

# A Unified Optimization Perspective to Single/Multi-observation Blur-Kernel Estimation with Applications to Camera-Shake Deblurring and Nonparametric Blind Super-Resolution

Wen-Ze Shao<sup>1,2,3</sup> · Qi Ge<sup>1</sup> · Hai-Song Deng<sup>4</sup> · Zhi-Hui Wei<sup>5</sup> · Hai-Bo Li<sup>1</sup>

Received: 5 December 2014 / Accepted: 18 August 2015 / Published online: 1 September 2015  
© Springer Science+Business Media New York 2015

**Abstract** The nonparametric blur-kernel estimation, using either single image or multi-observation, has been intensively studied since Fergus et al.'s influential work (ACM Trans Graph 25:787–794, 2006). However, in the current literature there is always a gap between the two highly relevant problems; that is, single- and multi-shot blind deconvolutions are modeled and solved independently, lacking a unified optimization perspective. In this paper, we attempt to bridge the gap between the two problems and propose a rigorous and unified minimization function for single/multi-shot blur-kernel estimation by coupling the maximum-a-posteriori (MAP) and variational Bayesian (VB) principles. The new function is depicted using a directed graphical model, where the sharp image and the inverse noise variance associated with each shot are treated as random variables, while each blur-kernel, in difference from existing VB methods, is just modeled as a deterministic parameter. Utilizing a universal, three-level hierarchical prior on the latent sharp image and a Gamma hyper-prior on each inverse noise variance,

single/multi-shot blur-kernel estimation is uniformly formulated as an  $\ell_{0,5}$ -norm-regularized negative log-marginal-likelihood minimization problem. By borrowing ideas of expectation-maximization, majorization-minimization, and mean field approximation, as well as iteratively reweighted least squares, all the unknowns of interest, including the sharp image, the blur-kernels, the inverse noise variances, as well as other relevant parameters are estimated automatically. Compared with most single/multi-shot blur-kernel estimation methods, the proposed approach is not only more flexible in processing multiple observations under distinct imaging scenarios due to its independence of the commutative property of convolution but also more adaptive in sparse image modeling while in the meanwhile with much less implementational heuristics. Finally, the proposed blur-kernel estimation method is naturally applied to two low-level vision problems, i.e., camera-shake deblurring and nonparametric blind super-resolution. Experiments on benchmark real-world motion blurred images, simulated multiple-blurred images, as well as both synthetic and realistic low-resolution blurred images are conducted, demonstrating the superiority of the proposed approach to state-of-the-art single/multi-shot camera-shake deblurring and nonparametric blind super-resolution methods.

✉ Wen-Ze Shao  
shaowenze1010@163.com

<sup>1</sup> College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

<sup>2</sup> Department of Computer Science, Technion-Israel Institute of Technology, 32000 Haifa, Israel

<sup>3</sup> Jiangsu Key Laboratory of Image and Video Understanding for Social Safety, Nanjing University of Science and Technology, Nanjing 210094, China

<sup>4</sup> School of Science, Nanjing Audit University, Nanjing 211815, China

<sup>5</sup> School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China

**Keywords** Blur-kernel estimation · Variational Bayesian · Camera-shake removal · Blind deconvolution · Blind super-resolution

## 1 Introduction

The nonparametric blur-kernel estimation problem, using either single image or multi-observation, has been intensively studied since the influential work of Fergus et al. [1].

In the current literature, it is also commonly believed that [2], multi-observation blur-kernel estimation is better posed than its single counterpart. Due to this possible cause, single- and multi-observation blur-kernel estimations are generally modeled and solved in an independent manner, to be observed from the brief overview in the following. Note that, this paper only focuses on the spatially invariant blur-kernel estimation, which forms the basic step for more challenging scenarios of spatially variant and noisy blur-kernel estimation problems [3–7].

## 1.1 Nonparametric Blur-Kernel Estimation Methods

### 1.1.1 Single-shot Case

Roughly speaking, single-shot blur-kernel estimation is currently addressed by two types of inference schemes, i.e., variational Bayesian (VB) and maximum-a-posteriori (MAP). VB is demonstrated, both empirically [1] and theoretically [8] in the scenario of single-shot blur-kernel estimation, superior to MAP as they are regularized by common natural image statistics, e.g., total variation [9–11]. Readers can refer to [12–14] for detailed information on the VB-based signal and image processing approaches. Recently, VB-based blur-kernel estimation methods are reformulated equivalently into their counterpart MAP versions in [15] (regularized by a coupled prior on the sharp image and the blur-kernel), using the convex type of variational representation of image priors. