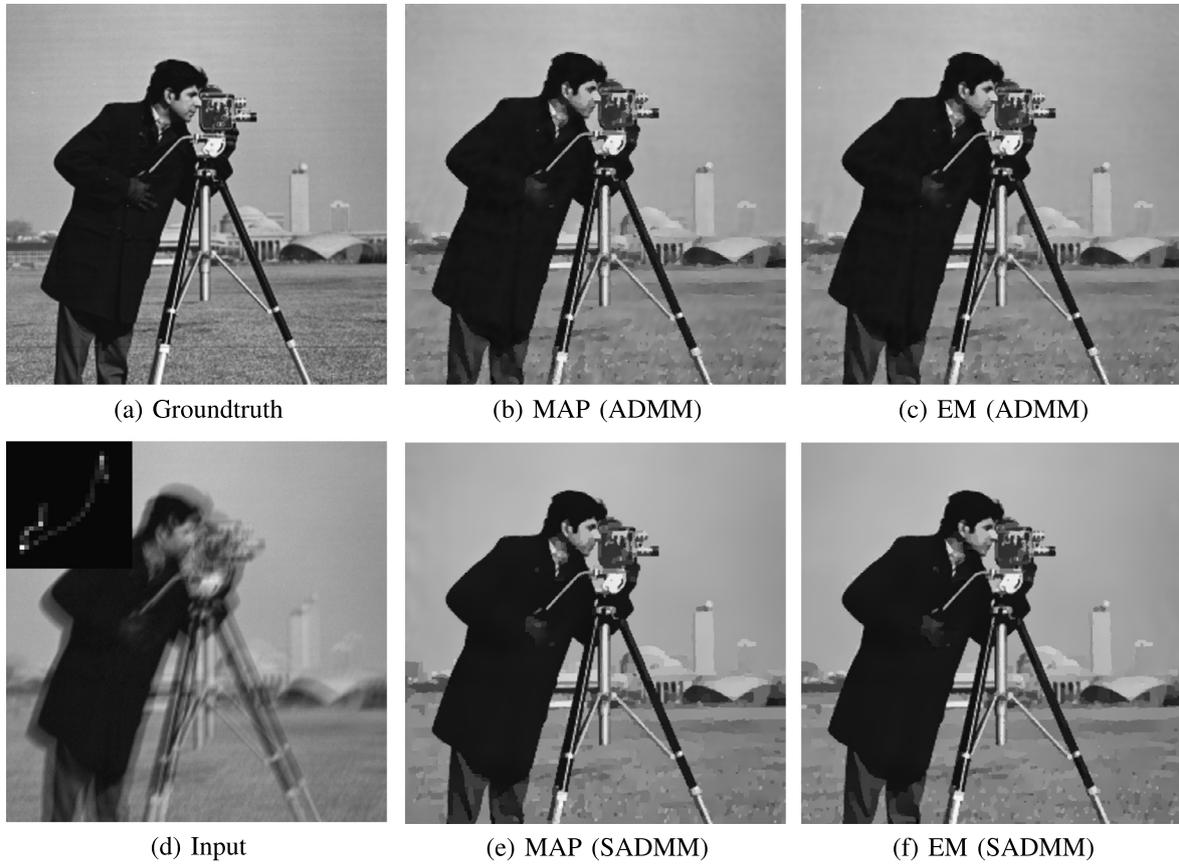
Fig. 2. SADMM vs. ADMM in nonblind deconvolution: restorations. (a) Groundtruth. (b) Restoration by MAP (ADMM). (c) Restoration by EM (ADMM). (d) Degraded image with groundtruth kernel, noise level $\sigma = 0.01$. (e) Restoration by MAP (SADMM). (f) Restoration by EM (SADMM).

one ADMM iteration, however we have not experienced any convergence problems in the performed experiments.

To demonstrate the advantages of SADMM (fewer ADMM iterations, 100×1, and a large β_v equal to 0.1) over EM (more ADMM iterations, 100×4, and a small β_v equal to 0.001), we adopt the log prior and test the SADMM and ADMM in nonblind image deconvolution using the synthetic degradation in Fig. 2(a), whose groundtruth image is shown in Fig. 2(d). We use the MAP cost function, that is, $F(x) = \frac{1}{2}\|Hx - y\|_2^2 + R(x)$, where $R(x) = \sigma^2 \sum_i \sum_{\gamma=1}^{L} \log_\epsilon (|x_\gamma(i)|)$ and $\log_\epsilon$ $(\epsilon = 0.001)$ is the Huber log function. Fig. 2(b) shows the recovered image by MAP (ADMM), which is not very smooth and contains a small amount of boundary artifacts. In contrast, the restoration by MAP (SADMM) (Fig. 2(e)) is smoother and with many sharp step edges, including the grass region. The cost functions, presented in Figs. 3(a) and (c), indicate that the proposed SADMM converges faster than the ADMM. It should be noted that ADMM needs 800 FFTs (including 400 IFFTs) to generate such a smooth restoration,[1] whereas SADMM requires only 200 FFTs (including 100 IFFTs). The same conclusions can also be drawn for the EM case, as shown in Figs. 2(c), 2(f), 3(b), and 3(d).

We now analyze the penalty weights for the ADMM procedure. Fig. 4 shows the impact of β_v on the SADMM results, for the input image in Fig. 2(a). It is clear that the recovered image becomes smoother as β_v increases. However, if β_v is too large, it can not recover sufficient (sharp) image edges and the convergence speed becomes slow (the image evolves slowly in each iteration),

as shown in Figs. 4(a) and (f), respectively. Fig. 4(c) suggests that $\beta_v \in [0.1, 1]$ is a good choice for Algorithm 1, since it can enforce adequately strong smoothness and also recover many sharp step edges. In short, β_v controls the image smoothness and convergence speed.

