Variational Dirichlet Blur Kernel Estimation

Xu Zhou, *Student Member, IEEE,* Javier Mateos, *Member, IEEE,* Fugen Zhou, *Member, IEEE,* Rafael Molina, *Member, IEEE,* and Aggelos K. Katsaggelos *Fellow, IEEE*

***Abstract*—Blind image deconvolution involves two key objectives, latent image and blur estimation. For latent image estimation, we propose a fast deconvolution algorithm, which uses an image prior of nondimensional Gaussianity measure to enforce sparsity and an undetermined boundary condition methodology to reduce boundary artifacts. For blur estimation, a linear inverse problem with normalization and nonnegative constraints must be solved. However, the normalization constraint is ignored in many blind image deblurring methods, mainly because it makes the problem less tractable. In this paper, we show that the normalization constraint can be very naturally incorporated into the estimation process by using a Dirichlet distribution to approximate the posterior distribution of the blur. Making use of variational Dirichlet approximation, we provide a blur posterior approximation that takes into account the uncertainty of the estimate and removes noise in the estimated kernel. Experiments with synthetic and real data demonstrate that the proposed method is very competitive to state-of-the-art blind image restoration methods.**

***Index Terms*—Blind Deconvolution, image deblurring, variational distribution approximations, Dirichlet distribution, constrained optimization, point spread function, inverse problem.**

I. Introduction

BLIND image deconvolution (BID) refers to the problem of recovering the source image from a degraded observation when the blur kernel is unknown, using partial information about the imaging system [1]. Primary applications of BID include astronomical imaging, medical imaging, and computational photography [2].

Mathematically, the degraded observation $y \in \Re^M$ can be approximately modeled as [2]

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

where $x \in \Re^N$ is the original image, H of size $M \times N$ is the convolution matrix whose row elements are obtained from the blur kernel $h \in \Re^K$, and $n \in \Re^M$ is assumed to be additive zero-mean Gaussian white noise. BID is a severely ill-posed inverse problem, since multiple pairs of kernels and original images can match the model (1) equally well. To overcome the ill-posed nature of BID, additional constraints or assumptions on both x and h must be introduced.

Many earlier works assume that the blur kernel follows a parametric model [1] [2] or satisfies some additional constraints, *e.g.*, centrosymmetric, nonnegative and normalization constraint [3] [4]. For the latent image, many BID methods [5]–[23] utilize image sparsity to estimate the image.

Fergus *et al*. [5] assume that the gradient of the latent image ∇x obeys a heavy-tailed distribution, meaning that most elements of ∇x are zero or very small but a number of elements are quite large. Specifically, they use a Mixture of Gaussians (MoG) to approximate a heavy-tailed distribution. However, as pointed out in [5], [24], and [17], using the maximum *a posteriori* (MAP) approach often yields delta kernels because the cost function favors a blurry explanation over sharp reconstructions. To avoid this problem, the Variational Bayesian (VB) approach [25] has been adopted in [5], see also [26] [27] [17]. Using the VB approach, a general framework based on the Super Gaussian priors is proposed by Babacan *et al*. [17]. Instead of using MAP estimation, Levin *et al*. [24] [11] suggest a MAP_h approach, namely marginalizing over latent images and then estimating the kernel alone. The rationale behind MAP_h is that estimating h alone is much better conditioned than estimating h and x together. Recently, following the MAP_h approach and using a majorization-minimization approach, Zhang and Wipf [23] propose a parameter free BID method for removing camera shake, in which the gradient image is modeled by an independent Gaussian distribution with zero mean and spatially varying variance. Interested readers are referred to [28] for a review of of Bayesian blind deconvolution methods.

An alternative to enforce image sparsity in the transform domain is to use the shock filter [29], see [7] and [9]. In fact, the shock filter not only promotes image sparsity, but also predicts step edges. Following this idea, Xu and Jia [14] point out that not all predicted step edges are helpful. Therefore, they suggest an edge selection strategy for removing texture edges, which improves the result significantly. Instead of predicting step edges, a probably better idea is to model step edges as local minima of a sparsity measure, *e.g.*, $\|\nabla x\|_1/\|\nabla x\|_2$, $\|\nabla x\|_0$, $\sum_i |\nabla x(i)|/(|\nabla x(i)|+\|\nabla x\|_1/N)$, see [12], [15] and [22] for more details. Sun *et al*. [19] also propose a novel parameterized patch model to regularize step edges, using learned image statistics or synthetic structures. Making use of directional filters, Zhong *et al*. [20] show that a good kernel can be estimated from a blurred image with $1\% - 10\%$ noise. Step edge based methods are robust to noise since noise and small gradients are smoothed out, but may fail if the image is dominated by textures. By combining step edge prediction [9]

This work was sponsored in part by National Natural Science Foundation of China (61233005), Ministerio de Ciencia e Innovación under Contract TIN2013-43880-R, the European Regional Development Fund (FEDER), the CEI BioTic with the Universidad de Granada and the Department of Energy grant DE-NA0002520.

X. Zhou and F. Zhou are with the Image Processing Center, Beihang University, Beijing 100191, China (e-mail: xuzhou@buaa.edu.cn; zhfugen@buaa.edu.cn). X. Zhou performed the work while at Universidad de Granada, as a joint-PhD student sponsored by Chinese Scholarship Council.

J. Mateos and R. Molina are with Departamento de Ciencias de la Computación e I.A., Universidad de Granada, Granada 18071, Spain (e-mail: jmd@decsai.ugr.es; rms@decsai.ugr.es).

A. Katsaggelos is with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208-3118 USA (e-mail: aggk@eecs.northwestern.edu).

with covariance image [11] used in kernel estimation, Wang *et al* [18] propose a robust BID method.

In this work, a BID algorithm that alternates between the estimation of image and blur is proposed. For the latent image estimation, we use the Nondimensional Gaussianity Measure (NGM) proposed in [22], to promote image sparsity. As the optimization algorithm of [22] does not make use of the FFT, we propose a fast deconvolution algorithm with undetermined boundary condition [30] to estimate the latent sharp image, which uses the variable splitting [31] and mask decoupling [32] [33] approaches. For the blur estimation part we propose a Variational Dirichlet (VD) method. This changes the regularization based approach to a variational Bayes approximation to estimate the posterior distribution of the blur. The benefit of using VD is that the resulting optimization problem does not have any equality constraint but lower bound constraints only, so that it can be efficiently solved by the gradient projection method [34]. The other benefit of VD is a new adaptive sparsity promoting term associated with the image, which is helpful to remove the kernel noise. Furthermore as we will later show, the VD approach can be applied to regularization and variational based BID methods that alternate between image and blur estimation, where the blur energy function (either from regularization or blur prior) is quadratic. This is the case of many state-of-the-art BID methods. In summary, the main contributions of this paper include a novel efficient optimization algorithm in the image space for NGM based deconvolution and a novel VD approach for kernel estimation.

This paper is organized as follows. In Section II, we introduce the framework to be used in BID. In Section III, we propose a fast nonblind image deconvolution algorithm formulated in the image domain for the NGM model in [22]. Section IV contains our blur kernel estimation algorithm. We derive the VD approximation and show the connection between VD approximation and MAP estimation. An efficient gradient projection algorithm is proposed to solve the resulting optimization problem. This section also discusses the applications of the proposed VD in kernel estimation when the blur prior is quadratic. A multiscale implementation of the VD based BID method is presented in Section IV-D, followed by a discussion on parameter setting and implementation in Section IV-E. In Section V, we compare our BID method with state-of-the-art BID methods. Experimental results on synthetic and real data show that the proposed method is very competitive in comparison with state-of-the-art BID algorithms [15], [17], [19], [20]. Finally, Section VI concludes this paper.

II. Regularization Based Blind Image Deconvolution

Regularization based BID methods, [3], [4], [7], [10], [21], [26], [27], just to name a few, alternatively solve the following two optimization problems,

$$x^{k+1} = \arg\min_x \frac{1}{2}\|H^k x - y\|_2^2 + \lambda_x R_x(x), \tag{2}$$

$$h^{k+1} = \arg\min_h \frac{1}{2}\|X^{k+1} h - y\|_2^2 + \frac{\lambda_h}{2} R_h(h), \tag{3}$$

$$\text{subject to} \quad h \geq 0 \tag{4}$$

$$\sum_j h(j) = 1 \tag{5}$$

where k denotes iteration index, H^k is the convolution matrix formed by the estimated impulse response of the blur at the k-th iteration step h^k, X^{k+1} is the convolution matrix formed by the restored image x^{k+1}, R_x and R_h are regularization functions, $\lambda_x > 0$ and $\lambda_h \geq 0$ are penalty weights controlling the tradeoff between the data fitting term and regularization.

Recent published papers, see for instance [9], [11], [15], [17] have shown, however, that using gradient images leads to better kernel estimation. Furthermore, the final output image of (2) is often highly smoothed mainly due to the large penalty weight λ_x, which is not a desirable result. After kernel estimation, to reconstruct the final sharp image a nonblind image deconvolution method is required.

In this work we replace the blur estimation problem defined in (3)-(5) by the filter space formulation

$$h^{k+1} = \arg\min_h \sum_i \frac{1}{2}\|\nabla_i X^{k+1} h - \nabla_i y\|_2^2 + \frac{\lambda_h}{2} R_h(h), \tag{6}$$

$$\text{subject to} \quad h \geq 0 \tag{7}$$

$$\sum_j h(j) = 1 \tag{8}$$

where now $\nabla_i X^{k+1}$ is the matrix formed by the gradient of the image x^{k+1} utilizing the i-th filter (details on the filters used will be provided later).

Note that we are formulating the image estimation in the spatial domain and the blur estimation in the gradient domain. Xu *et al.* [15] indicate that using this spatial/filter formulation is better than using the same space formulation. One of the merits of using the gradient domain for kernel estimation is that the boundary artifacts are reduced significantly since most of the gradients at the boundaries are zero. This spatial/filter formulation prevents us from using the same cost function to be minimized on the image and blur and consequently makes it harder to establish the convergence of the iterative method, however, as we will show in the experimental section this dual formulation produces better image and blur estimates.

The above framework has two steps, namely, the latent image estimation step according to (2) with a given h, and the kernel estimation step according to (6)-(8) with a given x. By alternatively solving (2) and (6)-(8), one obtains a good estimation of the kernel, provided that a good image regularization $R_x(x)$ and a good kernel regularization $R_h(h)$ are used. It should be noted that both $R_x(x)$ and $R_h(h)$ might evolve during the iterations (*e.g.*, edge-predicting based method [9] and VB method [17]).

When the uncertainty of the image and blur are not taken into account and $R_x(x)$ and $R_h(h)$ do not incorporate information on the covariance of x and h respectively, this iterative framework is also known as MAP approach in the Bayesian formulation. It may not work well depending on the

initialization of the kernel, images priors/regularizations and kernel priors/regularizations. As shown in Levin *et al.* [24], the MAP approach may suffer from the delta kernel in the case of l_p regularization ($R_x(x) = \sum_j |\nabla x(j)|^p$ with $p \in (0, 2]$), since the cost function favors no-blur explanation (h is the delta kernel and $x = y$) over other solutions.

III. Fast Image Deconvolution using NGM

Let us now proceed with the solution of problem (2), that is, the image estimation part. Let us assume that we have a blur estimation and face the problem of estimating the latent image from the blurred observation.

In [22], to enforce sparsity, we proposed a Nondimensional Gaussianity Measure (NGM), $\sum_{i=1}^{2}\sum_j |\nabla_i x(j)|/(|\nabla_i x(j)| + E[|\nabla_i x|])$, where $i = 1$ and 2 denote horizontal and vertical filtering respectively and $E[|\nabla_i x|]$ denotes the mean value of the vector $|\nabla_i x|$, see [22] for details.

Decreasing NGM concentrates the energy at a small number of elements with most of them being zero or very small values. In other words, NGM favors sharp step edges over blurry ones. An example of a 3-D NGM is shown in Fig. 1.

For a given kernel h, we propose here to estimate a sharp image as the minimizer of

$$\min_x \frac{1}{2}\|Hx - y\|_2^2 + \lambda_x \sum_{i=1}^{2}\sum_j \frac{|\nabla_i x(j)|}{|\nabla_i x(j)| + E[|\nabla_i x|]} \quad (9)$$

where $\lambda_x > 0$ is a penalty weight that controls the degree of sparsity. This model differs from the one in [22], since the image estimation step in [22] (Eq. (7)) is formulated in the gradient domain.

To mitigate boundary artifacts, we adopt here undetermined boundary conditions [30] (see [35] for other boundary conditions). That is to say, $H = M_p T_h$, where M_p is an $M \times N$ matrix that extracts the Field of View (FOV) of image x and T_h is an $N \times N$ circulant matrix that performs circular convolution with kernel h (see [32], [33] and [35] for more details).

Making use of variable splitting [31] and mask decoupling [32], [33], we show how (9) can be solved efficiently. As in [32], [33] and [31], we introduce two auxiliary variables $u = T_h x$ and $v = \nabla x$ (specifically, $v_i = \nabla_i x$). Using the augmented Lagrangian and penalty approach for the constraints $u = T_h x$ and $v = \nabla x$, respectively, we obtain the unconstrained problem

$$\begin{aligned}\min_{x,u,v} &\frac{1}{2}\|M_p u - y\|_2^2 + \frac{\lambda_u}{2}\|u - T_h x - du\|_2^2 \\ &+ \sum_{i=1}^{2}\frac{\lambda_v}{2}\|v_i - \nabla_i x\|_2^2 + \lambda_x \sum_j \frac{|v_i(j)|}{|v_i(j)| + E[|v_i|]}, \quad (10)\end{aligned}$$

where λ_u and λ_v are the penalty weights and du is the Lagrangian multiplier. To obtain a good approximate solution to problem (9), λ_v has to be very large. In practice, a large penalty weight makes the algorithm very inefficient, so a continuous strategy, like the one in [31], is adopted for the penalty weight λ_v. We use an alternative minimization scheme, namely minimizing over one variable with the other two fixed. Consequently, we have to solve three subproblems as follows.

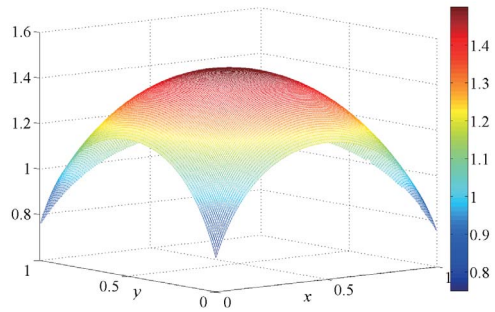

Fig. 1. 3-D $NGM = \frac{x}{x+1/3} + \frac{y}{y+1/3} + \frac{1-x-y}{1-x-y+1/3}$. It is clear that minimizing NGM pushes $(x, y, 1-x-y)$ away from $(1/3, 1/3, 1/3)$ and makes it more sparse, namely the third component increases while the other two decrease.

A. u-subproblem

Given x, du, and v, we need to solve

$$\min_u \frac{1}{2}\|M_p u - y\|_2^2 + \frac{\lambda_u}{2}\|u - T_h x - du\|_2^2, \quad (11)$$

Clearly, problem (11) has a closed form solution, given by

$$u = (M_p^T M_p + \lambda_u I)^{-1}(M_p^T y + \lambda_u(T_h x + du)), \quad (12)$$

where M_p^T pads y to an image of size N with zeros in both dimensions. Note that $M_p^T M_p$ is a diagonal matrix and hence, the inversion $(M_p^T M_p + \lambda_u I)^{-1}$ is trivial.

After updating u, we update du by

$$du = du + T_h x - u. \quad (13)$$

B. x-subproblem

Given u, du and v, we need to solve

$$\min_x \frac{\lambda_u}{2}\|u - T_h x - du\|_2^2 + \sum_{i=1}^{2}\frac{\lambda_v}{2}\|v_i - \nabla_i x\|_2^2. \quad (14)$$

Again, this problem also has a closed form solution. Assuming periodic boundary conditions for ∇_i, the solution can be efficiently computed with the use of a 2-D FFT.

C. v-subproblem

Given x, du, and u, we need to solve

$$\min_{v_i} = \frac{\lambda_v}{2}\|v_i - \nabla_i x\|_2^2 + \lambda_x \sum_j \frac{|v_i(j)|}{|v_i(j)| + E[|v_i|]}, \quad (15)$$

where $i = 1, 2$. The above cost function is nonconvex and inseparable. However, if we fix $E[|v_i|]$ from the previous iteration, it can be simplified as a single variable minimization problem, given by

$$\min_z g(z) = \frac{1}{2}\|z - w\|_2^2 + \lambda\frac{|z|}{|z| + E}, \quad (16)$$

where $\lambda = \frac{\lambda_x}{\lambda_v}$, $E = E[|v_i|]$ and w represents a component of $\nabla_i x$. Note that the minimizer z^* must have the same sign as w,

Algorithm 1 Fast NGM Deconvolution

Require: y, x^0, λ_x, λ_u, λ_v.
precompute the *Fourier* transform of h, ∇_1 and ∇_2
$u = M_p^T y$, $x = x^0$, $du = 0$
repeat
  update u using (12)
  update du using (13)
  $v_i = \nabla_i x$, $s_i = \text{sign}(\nabla_i x)$, $i = 1, 2$
  for $j = 1$ to 2 **do**
    $E_i = \text{mean}(|v_i|)$, $i = 1, 2$
    $v_i = s_i \max(|\nabla_i x| - \frac{\lambda_x E_i}{\lambda_v (|v_i| + E_i)^2}, 0)$, $i = 1, 2$
  end for
  $\lambda_v = \min(\lambda_v \sqrt{2}, 1)$
  update x by solving (14) in *Fourier* domain
until convergence
return x

since replacing z by $z_s = \text{sign}(w)|z|$ we have $g(z) \geq g(z_s)$. Let us examine the case $w \geq 0$, we aim at finding

$$\min_{z \geq 0} g(z) = \frac{1}{2}\|z - w\|_2^2 + \lambda \frac{z}{z + E}. \tag{17}$$

Letting $g'(z) = 0$, we have

$$z = \max(w - \frac{\lambda E}{(z + E)^2}, 0) \tag{18}$$

This equation implies a fixed point iteration formula $z^{t+1} = \max(w - \lambda E/(z^t + E)^2, 0)$. Together with the case $w < 0$, we have the ultimate fixed point iteration for (16)

$$z^{t+1} = \text{sign}(w) \max(|w| - \frac{\lambda E}{(|z^t| + E)^2}, 0), t = 0, 1, 2... \tag{19}$$

The iterative formula (19) implies $|z^*| \leq |w|$, which guides us to use $z^0 = w$. Noting that $|w| - \lambda E/(|z| + E)^2$ is monotonically decreasing as $|z|$ decreases, it can be shown that the sequence $\{z^t\}$ generated by (19) with the initialization $z^0 = w$ is also monotonic, and it converges to a stationary point of g or zero. We observe that only two iterations are usually needed to obtain a good approximate solution.

Equation (19) can be viewed as a shrinkage operator, since it always moves the input towards zero. Unlike the shrinkage operator in (10) of [22], it has a new spatially variant threshold that better preserves edge and promotes smoothness. More importantly, we fix only a scalar $E[|v_i|]$ in each inner iteration, whereas in [22] the whole denominator $|v_i| + E[|v_i|]$ is fixed, which probably alters the properties of NGM.

D. Algorithm

The alternating minimization algorithm for (10) is presented in Alg. 1 which achieves state-of-the-art speed, since the updates for u and v are element-wise operations and the x-update can be efficiently implemented using a 2-D FFT. The penalty weight λ_u is set to 0.1, more choices can be found in [32], and λ_v is set initially to a small value, i.e., 0.001. Since the mean value $E[|v_i|]$ is not a constant in the whole iterative process, we can not prove the convergence of Alg. 1 theoretically, but we show its convergence empirically in the experimental section.

IV. Variational Dirichlet

In this section, we approach the solution of (6) with constraints (7) and (8) using variational inference, under the assumption that $R_h(h) = h^T Q h$ with Q being a $K \times K$ semidefinite symmetric matrix. This is a regularization model frequently found in the BID literature where, for instance, $R_h(h)$ takes into account the covariance matrix of the image estimate (see Levin *et al*. [11] and Babacan *et al*. [17]).

The solution of (6) without considering the constraints in (7) and (8) can be rewritten as a quadratic optimization problem (for simplicity, the iteration index is removed),

$$h^* = \arg\max_h f_x(h) \equiv \frac{1}{2} h^T A h + b^T h \tag{20}$$

where

$$A = \sum_{i=1}^{2} \nabla_i X^T \nabla_i X + \lambda_h Q \tag{21}$$

$$b = -\sum_{i=1}^{2} \nabla_i X^T \nabla_i y \tag{22}$$

The cost function $f_x(h)$ can be derived from the MAP approach, for the Gaussian conditional probability $p(y|h) \propto e^{-\sum_{i=1}^{2} \frac{1}{2}\|\nabla_i X h - \nabla_i y\|_2^2}$ and the blur Gaussian prior $p(h) \propto e^{-\frac{\lambda_h}{2} h^T Q h}$. A shortcoming of the MAP approach is that the constraints (7) and (8) cannot be integrated with the cost function.

The Variational Dirichlet approach we are about to propose aims at approximating the solution in (20) by finding a Dirichlet distribution that is the closest one, in the Kullback−Leibler (KL) divergence sense, to the posterior distribution of the blur given the observation. The Dirichlet distribution generates values always greater than zero, with their sum equal to one. The use of the Dirichlet distribution incorporates in a natural way the constraints in (7) and (8) without the need to explicitly include them into the problem. The parameters of the distribution indicate how sparse the obtained solution will be.

Let $S = \{h | h_i > 0, i = 1, 2, ...K, \sum_i h_i = 1\}$ be the $K-1$ dimensional simplex and $q_\alpha(h)$ the Dirichlet probability density function (see p. 261 in [36]), defined as

$$q_\alpha(h) = \frac{1}{B(\alpha)} \prod_{i=1}^{K} h_i^{\alpha_i - 1}, \ (h \in S \text{ and } \alpha_i > 0), \tag{23}$$

where $B(\alpha)$ is called the *multinomial Beta function* and has the form

$$B(\alpha) = \frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma(S_\alpha)}, \tag{24}$$

with Γ denoting the *gamma* function and $S_\alpha = \sum_{i=1}^{K} \alpha_i$.

Then, we aim at finding the Dirichlet distribution

$$q_{\hat{\alpha}}(h) = \arg\min_{q_\alpha \in \Omega} KL(q_\alpha(h), p(h|y)) \tag{25}$$

where Ω denotes the set of Dirichlet distribution in (23) with $\alpha = (\alpha_1, ..., \alpha_K)$ and all its components greater than

a lower bound $lb > 0$. This is equivalent to finding a factor $\hat{\alpha}$ satisfying

$$\begin{aligned}\hat{\alpha} &= \arg\min_{\alpha \geq lb} KL(q_\alpha(h), p(h|y)) \\ &= \arg\min_{\alpha \geq lb} \int_S q_\alpha(h) \log \frac{q_\alpha(h)}{p(h|y)} dh \\ &= \arg\min_{\alpha \geq lb} \int_S q_\alpha(h) \log \frac{q_\alpha(h)}{p(y|h)p(h)} dh \\ &= \arg\min_{\alpha \geq lb} \int_S q_\alpha(h) \log \frac{q_\alpha(h)}{e^{-\frac{1}{2}h^T Ah - b^T h}} dh \\ &= \arg\min_{\alpha \geq lb} E_{q_\alpha}[\log q_\alpha(h)] + \frac{1}{2} E_{q_\alpha}[h^T Ah] + E_{q_\alpha}[b^T h] \quad (26)\end{aligned}$$

where $\alpha \geq lb$ should be understood as α having all its components greater or equal to lb. Theoretically, $\alpha > 0$ is adequate. However, in practice, we need a lower bound for α to avoid numerical instability.

After finding a solution $\hat{\alpha}$ to (26), one can use the expectation $\hat{\alpha}/S_{\hat{\alpha}}$ as an estimation of the kernel. However, preliminary experiments show that model (26) always leads to very poor kernel estimation. Such failures, as we will show soon, are indeed caused by the over-weighted negentropy term $E_{q_\alpha}[\log q_\alpha(h)]$. To avoid this problem, we introduce a weight γ on this term. As a result, we obtain a more general model

$$\arg\min_{\alpha \geq lb} \gamma E_{q_\alpha}[\log q_\alpha(h)] + \frac{1}{2} E_{q_\alpha}[h^T Ah] + E_{q_\alpha}[b^T h], \quad (27)$$

which includes (26) as a special case when $\gamma = 1$.

The rationale behind the introduction of γ is the following. Eq. (27) is equivalent to minimizing $KL(q_\alpha(h), p_\gamma(h|y))$ where $p_\gamma(h|y)) \propto e^{-\sum_{i=1}^2 \frac{1}{2\gamma}\|\nabla_i X h - \nabla_i y\|_2^2 - \frac{\lambda_h}{2\gamma} h^T Q h}$. By making γ small we approach a degenerate posterior distribution on h, $p_\gamma(h|y)$. So $q_{\hat{\alpha}}(h)$ minimizing (27) will provide a Dirichlet approximation to a neighbour of the MAP. Furthermore, since we are minimizing the reverse KL, also known as *I-projection* or *information projection* (see [37] for details), where $p_\gamma(h|y)$ is small $q_{\hat{\alpha}}(h)$ will also have to be small.

As shown in [36], the negentropy term has the form

$$E_{q_\alpha}[\log q_\alpha(h)] = (\alpha - \mathbf{1})^T(\psi(\alpha) - \psi(S_\alpha)\mathbf{1}) - \log(B(\alpha)), \quad (28)$$

where $\psi = \Gamma'/\Gamma$ is the *digamma* function. Making use of the expectation $E_{q_\alpha(h)}[h] = \alpha/S_\alpha$ and covariance matrix Σ_α given in [36], we obtain in a straightforward way

$$\begin{aligned}E_{q_\alpha}[h^T Ah] &= Tr(A\Sigma_\alpha) + E_{q_\alpha(h)}[h]^T A E_{q_\alpha(h)}[h] \\ &= \frac{S_\alpha A_d^T \alpha - \alpha^T A\alpha}{S_\alpha^2(S_\alpha + 1)} + \frac{\alpha^T A\alpha}{S_\alpha^2} \\ &= \frac{\alpha^T A\alpha + A_d^T \alpha}{(S_\alpha + 1)S_\alpha}, \quad (29)\end{aligned}$$

$$E_{q_\alpha}[b^T h] = \frac{\alpha^T b}{S_\alpha}, \quad (30)$$

where A_d is the vector formed with the diagonal elements of A, *i.e.*, $A_d(i) = A_{ii}, i = 1, 2, ...K$. Together with (28), (29) and (30), the cost function in (27) has the following form

$$\begin{aligned}L(\alpha) = \gamma[(\alpha - \mathbf{1})^T(\psi(\alpha) - \psi(S_\alpha)\mathbf{1}) - \log(B(\alpha))] \\ + \frac{\alpha^T A\alpha + A_d^T \alpha}{2(S_\alpha + 1)S_\alpha} + \frac{b^T \alpha}{S_\alpha}, \quad (31)\end{aligned}$$

whose gradient is

$$\begin{aligned}\nabla L(\alpha) = \gamma(\alpha - \mathbf{1}) \circ (\psi'(\alpha) - \psi'(S_\alpha)\mathbf{1}) + \frac{2A\alpha + A_d}{2(S_\alpha + 1)S_\alpha} \\ + \frac{b}{S_\alpha} - \frac{(\alpha^T A\alpha + A_d^T \alpha)(2S_\alpha + 1)}{2(S_\alpha + 1)^2 S_\alpha^2} - \frac{b^T \alpha}{S_\alpha^2}, \quad (32)\end{aligned}$$

where $\circ$ denotes element-wise product.