Regarding the robustness parameter ϵ used in the prior, we note that it plays an important role on the image sparsity as well as the numerical stability of the estimation procedure. The HSG prior favors an image of smaller sparsity than the SG prior, depending on the value of ϵ. From (10) and (18), it is clear that increasing ϵ promotes less sparsity, since the weights ξ will be upper bounded by a smaller value. Fig. 5 shows the restorations with 5 different values of ϵ ranging from 0 to 0.02. As we can see in Figs. 5(a)–(e), the strong edges are basically the same, however the weak edges are different. As ϵ increases, more weak edges appear and the image gets somewhat blurry, see the selected patches in Figs. 5(d) and (e). Figs. 5(a) and (b) are quite similar, suggesting that a small ϵ like 0.002 is enough to promote very strong sparsity for image deblurring. This figure shows that a small ϵ is crucial for a smooth and blur free restoration. However, as ϵ drops to zero, a small change in the image could cause a huge change in the value of the cost function. More importantly, as ϵ increases, the Lipschitz constant of $\log_\epsilon$ decreases and fewer iterations are required for convergence. Fig. 6 shows that with $\epsilon = 0$, the cost function does not reach the limit (not flat) in 100 iterations and is not as stable as the others. A tradeoff between image sparsity and numerical stability is provided for ϵ in the range $[0.001, 0.004]$.

The noise variance σ^2 controls the tradeoff between image details and smoothness. Increasing σ^2 leads to larger smoothness

---

[1] With 25 IRLS iterations, the output image is not so smooth and exhibits more (boundary) artifacts.
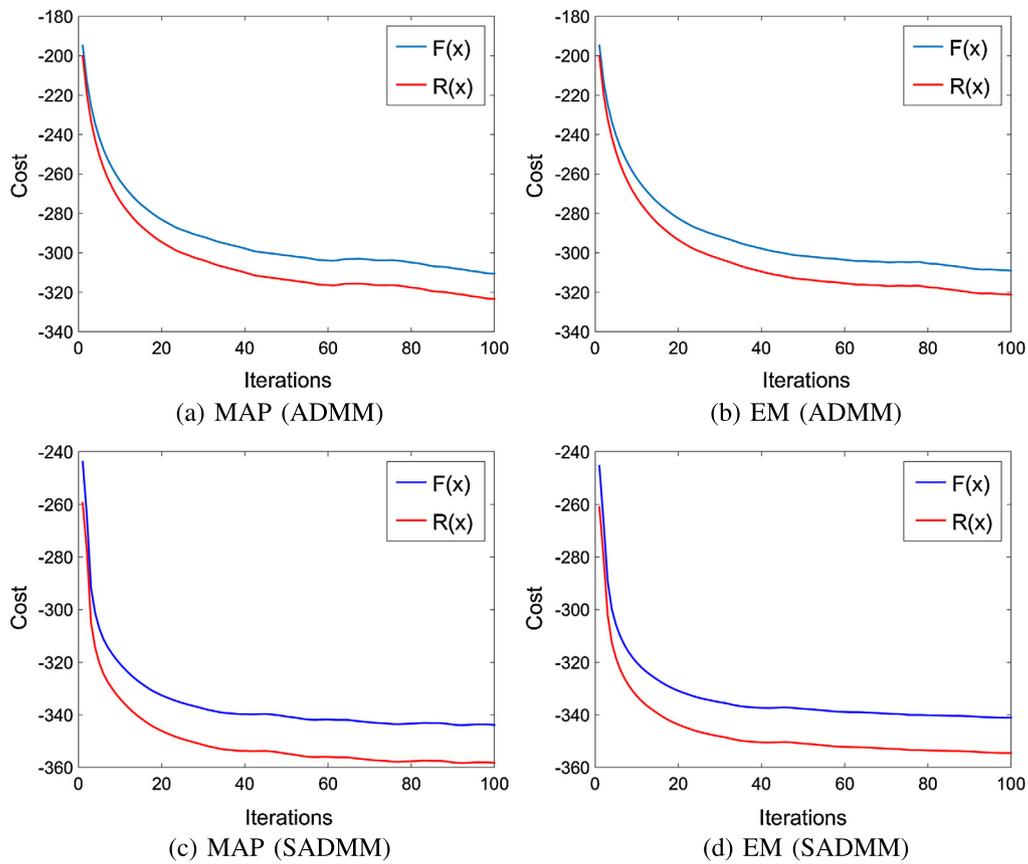
Fig. 3. SADMM vs. ADMM in nonblind deconvolutions: (a) Cost profile by MAP (ADMM); (b) Cost profile by EM (ADMM); (c) Cost profile by MAP (SADMM); (d) Cost profile by EM (SADMM).
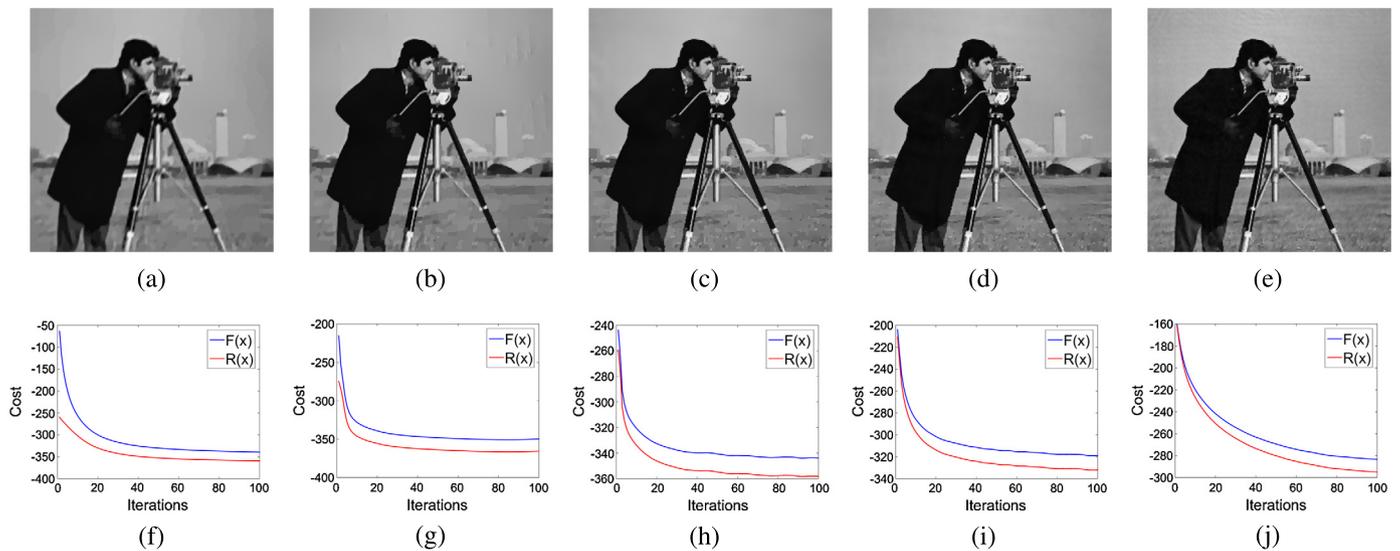

Fig. 4. Restorations with different β_v. (a) $\beta_v = 10$. (b) $\beta_v = 1$. (c) $\beta_v = 0.1$. (d) $\beta_v = 0.01$. (e) $\beta_v = 0.001$. (f)–(j) are the cost profiles of (a)–(e), respectively.

and fewer details. For blur estimation, it is helpful to start with a large σ^2 in which case the blur kernel is not accurate, and then gradually decrease it to the desirable value as the kernel estimate improves [10,13,27]. The rationale behind this strategy is that a very smooth image can facilitate the quick identification of the blur and hence speed up convergence. In our implementation, we adapt σ only in the first 5 iterations and then keep it fixed to 0.01, similarly to [27].

5. Experiments

We first use synthetic data to demonstrate that the proposed EM-SADMM approach is more robust than EM-ADMM, then we compare our EM-SADMM BID method with state-of-the-art methods, [13,14,25,27]. For simplicity, we denote the proposed BID methods by the form log+ADMM or log+SADMM, where log+ADMM means Huber log prior and ADMM optimiza-
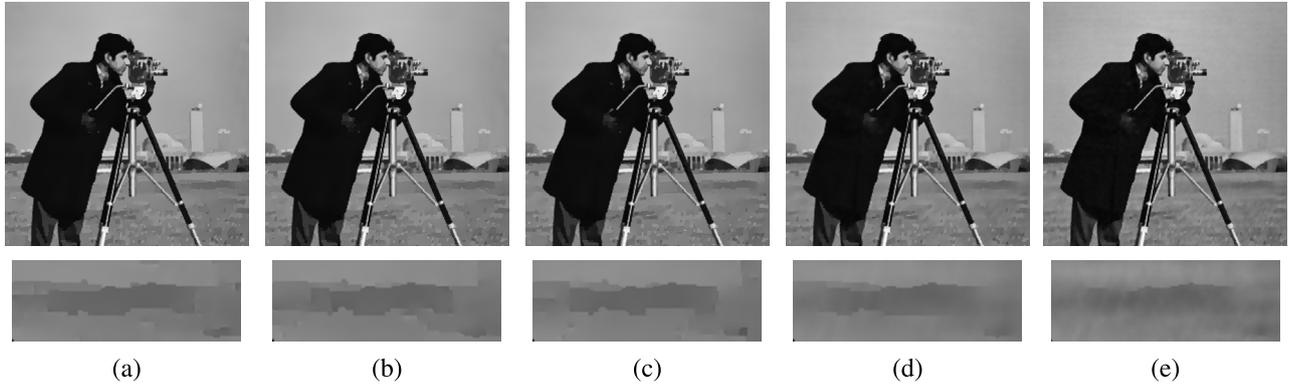
(a)　　　　　　　(b)　　　　　　　(c)　　　　　　　(d)　　　　　　　(e)

Fig. 5. Restorations with different ϵ: (a) $\epsilon = 0$; (b) $\epsilon = 0.002$; (c) $\epsilon = 0.004$; (d) $\epsilon = 0.01$; (e) $\epsilon = 0.02$. Increasing ϵ makes weak edges less steep and more blurry, as shown in close-up views in the 2nd row.
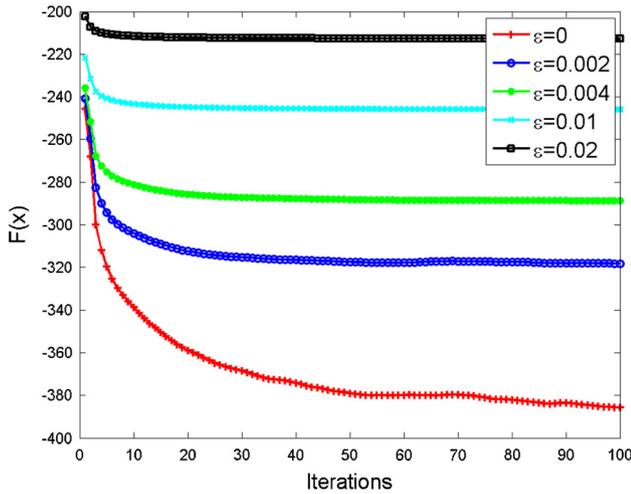

Fig. 6. Cost profiles with different values of ϵ. Increasing ϵ leads to better convergence and numerical stability.

tion scheme, and log+SADMM means Huber log prior and ADMM optimization scheme.

To estimate the blur kernel, only the first order derivative filters, $f_1 = [1, -1]$ and $f_2 = [1, -1]^T$, are used. For the final image reconstruction, we adopt not only the first order but also the second order derivative filters, i.e., $f_3 = [-1, 2, -1]$, $f_4 = [-1, 2, -1]^T$, $f_5 = [1 - 1; -1, 1]$. For fair comparison, we use our fast nonblind deconvolution algorithm[2] with the same parameters to recover the final image for all the BID methods, that is, to solve (31). All experiments are carried out using MATLAB on a PC with the Intel Core i7-4790 CPU @ 3.60 GHz. The code is available at https://sites.google.com/site/fbdhsgp/.

5.1. EM-SADMM vs. EM-ADMM in blind deconvolution

We first show that the proposed EM-SADMM approach outperforms the EM-ADMM approach in terms of restoration quality, speed, and objective function value. We choose the 5 standard images (all of size 512×512): Cameraman, Lena, Boat, Airplane, and Pepper, and use the 8 motion kernels in [35] to convolve each image and add 1% Gaussian noise as in [14]. Subsequently, we obtain a dataset of 40 degraded images. Table 1 shows the average PSNR (the highest PSNR after a shift operation, see Eq. (12) in [19]), SSIM [36], SSD (Sum of Squared Differences [35]) error, error ratio

(ratio between the SSD restoration error with the estimated kernel and the SSD restoration error with the groundtruth kernel), running time, regularization $R(\hat{x})$ (smaller indicates smoother), and objective $F(\hat{x})$ (smaller is better). It is clear that the EM-SADMM approach yields better results than the EM-ADMM approach for all figures of merit and also is about 1.5 times faster. In addition, it shows that $\epsilon = 0.004$ leads to the best performance in terms of the 4 figures of merits and running time. However, further increasing ϵ makes the overall performance worse.