Before showing how to optimize $L(\alpha)$, let us discuss the connection between γ and S_α. As γ goes to zero, $p_\gamma(h|y)$ becomes degenerate and hence, $p_{\hat{\alpha}}$ will also have to be close to a degenerate distribution. For a Dirichlet distribution to be degenerate, we need S_α to be infinity. The larger the value of S_α, the closer we are to the MAP solution and therefore the smaller the approximation of the distribution around the MAP. As we will see in the experimental section, approximating the distribution around the MAP rather than finding the MAP leads to reduced noise in the estimated kernel.

$L(\alpha)$ has three terms, the negentropy, quadratic and linear terms. Notice that, since $\psi'(\alpha)$ is strictly decreasing, the negentropy term has a unique stationary point $\alpha = \mathbf{1}$ and, hence, favors the uniform distribution, namely $q(h) = const$. Consequently, decreasing the negentropy term will push α to the point $\mathbf{1}$. $A_d^T \alpha$ in the quadratic term can be view as a weighted l_1 term while $\frac{A_d^T \alpha}{2(S_\alpha+1)S_\alpha}$ adaptively promotes sparsity. Since the weights in A_d are formed from $\nabla_i X^T \nabla_i X$ and Q, sparsity is promoted adaptively regardless of the scale of $\nabla_i X^T \nabla_i X$. Of course, if S_α goes to infinity, this term will vanish. $\frac{\alpha^T A\alpha}{2(S_\alpha+1)S_\alpha} + \frac{b^T \alpha}{S_\alpha}$ can be viewed as the data fidelity term with regularization R_h.

The benefit of the new model is twofold. On one hand, it has an adaptive sparsity regularization term, which can be useful in sparse kernel estimation. On the other hand, it does not have the equality constraint, which allows the optimization problem to be efficiently solved by the gradient projection method [34]. Note that, with the equality constraint, we can not use the gradient projection method. As a disadvantage, the cost function becomes nonconvex.

A. Optimization Algorithm

Let us minimize $L(\alpha)$ with lower bound constraint $\alpha \geq lb$. As $L(\alpha)$ is nonconvex, we search for a local minimum. With the gradient $\nabla L(\alpha)$ defined in (32) at hand, we can use the gradient projection method [34], which is very suitable for box constrained optimization problems. Given a current point α, we need to find a step size s such that the projection α_p (along the direction of negative gradient)

$$\alpha_p = \max(\alpha - s\nabla L(\alpha), lb), \quad (33)$$

satisfies

$$L(\alpha_p) \leq L(\alpha) + \delta_1(\alpha_p - \alpha)^T \nabla L(\alpha). \quad (34)$$

Algorithm 2 Backtracking Gradient Projection Algorithm

Require: $\nabla_i y$, $\nabla_i x$, α_0, λ_h, γ, lb.
1: precompute A_d, b, $\alpha = \max(\alpha_0, lb)$, $s = \sum_{i=1}^{K} \alpha_i$
2: **repeat**
3: $s = \min(\sum_{i=1}^{K} \alpha_i, 1.2s)$
4: $\alpha_p = \max(\alpha - s\nabla L(\alpha), lb)$
5: **while** $L(\alpha_p) > L(\alpha) + 0.01 * (\alpha_p - \alpha)^T \nabla L(\alpha)$ **do**
6: $s = 0.5 * s$
7: $\alpha_p = \max(\alpha - s\nabla L(\alpha), lb)$
8: **end while**
9: $\alpha = \alpha_p$,
10: **until** convergence
11: **return** α

If (34) does not hold, we shrink the step size by

$$s = s * \delta_2, \tag{35}$$

until (34) holds. Note that, since $L(\alpha)$ is continuous for all $\alpha > 0$, we can always find such a small s for (34) to hold. This line search strategy is the so-called backtracking approach. It is used in the proposed gradient projection algorithm for kernel estimation shown in Alg. 2. The parameters $lb = 1$ and $\gamma = 10^{-6}$ are chosen as default.

An important question is how to choose a good initial step size. If it is too small, then the cost function drops very slowly at each iteration. If the step is too large, we may need to shrink the step size many times for (34) to hold. In fact, one can see that ∇L is nearly zero when S_α is large. However, it is not the case for $S_\alpha \nabla L$. Based on this fact, we choose S_α as the initial step size. In step 3 of Alg. 2, we choose $\min(S_\alpha, 1.2s)$ as initial step size instead of S_α, because it saves computation when s is shrunk many times in the last iteration. In many experiments, we often observe that $L(\alpha)$ drops quickly and no extra calls to $L(\alpha)$ are needed.

B. VD approximation for other quadratic models

The proposed VD can also be applied with BID methods, such as Levin *et al*. [11] and Babacan *et al*. [17], whose kernel estimation models are quadratic. For example, in Babacan *et al*. [17], the kernel estimation model (see Eq. (14) in [17]) has the form

$$\begin{aligned} &\arg\min_h \frac{1}{2} h^T A h + b^T h \\ &\text{subject to } h \geq 0 \text{ and } \sum_j h(j) = 1, \end{aligned} \tag{36}$$

where $A(m,n) = \frac{1}{\sigma^2}\sum_i \sum_j \nabla_i x(m+j)\nabla_i x(n+j) + C_{x_i}(m+j, n+j)$ with C_{x_i} being the covariance matrix of $\nabla_i x$ and $b(m) = -\frac{1}{\sigma^2}\sum_i \sum_j \nabla_i x(m+j)\nabla_i y(j)$. In practice, C_{x_i} is approximated by the inverse of the diagonal of $C_{x_i}^{-1}$ (see [17] for more details).

C. Multiscale Blur Kernel Estimation

So far, we have shown how to estimate a blur kernel for a given image and an image for a given kernel, which are the two key steps in most BID methods. By alternating the two steps, we can obtain an estimation of the blur kernel and the image. Unfortunately, directly applying this strategy to the input image may not work if the blur has large support.

To handle large blur supports, most of the existing BID methods use a multiscale scheme which was first applied in motion deblurring by Fergus *et al*. [5]. They point out that single scale BID may suffer from local minimum, particularly for large blur. At the coarsest level, the blur is reduced significantly and hence, it is easy to estimate the kernel from downsampled image. At the next finer level, we can upsample the estimated kernel and use it as a good initial guess for single scale BID. Repeating this process until the finest level, we can obtain a better kernel estimate. In short, the multiscale approach can alleviate the ill-posed nature of BID substantially.

After the kernel is estimated, we need to reconstruct the final sharp image with the estimated kernel using a nonblind deconvolution method (*e.g.*, [38] and [35]), since the image estimated by Alg. 1 is rather smooth.

D. Parameter Settings & Implementation Details

For Alg. 2, we set $Q = C^T C$ with C being the identity or the Laplacian operator (see (21) in [2]). We stop the iteration when the relative variation of cost function is less than 10^{-5} or the number of iterations exceeds 20. We note that the time and space complexity of computing $\nabla_i X^T \nabla_i X$ in (20) is $O(K^2 N)$ and $O(KN)$, respectively, indicating that it is very time and memory consuming to calculate the matrix A. In our algorithm, we do not need to compute A, but $\nabla_i X\alpha$ and $\nabla_i X^T \nabla_i X\alpha$ which can be obtained via a concatenation of 2-D convolutions. If there is no extra call to $L(\alpha)$, each iteration of Alg. 2 costs four 2-D convolutions. We note that the most time consuming parts in computing ∇L, including $\nabla_i X\alpha$ and $\alpha^T \nabla_i X^T \nabla_i X\alpha$, have already been calculated while computing $L(\alpha)$.

The values for λ_x and λ_h are crucial to obtain good kernel estimates, since the deblurring model depends on the two weights. λ_h controls the smoothness of the kernel and also helps avoid delta kernels and over-fitting of the data fidelity term. As λ_h increases, the kernel gets wider and exhibits more noise. Generally, λ_h is proportional to the noise level and image size. Note that, since the image size is varying at each scale, λ_h must be adjusted accordingly. The choice of λ_x depends on many factors, such as the amount of edge information, the degree of blur, noise level, etc. Typical choices for λ_x are in the range $[0.0001, 0.001]$. 0.0002 is chosen as the default value for λ_x. To understand how to choose a suitable λ_x, readers are referred to Section V-C.

V. Experiments

In this section we carry out a comprehensive set of experiments to analize the performance of the proposed BID approach. We begin by showing that formulating the kernel estimation in the filter space provides better results than using the image space. Then, assuming that the real underlying image is known, we analyze the importance of the blur normalization constraint in Alg. 2. Next, we show how Alg.

TABLE I
IMAGE SPACE VS. FILTER SPACE FOR KERNEL ESTIMATION

	Filter space	Image space
Average Error Ratio	1.2257	2.1916
Average SSDs	31.9840	55.8135

2 can be applied to the quadratic regularization in [17] as an example of how our method can be applied to other quadratic regularization methods. Before comparing our method with state-of-the-art BID algorithms [15], [17], [19] and [20], we also conduct two additional experiments aimed at analyzing the numerical performance of the image estimation procedure and the impact of the regularization parameter.