The EM-SADMM approach leads to better BID performance than the EM-ADMM approach, mainly because the EM-SADMM approach better avoids undesirable saddle point and local minima than the EM-ADMM approach, as we have shown in Fig. 2. To further illustrate this point, we present the worst two EM-ADMM ($\epsilon = 0.001$) restorations in Fig. 7(b) and the corresponding image estimates $\hat{x}$ (see Eq. (17)) in Fig. 7(d). The input images are shown in Fig. 7(a). As we can see from Figs. 7(b) and (d), it is conceivable that the ringing artifacts and blur in Fig. 7(b) are indeed caused by the ringing artifacts and blur in the image estimates $\hat{x}$ in Fig. 7(d), which are typical bad local minima or saddle point since the cost function $F(x)$ is not small (see the caption in Fig. 7). Interestingly, by calculating $F(\hat{x}) - R(\hat{x})$, we emphasize that the image and kernel estimates obtained by log+ADMM better fit the observation data than the estimates obtained by log+SADMM, although the image and kernel estimates obtained by log+ADMM are less accurate. Take the *Boat* image in Fig. 7 as an example; the data fitting term for log+ADMM is about 2.41, smaller than 2.76, the data fitting term for log+SADMM. Therefore, we can conclude that the image and kernel estimates by log+ADMM in Fig. 7 are indeed poor stationary points. By enforcing more smoothness during the optimization process, the log+SADMM approach leads to smoother image estimates and hence is more likely to avoid ringing and blur, as shown in Figs. 7(c) and (e).

Table 2 shows the running time of the 5 methods to estimate three kernels of different support from a 512×512 image. As we can see, the proposed method log+SADMM with $\epsilon = 0.002$ is about 10–20 times faster than [10] and [13], depending on the kernel size. Thanks to the use of 2-D FFT, the running time of the proposed method does not increase much as the kernel gets larger, whereas both [10] and [13] become very time consuming as the kernel size increases. Due to the log prior intrinsic computation cost and the MATLAB implementation, the proposed algorithm is still much slower than the MAP method [17]. But, we argue that the BID method [17] is an "early stop" algorithm and can not achieve current state-of-the-art performance, as shown in [27]. The computational complexity of the proposed algorithm is $O(N_F N \log N)$, similar to [17], where N_F denotes the number of FFTs and IFFTs, since the computations for the image and kernel estimation are both dominated by FFTs and IFFTs. In contrast,

---

[2] Code available at http://decsai.ugr.es/vip/resources/VDBKE.html.

Table 1
Quantitative comparison on our test set of 40 images.

| Methods	PSNR	SSIM	SSD	Error ratio	Time	$R(\hat{x})$	$F(\hat{x})$
With known h	32.90	0.8668	105.76	1	–	–	–
$\log_{0.001}$+ADMM	30.05	0.8529	167.20	1.5825	25.21	−312.62	−310.03
$\log_{0.002}$+ADMM	30.04	0.8528	167.23	1.5832	25.61	−302.66	−300.07
$\log_{0.004}$+ADMM	30.09	0.8531	167.80	1.5914	25.51	−285.95	−283.37
$\log_{0.01}$+ADMM	30.00	0.8507	172.95	1.6334	25.39	−253.07	−250.52
$\log_{0.001}$+SADMM	30.10	0.8600	128.49	1.2049	16.64	−339.52	−336.72
$\log_{0.002}$+SADMM	30.14	0.8598	130.04	1.2235	16.12	−319.40	−316.61
$\log_{0.004}$+SADMM	30.24	0.8600	128.55	1.2034	14.80	−294.14	−291.38
$\log_{0.01}$+SADMM	30.08	0.8535	158.15	1.5040	9.88	−254.69	−252.01

(a)   (b)   (c)   (d)   (e)
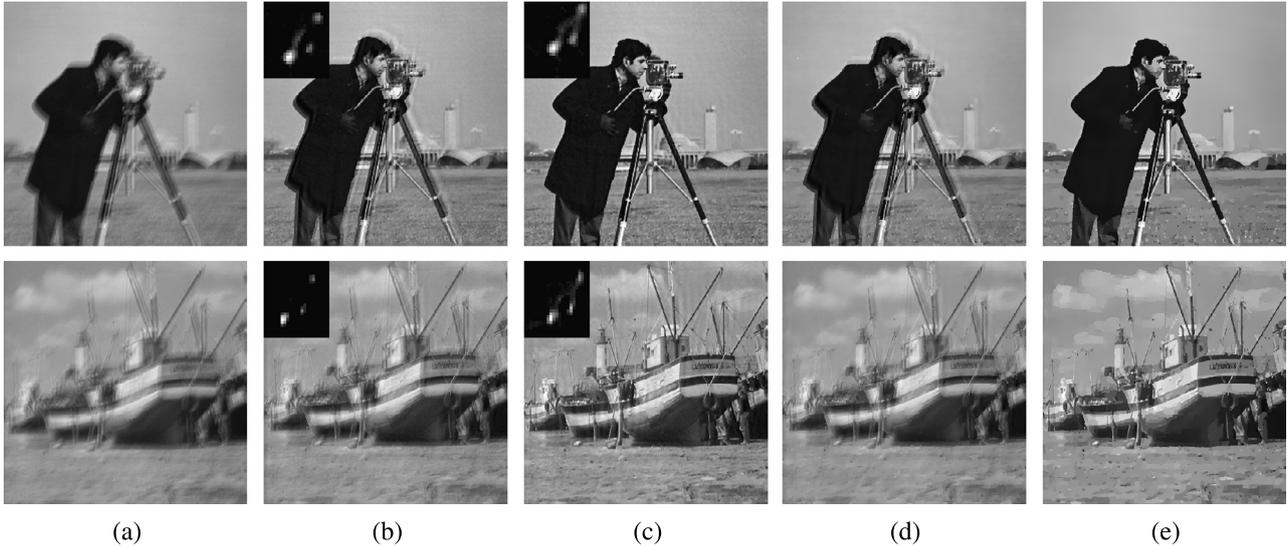
Fig. 7. The two worst deblurred results on the 40 dataset. (a) Input. (b) x^* (see Eq. (31)) with the kernel estimated by log+ADMM. (c) x^* with the kernel estimated by log+SADMM. (d) The estimate $\hat{x}$ by log+ADMM, see Eq. (17), Cameraman: $F(\hat{x}) = -318.35$, $R(\hat{x}) = -320.89$; Boat: $F(\hat{x}) = -295.14$, $R(\hat{x}) = -297.65$. (e) The estimate $\hat{x}$ by log+SADMM, Cameraman: $F(\hat{x}) = -352.34$, $R(\hat{x}) = -355.10$; Boat: $F(\hat{x}) = -331.61$, $R(\hat{x}) = -334.54$.

Table 2
Running time (in seconds) of different methods.

Kernel support	11×11	21×21	27×27
Babacan et al. [13] (MATLAB)	116.48	302.85	535.71
Levin et al. [10] (MATLAB)	111.75	240.72	425.58
Perrone&Favaro [25] (MATLAB)	2382.85	3447.20	3745.64
Xu et al. [17] (C++)	2.24	3.13	2.93
log+SADMM (MATLAB)	9.45	15.6	15.12

the computational complexity of the methods in [10,13,25] is of the form $O(N_C N K)$, with N_C being the number of 2-D spatial convolutions, since 2-D convolutions dominate the computations. Notice that we have used here K, the blur size, instead of the image size, N. In general, due to the use of ADMM, N_F can be 3–5 times smaller than N_C.

5.2. Comparison on the Sun et al. [14] dataset

We test the proposed methods on the dataset in [14], which contains 640 blurry and noisy images ($\sigma = 0.01$) generated from 80 natural images and 8 motion kernels. For fair comparison, we use the nonblind deconvolution method EPLL [37] to reconstruct the final restoration. Fig. 8 shows the success percent of 5 methods, including the VB methods (the proposed log+SADMM and Babacan et al. [13]) and the MAP methods, Sun et al. [14], Perrone and Favaro [25] and Zhou et al. [27]. We observe that a restoration with error ratio less than 2 is visually satisfying (without noticeable ringing artifacts and blur). Fig. 8 suggests that our method significantly outperforms the competitors, with over 75% excellent restorations (error ratio <1.5), 95% successful restorations (error
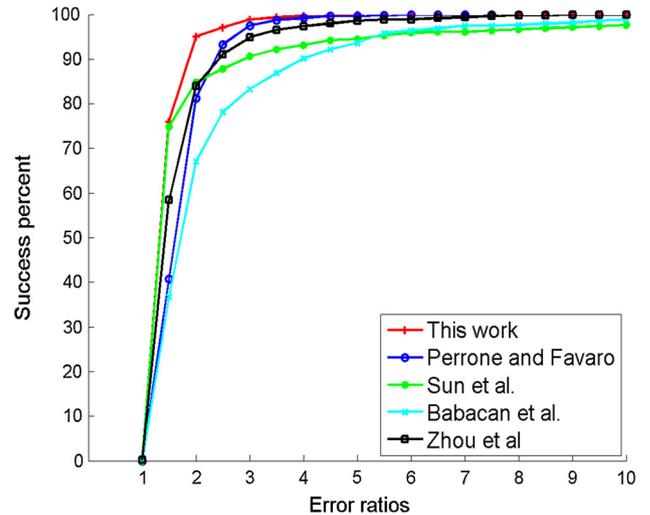

Fig. 8. Cumulative histograms of the error ratios across the dataset [14].

ratio <2), and 100% acceptable restorations (error ratio <5). Sun et al. [14] can also achieve about 75% excellent restorations, but is not so robust as Perrone and Favaro [25] and Zhou et al. [27], which have over 99% and 98% acceptable restorations, respectively. Table 3 shows the average SSD error and error ratio of those methods. It is clear that the proposed method takes the lead, with an average error ratio of 1.41 and SSD error of 590.

Fig. 9 shows 5 selected images for visual comparison, in which only image 5 contains adequate step edges and the rest lack step

Table 3
Average scores of different methods.

Methods	SSD error	Erorr ratio
With known h	392.15	1
log+SADMM	590.39	1.41
Perrone & Favaro [25]	635.27	1.71
Sun et al. [14]	618.92	2.01
Babacan et al. [13]	792.27	2.27
Zhou et al. [27]	686.29	1.66

edges. As we can see from Fig. 9, the step edge based method [14] works well for image 5, but fails for the other natural images. The algorithm by Babacan et al. [13] works well for image 5 but generates noticeable ringing artifacts for the rest. Our method significantly outperforms Babacan et al. [13] mainly because the proposed EM-SADMM approach better avoids poor stationary points

than the EM approach, which uses CG or ADMM to solve (17) fully and cannot guarantee a smooth image estimate as SADMM does. Perrone and Favaro [25] and Zhou et al. [27] yield good restorations but still exhibiting visible ringing artifacts or blur, and the estimated kernels are not as accurate as ours, see image 44 for an example. The proposed method shows superior performance for all these challenging images, without any visible ringing artifacts in the deblurred results.