All results are obtained by the default parameters unless otherwise specified. All experiments are carried out using Matlab 7.11 with the Intel Core i5-337U CPU @ 1.8GHz. The proposed BID method takes about 20 seconds to estimate a 19×19 blur from a 256×256 image.

A. Image vs. filter space for kernel estimation

The kernel estimation problem can be formulated either in the image space (3)-(5) or the filter space, (6)-(8). In this manuscript, we are formulating the image estimation in the image space and the kernel estimation in the filter space. To show the superiority of this formulation, we conducted experiments on the dataset [24], which consists of 32 blurred images, corresponding to 4 groundtruth images and 8 motion blur kernels.

Table I reports the average SSD (sum of squared differences) and average error ratio (ratio between SSD errors of the deconvolution with the estimated kernel and the deconvolution with the groundtruth kernel [24]) of all 32 restorations. It is clear that using filtered images for kernel estimation provides better results according to these two criteria. This conclusion is in agreement with the results obtained by Xu *et al*. [15].

B. Kernel Estimation with Groundtruth Image

We now evaluate the performance of the backtracking gradient projection algorithm in Alg. 2 for kernel estimation. For this purpose, we obtain the degraded image in Fig. 2(a) by blurring the *cameraman* image with the kernel h_0 in Fig. 2(b) and corrupting the blurred image by 5% Gaussian noise. The MATLAB function *quadprog* is used to solve the quadratic model (20) with and without the normalization constraint, where $C = I$. In contrast, Alg. 2 is used to find an approximate solution to (20) with different initial points.

Table II shows the l_2-norm errors between the true kernel and the ones estimated by *quadprog* and Alg. 2 with different values for λ_h. The kernels obtained with $\lambda_h = 1$ are presented in Fig. 2 for visual evaluation. As we can see from Table II, it is clear that *quadprog* with normalization constraint yields more accurate solution than without normalization constraint. This occurs because the normalization constraint acts as l_1 regularization and therefore promotes sparsity. Consequently, the normalization constraint suppresses noise, see the noticeable noise in the background of kernel h_1 in Fig. 2(c) and the little noise in kernel h_2 in Fig. 2(d) for comparison. As we can see from Fig. 2(g), kernel h_1 has larger errors than the others at many locations, especially at $(5, 12)$, where h_1 has an error over 0.01 while the errors of the rest kernels are about 0.005.

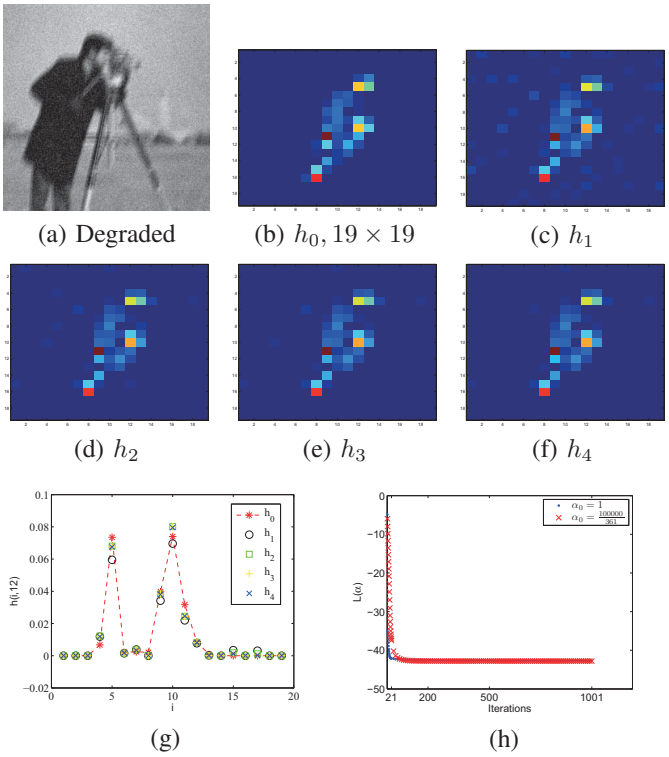

Fig. 2. Kernel estimation by solving model (20) with $\lambda_h = 1$. (a) Degraded image. (b) Groundtruth kernel. (c) *quadprog* with nonnegative constraints only. (d) *quadprog* with nonnegative and normalization constraints. (e) Alg.2 with $\alpha_0 = \frac{100000}{361}$. (f) Alg.2 with $\alpha_0 = 1$. (g) The $12th$ column vectors of the 5 kernels. (h) Evolutions of $L(\alpha)$ with different initial points. Plots (b)-(f) are best viewed on the screen.

TABLE II
l_2-NORM ERROR OF THE KERNELS ESTIMATED BY SOLVING THE QUADRATIC MODEL (20) WITH DIFFERENT λ_h AND OPTIMIZATION METHODS.

λ_h	***quadprog***		**Alg. 2**	
	with (7) only	with (7) and (8)	$\alpha_0 = \frac{100000}{361}$	$\alpha_0 = 1$
0	0.0462	0.0315	0.0311	0.0304
0.01	0.0462	0.0315	0.0311	0.0304
0.1	0.0462	0.0315	0.0311	0.0303
1	0.0463	0.0314	0.0310	0.0303
5	0.0465	0.0312	0.0308	0.0301
10	0.0468	0.0310	0.0306	0.0300
50	0.0502	0.0327	0.0324	0.0321
100	0.0561	0.0397	0.0395	0.0394

Table II and Fig. 2 also show that Alg. 2 produces slightly better results than *quadprog* with both constraints. This is because the cost function $L(\alpha)$, derived from the VD approach, has an adaptive sparsity term, which leads to a more sparse solution (see kernel h_2 in Fig. 2(d) and h_4 in Fig. 2(f) with less noise). Fig. 2(e) looks slightly more noisy than Fig. 2(f) because the adaptive sparsity term is diminished by

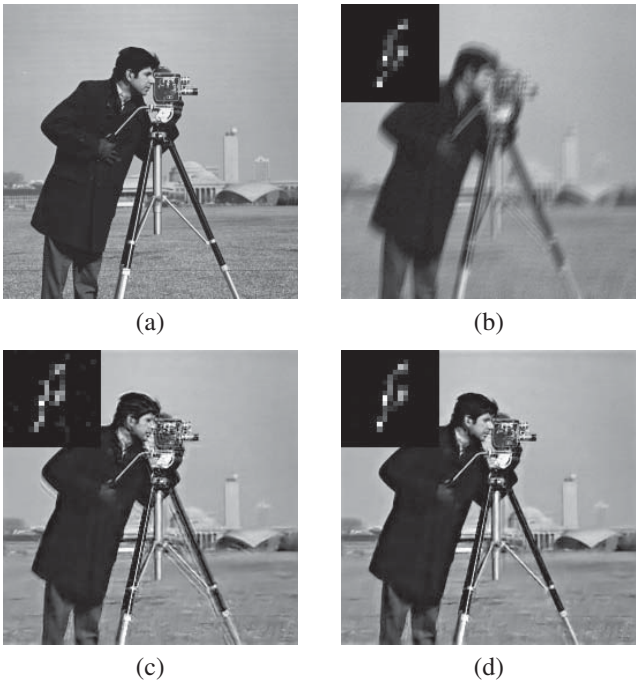

Fig. 3. Variational Dirichlet approximation vs. l_1 regularization with non-negative constraints. (a) $Cameraman$, 256×256. (b) Degraded image with the groundtruth kernel. (c) Restored image with the kernel estimated by [17], PSNR=24.37 dB, SSIM=0.8268. (d) Restored image with the kernel estimated by [17] + Alg. 2, PSNR=27.61 dB, SSIM=0.8732.

large S_α. To further show that an approximate solution to model in (20) can be obtained by Alg. 2, we observe that $||h_3 - h_2||_2 = 0.0017$ and $||h_4 - h_2||_2 = 0.0052$. $||h_3 - h_2||_2$ is 3 times smaller than $||h_4 - h_2||_2$, indicating that large α_0 yields a better approximate solution to model (20) since the adaptive sparsity term could be diminished by large S_α (see (31)). The evolution of the cost function $L(\alpha)$, shown in Fig. 2(h), shows that the proposed backtracking gradient projection algorithm is quite efficient, especially for small values of α_0 (*e.g.*, $\alpha_0 = 1$), since $L(\alpha)$ drops rapidly in less than 20 iterations.

In summary, this experiment indicates that the normalization constraint is important for a precise blur estimation and VD approximation leads to a less noisy kernel than MAP.

C. Application of Alg. 2 to Babacan et al. [17]

Since matrix A in (36) is symmetric and positive definite, we can use the proposed VD to find an approximate solution to (36). We just replace the kernel estimation in [17] with Alg. 2 (default parameters) while keeping the rest unchanged, and compare the resulting BID algorithm with [17]. We use the synthetic image in Fig. 3(b) obtained by blurring the image in Fig. 3(a) and adding a 1% Gaussian noise. To handle the noise in the images, once the kernels are obtained, we reconstruct the images using [38]. As we can see from Fig. 3(c) and (d), the proposed VD improves the result significantly, in terms of ringing artifacts, PSNR and SSIM [39]. Such a remarkable improvement is mainly due to the normalization constraint and the adaptive sparsity promoting term, which helps suppress the noise in kernel (see the kernels in Fig. 3(c) and (d) for comparison). In addition, the method in [17] needs 122.03 seconds to estimate the blur, whereas only 63.29 seconds are needed when our Alg. 2 is incorporated into [17].