5.3. Results on real data

Fig. 10(a) shows a real image with blur caused by camera shake [38]. Figs. 10(b)–(f) show the deblurred images and estimated kernels by various methods. As we can see in Fig. 10, noticeable ringing artifacts emerge in the results of [13] and Xu et al. [17], whereas fewer ringing artifacts appear in the results of Perrone and Favaro [25], Zhou et al. [27], and ours.
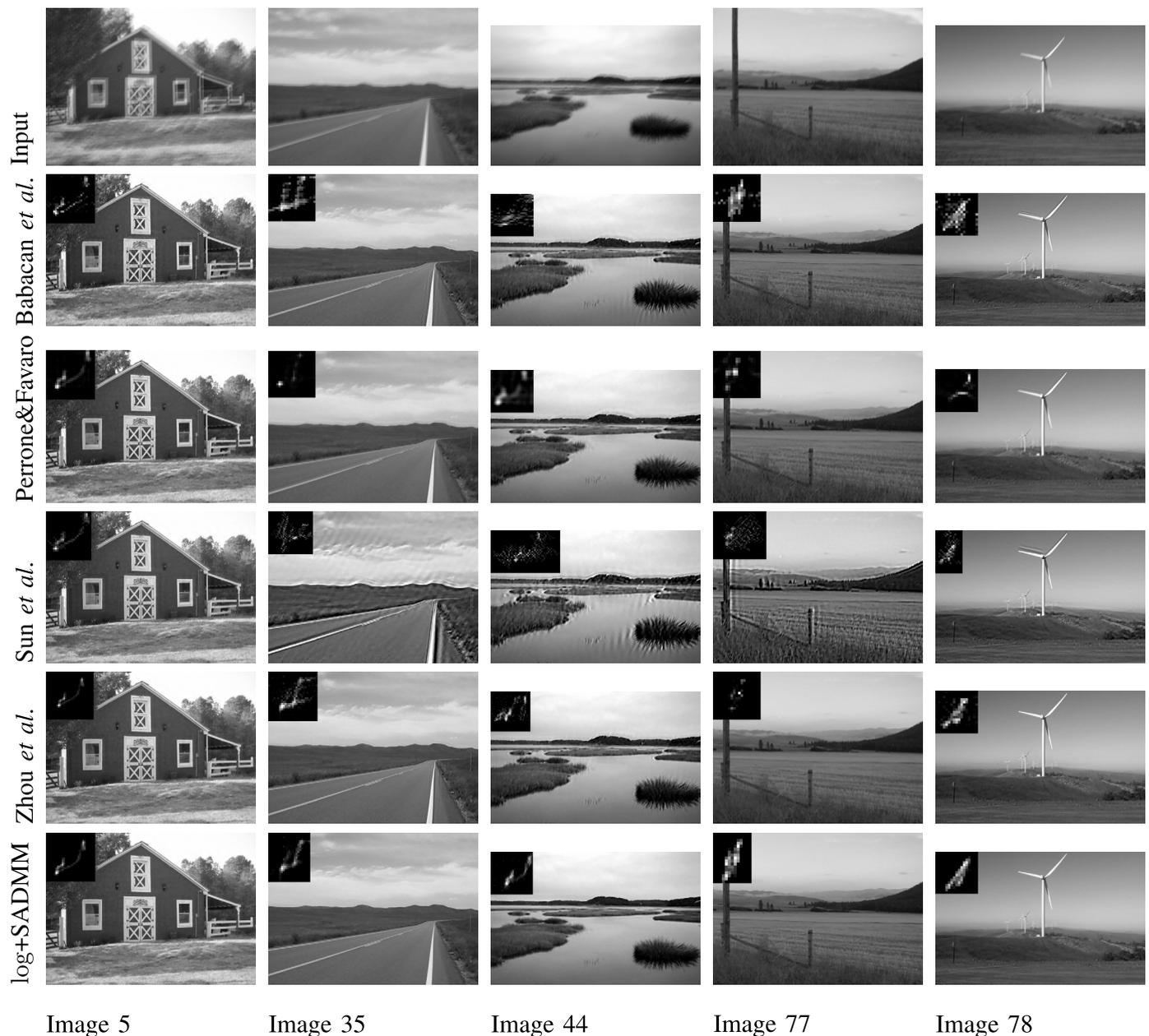

Fig. 9. Selected results on the dataset [14] for visual comparison. This figure is best viewed on the screen. Please zoom in for more details.
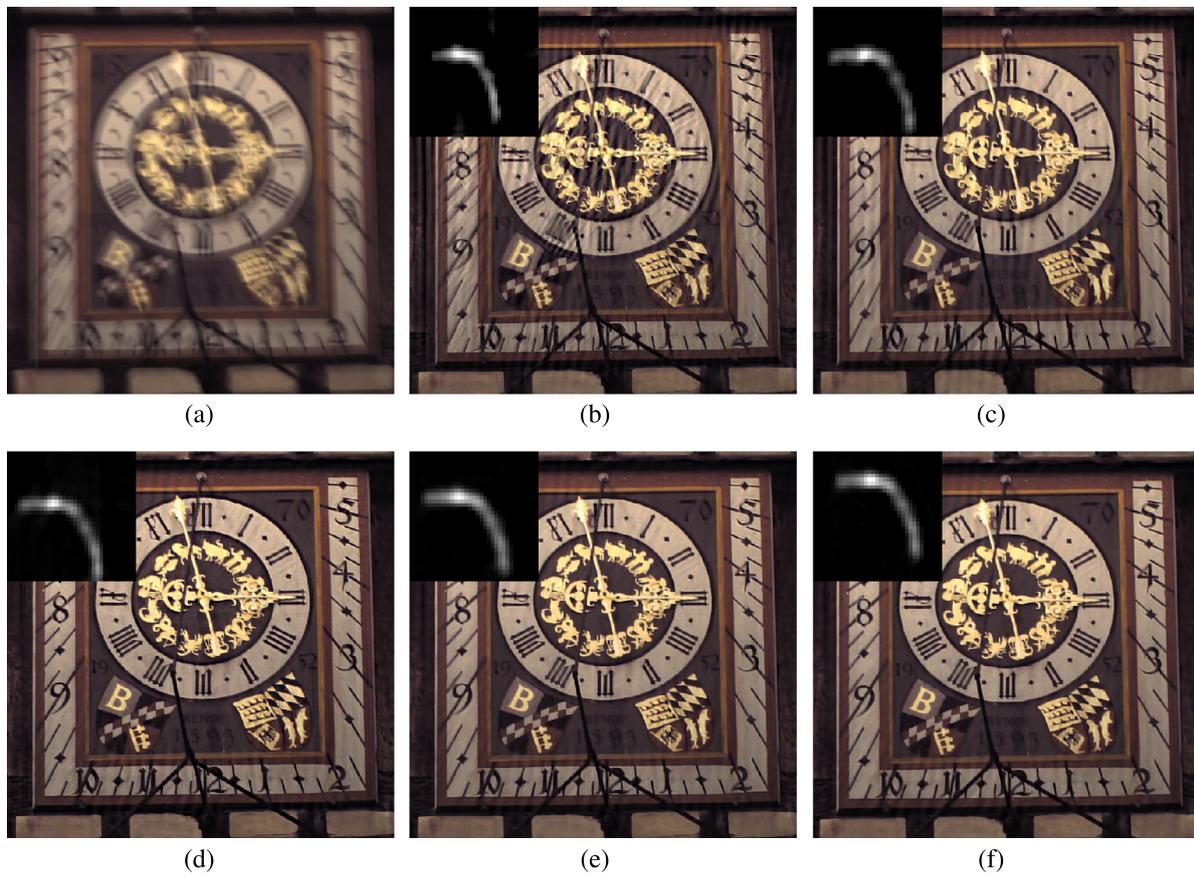
Fig. 10. Removing camera shake. (a) Input image, borrowed from [38]. (b) Babacan et al. [13]. (c) Xu et al. [17]. (d) Perrone and Favaro [25]. (e) Zhou et al. [27]. (f) The proposed log+SADMM.

6. Conclusion

In this paper, we have proposed first a new family of sparsity promoting priors, the so called Huber Super Gaussian priors, to address the numerical problem caused by the singularity at zero of some SG priors, and introduced an smooth optimization approach to deal with the poor stationary point problem caused by the use of nonconvex image priors. We show that the proposed EM-SADMM approach, which finds a smooth image mean rather than the exact image mean provided by the EM approach, better avoids undesirable saddle point and local minima. By making use of ADMM, a fast EM based BID method formulated in image space for these priors has been proposed. Experiments demonstrate that the proposed log+SADMM BID method significantly outperforms current state-of-the-art BID algorithms in terms of restoration quality and speed.

References

[1] M.R. Banham, A. Katsaggelos, Digital image restoration, IEEE Signal Process. Mag. 14 (2) (1997) 24–41.

[2] P. Ruiz, X. Zhou, J. Mateos, R. Molina, A.K. Katsaggelos, Variational Bayesian blind image deconvolution: a review, Digit. Signal Process. 47 (2015) 116–127.

[3] J. Miskin, D. MacKay, Ensemble learning for blind image separation and deconvolution, in: Advances in Independent Component Analysis, Springer, 2000.

[4] C. Likas, N. Galatsanos, A variational approach for Bayesian blind image deconvolution, IEEE Trans. Signal Process. 52 (8) (Aug. 2004) 2222–2233.

[5] R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, W.T. Freeman, Removing camera shake from a single photograph, ACM Trans. Graph. 25 (3) (2006).

[6] O. Shacham, O. Haik, Y. Yitzhaky, Blind restoration of atmospherically degraded images by automatic best step-edge detection, Pattern Recognit. Lett. 28 (15) (2007) 2094–2103.

[7] J.H. Money, S.H. Kang, Total variation minimizing blind deconvolution with shock filter reference, Image Vis. Comput. 26 (2) (2008) 302–314.

[8] S. Cho, S. Lee, Fast motion deblurring, ACM Trans. Graph. 28 (5) (2009).

[9] J. Cai, H. Ji, C. Liu, Z. Shen, Blind motion deblurring from a single image using sparse approximation, in: CVPR, IEEE, 2009.

[10] A. Levin, Y. Weiss, F. Durand, W.T. Freeman, Efficient marginal likelihood optimization in blind deconvolution, in: CVPR, IEEE, 2011.

[11] D. Krishnan, T. Tay, R. Fergus, Blind deconvolution using a normalized sparsity measure, in: CVPR, IEEE, 2011.

[12] A. Goldstein, R. Fattal, Blur-kernel estimation from spectral irregularities, in: ECCV, 2012.

[13] S. Babacan, R. Molina, M. Do, A. Katsaggelos, Bayesian blind deconvolution with general sparse image priors, in: ECCV, 2012.

[14] L. Sun, S. Cho, J. Wang, J. Hays, Edge-based blur kernel estimation using patch priors, in: ICCP, IEEE, 2013.

[15] L. Zhong, S. Cho, D. Metaxas, S. Paris, J. Wang, Handling noise in single image deblurring using directional filters, in: CVPR, IEEE, 2013.

[16] L. Xu, J.Y. Jia, Two-phase kernel estimation for robust motion deblurring, in: ECCV, 2010.

[17] L. Xu, S.C. Zheng, J.Y. Jia, Unnatural l0 sparse representation for natural image deblurring, in: CVPR, IEEE, 2013.

[18] M.S.C. Almeida, M.A.T. Figueiredo, Blind image deblurring with unknown boundaries using the alternating direction method of multipliers, in: ICIP, IEEE, 2013.

[19] X. Zhou, F. Zhou, X. Bai, Blind deconvolution using a nondimensional gaussianity measure, in: ICIP, IEEE, 2013.

[20] H. Zhang, D. Wipf, Non-uniform camera shake removal using a spatially-adaptive sparse penalty, in: NIPS, 2013.

[21] M. Vega, R. Molina, A. Katsaggelos, Parameter estimation in Bayesian blind deconvolution with super Gaussian image priors, in: EUSIPCO, 2014.

[22] R. Molina, J. Mateos, A. Katsaggelos, Blind deconvolution using a variational approach to parameter, image, and blur estimation, IEEE Trans. Image Process. 15 (12) (Dec. 2006) 3715–3727.

[23] D. Wipf, H. Zhang, Revisiting Bayesian blind deconvolution, J. Mach. Learn. Res. 15 (1) (2014) 3595–3634.

[24] S. Osher, L.I. Rudin, Feature-oriented image enhancement using shock filters, SIAM J. Numer. Anal. 27 (4) (1990) 919–940.

[25] D. Perrone, P. Favaro, A logarithmic image prior for blind deconvolution, Int. J. Comput. Vis. (2015) 1–14.

[26] T.F. Chan, C.-K. Wong, Total variation blind deconvolution, IEEE Trans. Image Process. 7 (3) (1998) 370–375.

[27] X. Zhou, J. Mateos, F. Zhou, R. Molina, A. Katsaggelos, Variational Dirichlet blur kernel estimation, IEEE Trans. Image Process. 24 (12) (Dec. 2015) 5127–5139.

[28] X. Zhou, R. Molina, F. Zhou, A. Katsaggelos, Fast iteratively reweighted least squares for l_p regularized image deconvolution and reconstruction, in: ICIP, IEEE, 2014.

[29] R.T. Rockafellar, Convex Analysis, Princeton University Press, 1996.

[30] M. Hintermüller, T. Wu, Nonconvex TV^q-models in image restoration: analysis and a trust-region regularization based superlinearly convergent solver, SIAM J. Imaging Sci. 6 (3) (2013) 1385–1415.