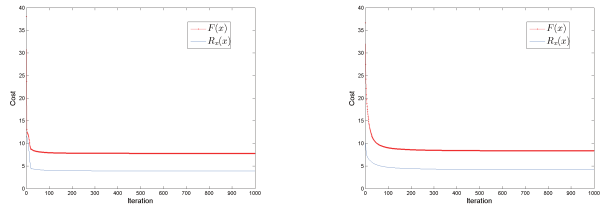

Fig. 4. The evolutions of $F(x)$ and $R_x(x)$, $\lambda_x = 0.0002$. (a) With continuation. (b) Without continuation.

D. Numerical performance of Alg. 1

To show the convergence of Alg. 1, we use the blurred image in Fig. 2(b) as the input image and test Alg. 1 with the groundtruth kernel and 1000 iterations. The evolutions of $F(x) = \frac{1}{2}\|Hx - y\|_2^2 + R_x(x)$ and $R_x(x) = \lambda_x \sum_{i=1}^{2}\sum_{j=1}^{N} |\nabla_i x(j)|/(|\nabla_i x(j)| + \sum_j |\nabla_i x(j)|/N)$ are shown in Fig. 4. As we can see from Fig. 4(a), where the continuation scheme $\lambda_v = \min(\sqrt{2}\lambda_v, 1)$ is applied, the values of both the objective function and regularization term drop quickly and converge rapidly, but with a small jump when λ_v reaches 1. In contrast, as shown in Fig. 4(b) where λ_v is fixed to 1, the two functions decrease slower but more smoothly.

E. The Impact of λ_x on Kernel Estimation

λ_x is the most important parameter in the proposed BID method. Increasing λ_x makes images smoother. As a result, less gradient information is used in kernel estimation. To utilize as much gradient information as possible, we should use a small λ_x. However, we do not want the noise to alter the kernel estimation. Hence, a tradeoff between removing noise and preserving edge information should be made. In general, if the noise standard deviation increases, λ_x should also be increased. Fig. 5 shows the impact of λ_x on kernel estimation, where the groundtruth and observation images are shown in Figs. 3 (a) and (b), respectively. Figs. 5(c) and 5(d) show the corresponding estimation of x obtained by the proposed method. Figs. 5(c) and 5(d) are highly smoothed and cartoon like. Fig. 5(d) is more suitable for kernel estimation than Fig. 5(c) because it has sufficient gradient information and much less noise. Clearly, λ_x determines how much and which gradient information is allowed to participate in kernel estimation.

F. Blind Image Deconvolution on Synthetic Data

We start with the widely used dataset [24], which consists of 32 blurred images, corresponding to 4 groundtruth images and 8 motion blur kernels. We compare the proposed BID method, using NGM and identity or Laplacian operator, with the state-of-the-art BID methods Cho and Lee [9], Levin *et al.* [11],

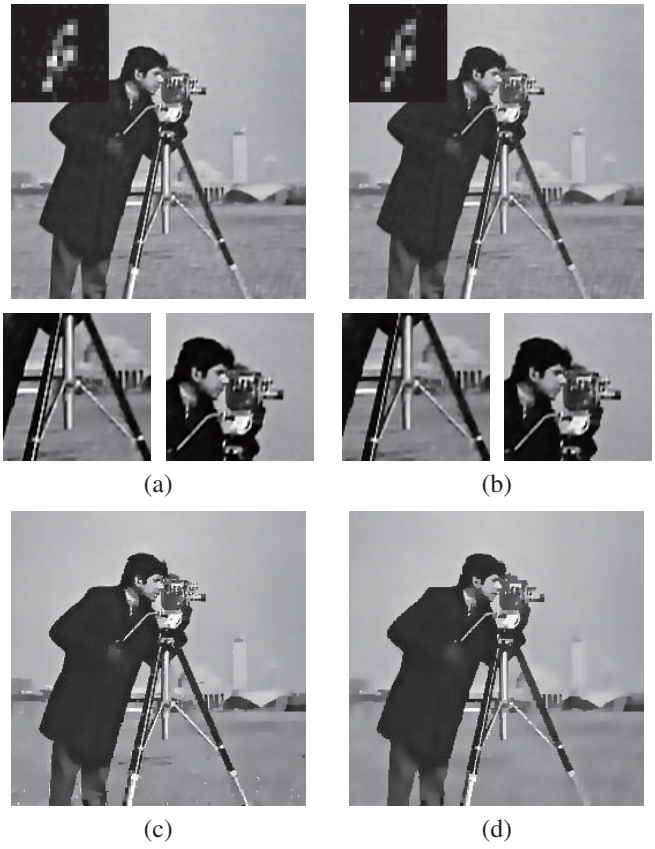

Fig. 5. (a) Restored image by $\lambda_x = 0.0002$ and $\lambda_h = 0.01$, PSNR=24.03 dB, SSIM=0.8418. (b) Restored image by $\lambda_x = 0.0005$ and $\lambda_h = 0.01$, PSNR=28.01, SSIM=0.8869. (c) The final x that produces the kernel shown in (a). (d) The final x that produces the kernel shown in (b).

Babacan *et al*. [17] and Sun *et al*. [19]. In [19], two image priors are proposed, *i.e.*, natural and synthetic priors (see [19] for details). Since the overall performance with both priors is similar, we report their results obtained by synthetic priors only. For fair comparison, once the kernel has been estimated by each method, we use the nonblind deconvolution method [40] with the same parameters in [11] to reconstruct the final sharp image. The parameters of the proposed algorithm are fixed to $\lambda_x = 0.00015$ and $\lambda_h = 0.01$ for all input images. Fig. 6(a) shows that the proposed method has the best performance in terms of error ratio with 93.75% of the results under error ratio 2 and 96.88% less than 3. One deblurring example, which is very challenging, is shown in Fig. 6(b-g). To the best of our knowledge, most of the existing BID methods do not reach an error ratio below 2 for this blurred image. In what follows, for simplicity, all of the kernels are obtained by using $C = I$.

As the dataset [24] is nearly noise free, we also carry out a set of experiments on Sun's dataset [19], which contains 640 images synthesized by blurring 80 natural images with 8 motion blur kernels borrowed from [24]. In this dataset, 1% Gaussian noise is added to every blurred image. Instead of using the whole dataset, we select 40 images synthesized from the 5 natural images shown in the top row of Fig. 7. Again, we fix the parameters $\lambda_x = 0.0002$ and $\lambda_h = 100$

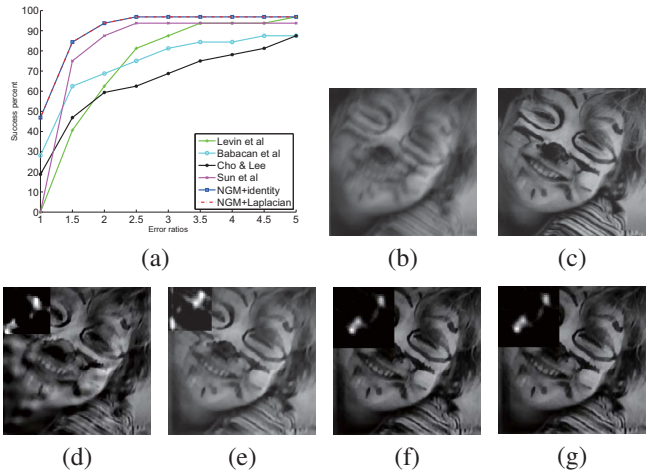

Fig. 6. Quantitative and qualitative evaluation on data set [24]. (a) Cumulative histograms of the error ratios across the data set [24]. (b) One blurred image of the data set. (c) Groundtruth. (d) Babacan *et al* [17], $error\,ratio = 15.327$. (e) Levin *et al* [11], $error\,ratio = 2.462$. (f) NGM+indentity, $error\,ratio = 1.656$. (g) NGM+Laplacian, $error\,ratio = 1.687$

for all input images. Note that a much higher value for λ_h is used to deal with noise and large image size. We compare the proposed method with Babacan *et al*. [17], Xu *et al*. [15] and Sun *et al*. [19]. Since [19] provides better results than many existing methods for this data set, including [9], [14], [11] and [12], we do not report the results by these methods. For fair comparison, we use the nonblind deconvolution method [38] to reconstruct the final sharp image with the PSF estimated with each method. We choose PSNR, SSIM [39] and error ratio [24], calculated by using the MATLAB code provided by [19], to evaluate the recovered images. Table III shows the overall performance of the four compared methods. The average error ratio in Table III shows that all methods can handle 1% Gaussian noise very well, except [17]. The results obtained by the proposed method are much better than the rest for the images number 5 and 27, in terms of all measures, and similar to the results of Sun *et al*. [19] for the other images. We present the estimated images and kernels in Fig. 7 for visual comparison. Our kernels are clean and visually accurate, thanks to the VD approximation, except for image 55 where the kernel is a bit too smooth. All the restored images by the proposed method are of high quality with fewer ringing artifacts. For example, for image 5, our result exhibits fewer ringing artifacts around the gate region than the others, indicating that our kernel is more accurate. For that particular image, we report that the PSNR is 30.36 dB, just about 0.8 dB less than the known kernel restoration (PSNR=31.12 dB), and the SSIM is 0.8439, just about 0.007 below the known kernel case (SSIM=0.8506).