[31] Y.L. Wang, J.F. Yang, W.T. Yin, Y. Zhang, A new alternating minimization algorithm for total variation image reconstruction, SIAM J. Imaging Sci. 1 (3) (2008) 248–272.

[32] A. Levin, R. Fergus, F. Durand, W.T. Freeman, Image and depth from a conventional camera with a coded aperture, ACM Trans. Graph. 26 (3) (2007).

[33] X. Zhou, F. Zhou, X. Bai, B. Xue, A boundary condition based deconvolution framework for image deblurring, J. Comput. Appl. Math. 261 (2014) 14–29.

[34] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (2011) 1–122.

[35] A. Levin, Y. Weiss, F. Durand, W.T. Freeman, Understanding and evaluating blind deconvolution algorithms, in: CVPR, IEEE, 2009.

[36] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

[37] D. Zoran, Y. Weiss, From learning models of natural image patches to whole image restoration, in: ICCV, IEEE, 2011.

[38] R. Köhler, M. Hirsch, B. Mohler, B. Schölkopf, S. Harmeling, Recording and playback of camera shake: benchmarking blind deconvolution with a real-world database, in: ECCV, 2012.
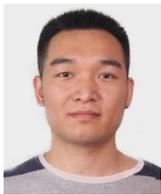
Xu Zhou received the B.S. degree in applied mathematics from the Central South University in 2009, and the Ph.D. degree in pattern recognition and intelligent system from the Beihang University, in 2016. From 2013 to 2014, he was a visiting student of the Visual Information Processing group, at the Department of Computer Science and Artificial Intelligence, University of Granada. He is currently a postdoctoral fellow with the School of Information Science and Technology, ShanghaiTech University. His research interests include image restoration and machine learning.

Miguel Vega was born in 1956, in Spain. He received the Bachelor Physics degree from the Universidad de Granada, Granada, Spain, in 1979, and the Ph.D. degree from the Universidad de Granada, in 1984. He is with the E.T.S. Ing. Informática of the Universidad de Granada (Departamento de Lenguajes y Sistemas Informáticos), where he is an Associate Professor and teaches software engineering. His research focuses on image processing (multichannel and super-resolution image reconstruction). He has collaborated at several projects from the Spanish Research Council.

Fugen Zhou received the B.S. degree in electronic engineering, M.S. degree and Ph.D. degree in pattern recognition and intelligent system from Beihang University, Beijing, China, in 1986, 1989, and 2006, respectively.

He joined the Image Processing Center at Beihang University in 1989. Since 2007, he has been a Professor with the School of Space, Beihang University. In 2004, as a Visiting Scholar, he cooperated with the research Institute of William Beaumont Hospital, USA. He was the Director of Image Processing Center at Beihang University (2000–2015). He currently serves as Associate Director of Medical Imaging Committee, Chinese Society of Medical Physics and a Standing Member of Chinese Society for Optical Engineering. One of his Ph.D. students, under his supervision, has been awarded the best doctoral thesis of BUAA in 2010 and the nomination prize of the national best doctoral thesis of China in 2011. His research interests include target detection and recognition, multi-modality image processing, biomedical image recognition. He has authored or coauthored over 70 research papers.

Rafael Molina was born in 1957. He received the degree in mathematics (statistics) and the Ph.D. degree in optimal design in linear models from the University of Granada, Granada, Spain, in 1979 and 1983, respectively. He became Professor of Computer Science and Artificial Intelligence at the University of Granada, Granada, Spain, in 2000. He is the former Dean of the Computer Engineering School at the University of Granada (1992–2002) and Head of the Computer Science and Artificial Intelligence department of the University of Granada (2005–2007). His research interest focuses mainly on using Bayesian modeling and inference in problems like image restoration (applications to astronomy and medicine), superresolution of images and video, blind deconvolution, computational photography, source recovery in medicine, compressive sensing, low-rank matrix decomposition, active learning, fusion, and classification.

Prof. Molina serves as an Associate Editor of Applied Signal Processing (2005–2007); the IEEE Transactions on Image Processing (2010–present); and Progress in Artificial Intelligence (2011–present); and an Area Editor of Digital Signal Processing (2011–present). He is the recipient of an IEEE International Conference on Image Processing Paper Award (2007) and an ISPA Best Paper Award (2009). He is a coauthor of a paper awarded the runner-up prize at Reception for early-stage researchers at the House of Commons.

Aggelos K. Katsaggelos received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Greece, in 1979, and the M.S. and Ph.D. degrees in Electrical Engineering from the Georgia Institute of Technology, in 1981 and 1985, respectively. In 1985, he joined the Department of Electrical Engineering and Computer Science at Northwestern University, where he is currently a Professor holder of the Joseph Cummings chair. He was previously the holder of the Ameritech Chair of Information Technology and the AT&T chair. He is also a member of the Academic Staff, NorthShore University Health System, an affiliated faculty at the Department of Linguistics and he has an appointment with the Argonne National Laboratory. He has published extensively in the areas of multimedia signal processing and communications (over 230 journal papers, 500 conference papers and 40 book chapters) and he is the holder of 25 international patents. He is the co-author of Rate-Distortion Based Video Compression (Kluwer, 1997), Super-Resolution for Images and Video (Claypool, 2007), Joint Source-Channel Video Transmission (Claypool, 2007), and Machine Learning Refined (Cambridge University Press, 2016). He has supervised 54 Ph.D. theses so far.

Prof. Katsaggelos was the Editor-in-Chief of the IEEE Signal Processing Magazine (1997–2002), a BOG Member of the IEEE Signal Processing Society (1999–2001), a member of the Publication Board of the IEEE Proceedings (2003–2007), and he is currently a Member of the Award Board of the IEEE Signal Processing Society. He is a Fellow of the IEEE (1998) and SPIE (2009) and the recipient of the IEEE Third Millennium Medal (2000), the IEEE Signal Processing Society Meritorious Service Award (2001), the IEEE Signal Processing Society Technical Achievement Award (2010), an IEEE Signal Processing Society Best Paper Award (2001), an IEEE ICME Paper Award (2006), an IEEE ICIP Paper Award (2007), an ISPA Paper Award (2009), and a EUSIPCO paper award (2013). He was a Distinguished Lecturer of the IEEE Signal Processing Society (2007–2008).