G. Blind Image Deconvolution on Real Data

First of all, we show that our deblurring algorithm removes large motion blurs. We select one severely blurred image (see Fig. 8(a)) from the dataset [41]. For this image, we set $lb = 0.1$ and $\lambda_h = 100$ to deal with large blur and image, respectively.

TABLE III
QUANTITATIVE EVALUATIONS ON THE DATASET IN [19]

	Average PSNR in (dB)					Average SSIM					Average Error Ratio				
Image Index	5	27	41	55	56	5	27	41	55	56	5	27	41	55	56
Groundtruth h	32.05	36.12	32.21	34.43	31.45	0.879	0.931	0.837	0.911	0.887	1.000	1.000	1.000	1.000	1.000
Sun [19]	28.70	33.63	**30.72**	**31.32**	29.14	0.851	0.921	**0.815**	0.899	**0.862**	1.408	1.247	**1.193**	**1.284**	1.642
Xu [15]	29.01	33.54	30.55	31.30	27.71	0.848	0.923	0.812	0.899	0.848	1.469	1.418	1.372	1.465	1.907
Babacan [17]	28.81	31.54	25.77	29.11	27.95	0.845	0.907	0.725	0.873	0.847	1.721	2.768	4.768	2.448	2.058
Proposed	**29.84**	**34.62**	30.40	31.03	**29.33**	**0.857**	**0.926**	0.801	**0.900**	0.859	**1.357**	**1.245**	1.488	1.516	**1.593**

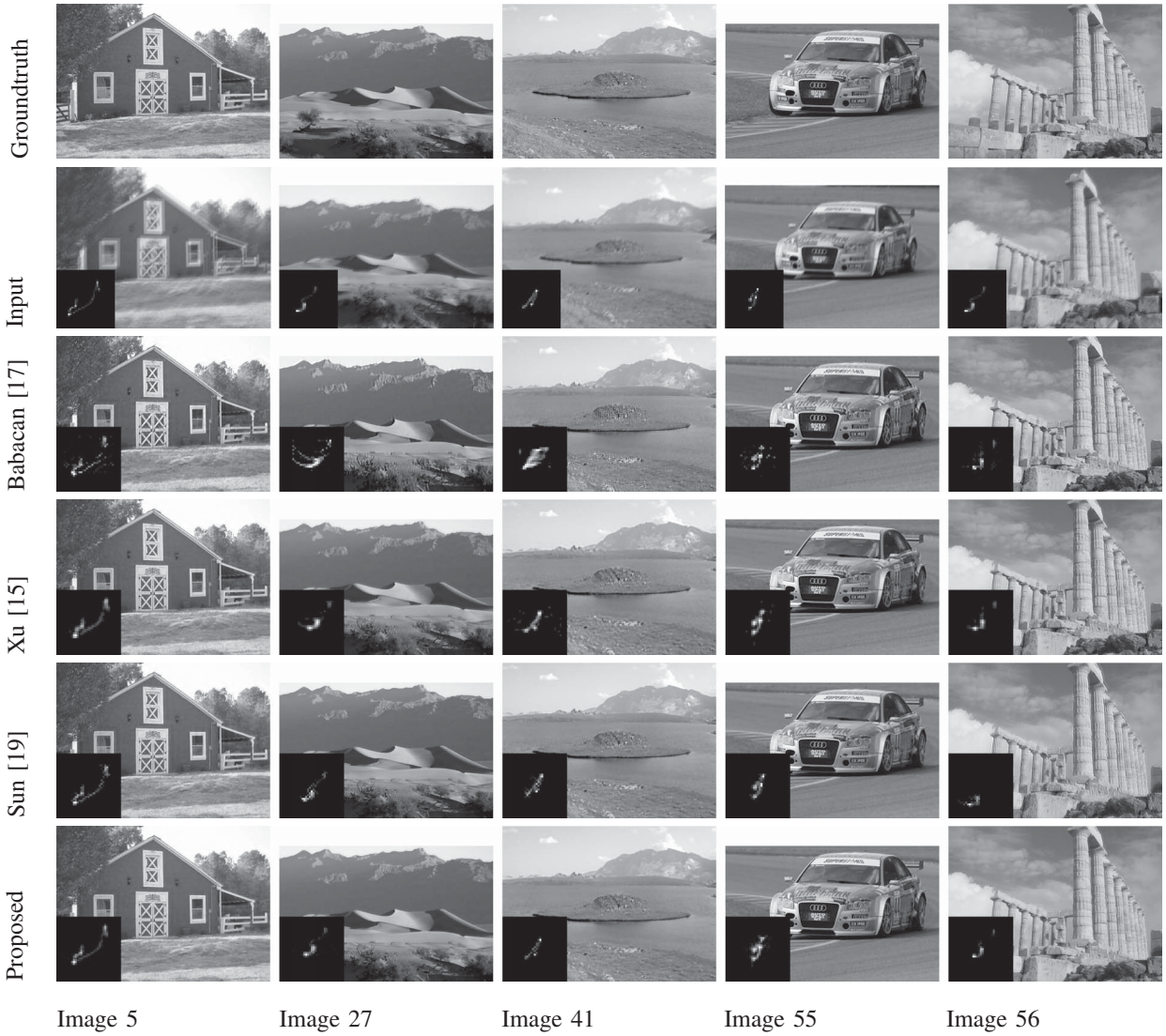

Fig. 7. Selected results on Sun's data set for visual comparison. This figure is best viewed on the screen. Please zoom in for more details.

We should emphasize that the final image is reconstructed by [35], in which an undetermined boundary condition is used to reduce boundary ringing artifacts. As we can see from Fig. 8, our result in Fig. 8(c) is better than Xu's [15] in Fig. 8(b), which produces a too sharpen image due to noise in the estimated kernel, see Fig. 8(d). The kernel obtained by [15] is quite noisy partially because no constraints are used in their kernel estimation step, whereas ours has less noise thanks to the VD approximation. Of course, a good image prior is the key for BID.

We also show that the proposed method can handle noise for real blurry images. Fig. 9(a) presents a blurry and noisy image borrowed from [20]. For this noisy image, we set $\lambda_x = 0.0005$ and $\lambda_h = 100$. Again, for fair comparison, we use [35] with the same parameters to recover the final image for the four methods, Zhong *et al*. [20], Babacan *et al*. [17], Xu *et al*. [15] and ours. We could not compare with Sun *et al*. [19] because their algorithm is not available online. As we can see from Fig. 9(c) and (e), Xu's method [15] and ours produce high quality images, exhibiting a small amount of noise but fewer

(a) Input (b) Xu *et al*. [15] (+ [35]) (c) Ours (+ [35]) (d) [15] (e) Ours

Fig. 8. Removing large motion blur due to camera shake. Please zoom in for more details.

(a) Input (b) Zhong *et al*. [20] (c) Xu *et al*. [15] (d) Babancan *et al*. [17] (e) Ours

Fig. 9. Removing motion blur from a real blurry and noisy image.

ringing artifacts. Babacan's result [17] Fig. 9(d) is satifactory, but with slightly more ringing artifacts and blur. Noticeable ringing artifacts are observed in Zhong's restoration [20] Fig. 9(b). In fact, Zhong's kernel is not very accurate, compared with ours and the one obtained by [15].

Finally, we test the proposed method on blurry images degraded by atmospheric turbulence [6]. Three atmospherically blurred images are shown in Fig. 10. We compare our method with [17] and [15]. The nonblind deconvolution algorithm [38] is adopted to recover the final sharp image. Fig. 10 shows the deblurred images of three methods as well as the estimated kernels. As we can see from the third column of Fig. 10, Xu *et al*. [15] fails totally for all images. The results by [17] are much better than those by Xu *et al*. [15], but still exhibit some noticeable ringing artifacts. Our results in the $4th$ column of Fig. 10, obtained by setting $\lambda_h = 10$, are clearly better than the others, in terms of ringing artifacts and image details. For example, the $5th$ row in Fig. 10 shows that our result exhibits the least amount of ringing artifacts over the car, and the tyres of car and the soldier on the left, barely visible in the other images, are better recovered.

In summary, from the above experiments, it is demonstrated that our proposed method obtains better results not only for motion blur but also for non-sparse blurs like atmospheric turbulence blur, compared with state-of-the-art methods.

VI. Conclusion

Nonnegative and normalization constraints are important for blur kernel estimation, especially in the case of noise and nonuniform blur, such as camera shake and atmospheric blur. With the constraints, l_1 regularization on the kernel does not add extra information on the kernel. For quadratic models with nonnegative and normalization constraints, we proposed a VD approach to find an approximate solution. The VD approach leads to a less noisy kernel estimate than MAP, thanks to the adaptive sparsity promoting term. For the NGM based deconvolution problem, we propose a fast nonblind deconvolution algorithm using mask decoupling and variable splitting. Combining NGM prior with VD, we show that the proposed algorithm is a clear competitor to state-of-the-art BID methods for motion deblurring and outperforms them when removing atmospheric blur.

Acknowledgment

The authors would like to thank Pablo Ruiz Matarán for his helpful discussions on variational Dirichlet and Libin Sun for providing us his MATLAB code for calculating PSNR and SSIM.

References

[1] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Process. Mag.*, vol. 13, no. 3, pp. 43–64, 1996.

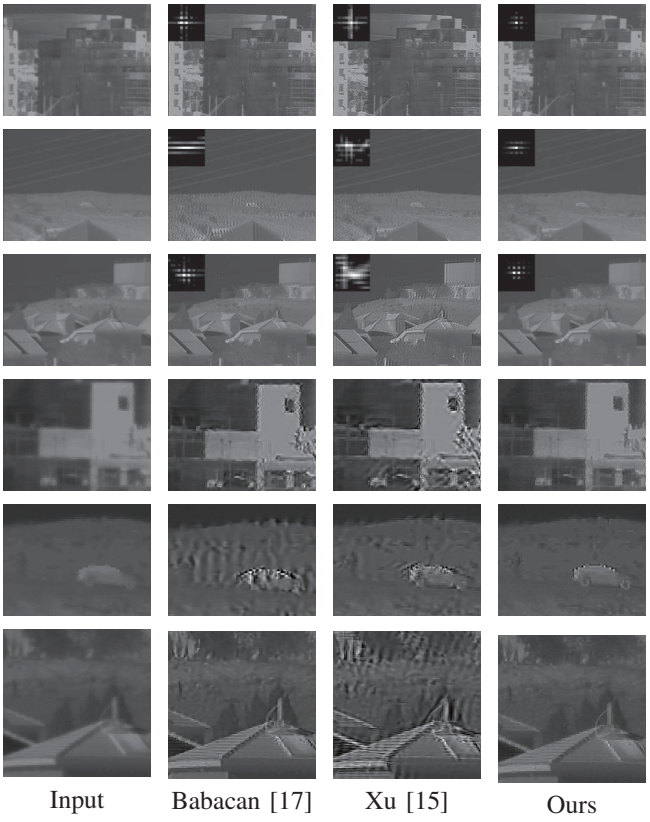

Fig. 10. Removing atmospheric blur from a single image. From left to right: input images with atmospheric blur (borrowed from [6]), results of Babacan *et al*. [17], results of Xu *et al*. [15] and ours. Rows forth to sixth: close-up views of the corresponding images at the first three rows, respectively. Close-up views are best viewed at the original resolution.

[2] M. R. Banham and A. Katsaggelos, "Digital image restoration," *IEEE Signal Process. Mag.*, vol. 14, no. 2, pp. 24–41, 1997.

[3] R. Molina, A. Katsaggelos, J. Abad, and J. Mateos, "A bayesian approach to blind deconvolution based on dirichlet distributions," in *ICASSP*. IEEE, 1997.

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

[5] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph*, vol. 25, no. 3, 2006.

[6] O. Shacham, O. Haik, and Y. Yitzhaky, "Blind restoration of atmospherically degraded images by automatic best step-edge detection," *Pattern Recognition Letters*, vol. 28, no. 15, pp. 2094–2103, 2007.

[7] J. H. Money and S. H. Kang, "Total variation minimizing blind deconvolution with shock filter reference," *Image and Vision Computing*, vol. 26, no. 2, pp. 302–314, 2008.

[8] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graph*, vol. 27, no. 3, 2008.

[9] S. Cho and S. Lee, "Fast motion deblurring," *ACM Trans. Graph*, vol. 28, no. 5, 2009.

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

[11] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," in *CVPR*. IEEE, 2011.

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

[13] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce, "Non-uniform deblurring for shaken images," *International Journal of Computer Vision*, vol. 98, no. 2, pp. 168–186, 2011.

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

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

[16] A. Goldstein and R. Fattal, "Blur-kernel estimation from spectral irregularities," in *ECCV*, 2012.

[17] S. Babacan, R. Molina, M. Do, and A. Katsaggelos, "Bayesian blind deconvolution with general sparse image priors," in *ECCV*, 2012.

[18] C. Wang, Y. Yue, F. Dong, Y. Tao, X. Ma, G. Clapworthy, H. Lin, and X. Ye, "Nonedge-specific adaptive scheme for highly robust blind motion deblurring of natural imagess," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 884–897, 2013.

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

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

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

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

[23] H. Zhang and D. Wipf, "Non-uniform camera shake removal using a spatially-adaptive sparse penalty," in *Advances in Neural Information Processing Systems*, 2013.

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

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

[26] C. Likas and N. Galatsanos, "A variational approach for bayesian blind image deconvolution," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2222–2233, Aug 2004.

[27] R. Molina, J. Mateos, and A. Katsaggelos, "Blind deconvolution using a variational approach to parameter, image, and blur estimation," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3715–3727, Dec 2006.

[28] P. Ruiz, X. Zhou, J. Mateos, R. Molina, and A. K. Katsaggelos, "Variational bayesian blind image deconvolution: A review," *to appear in Digital Signal Processing*, 2015.

[29] S. Osher and L. I. Rudin, "Feature-oriented image enhancement using shock filters," *SIAM Journal on Numerical Analysis*, vol. 27, no. 4, pp. 919–940, 1990.

[30] S. Reeves, "Fast image restoration without boundary artifacts," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1448–1453, Oct 2005.

[31] Y. L. Wang, J. F. Yang, W. T. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[32] M. S. C. Almeida and M. A. T. Figueiredo, "Deconvolving images with unknown boundaries using the alternating direction method of multipliers," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 884–897, 2013.

[33] A. Matakos, S. Ramani, and J. A. Fessler, "Accelerated edge-preserving image restoration without boundary artifacts," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 2019–2029, 2013.

[34] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York, NY: Springer New York, 1999.

[35] X. Zhou, F. Zhou, X. Bai, and B. Xue, "A boundary condition based deconvolution framework for image deblurring," *Journal of Computational and Applied Mathematics*, vol. 261, pp. 14–29, 2014.

[36] M. J. Beal, "Variational algorithms for approximate bayesian inference," Ph.D. dissertation, University of London, 2003.

[37] K. Murphy, *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[38] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *ICCV*. IEEE, 2011.

[39] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[40] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph*, vol. 26, no. 3, 2007.

[41] R. Köhler, M. Hirsch, B. Mohler, B. Schölkopf, and 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, Changsha, China, in 2009. He is currently pursuing the Ph.D. degree from the Image Processing Center, Beihang University, Beijing, China.

From October 2013 to October 2014, sponsored by the Chinese Scholarship Council, he was a Visiting Student with the Department of Computer Science and Artificial Intelligence, University of Granada, Spain. His research interests include image restoration, machine learning, and optimization.

Javier Mateos received the degree in computer science in 1991 and the Ph.D. degree in computer science in 1998, both from the University of Granada. He was an Assistant Professor with the Department of Computer Science and Artificial Intelligence, University of Granada, from 1992 to 2001, and then he became a permanent Associate Professor. He is coauthor of *Superresolution of Images and Video* (Claypool, 2006) and *Multispectral Image fusion Using Multiscale and Super-resolution Methods* (VDM Verlag, 2011). He received the IEEE International Conference on Signal Processing Algorithms, Architectures, Arrangements, and Applications best paper award (2013) and was finalist for the IEEE International Conference on Image Processing Best Student Paper Award (2010). He is conducting research on image and video processing, including image restoration, image and video recovery, super-resolution from (compressed) stills and video sequences, pansharpening and image classification.

Dr. Mateos is a member of the IEEE, Asociación Española de Reconocimento de Formas y Análisis de Imágenes (AERFAI) and International Association for Pattern Recognition (IAPR). He is editor of "Digital Signal Processing" and serves the IEEE as an Associate Editor for the "IEEE Transactions on Image Processing".

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 (formerly, Beijing University of Aeronautics and Astronautics), 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. Since 2000, He has been the Director of Image Processing Center at Beihang University. 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 processing & recognition. He has authored or coauthored over 70 research papers.

Rafael Molina (M88) was born in 1957. He received the degree in mathematics (statistics) in 1979 and the Ph.D. degree in optimal design in linear models in 1983. He became Professor of Computer Science and Artificial Intelligence at the University of Granada, Granada, Spain, in 2000. 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 in using Bayesian modeling and inference in problems like image restoration (applications to astronomy and medicine), super resolution of images and video, blind deconvolution, computational photography, source recovery in medicine, compressive sensing, low rank matrix decomposition, active learning, fusion and classification. See http://decsai.ugr.es/∼rms for publications, funded projects and grants.

Dr. Molina serves the IEEE and other Professional Societies: Applied Signal Processing, Associate Editor (2005-2007), IEEE Trans. on Image Processing, Associate Editor (2010–), Progress in Artificial Intelligence, Associate Editor (2011–) and Digital Signal Processing, Area Editor (2011–). He is the recipient of an IEEE International Conference on Image Processing Paper Award (2007), an ISPA Best Paper Award (2009) and 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 AT&T chair. He was previously the holder of the Ameritech Chair of Information Technology (1997-2003). He is also the Director of the Motorola Center for Seamless Communications, 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, forthcoming). He has supervised 50 Ph.D. theses so far.

Among his many professional activities Prof. Katsaggelos was 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).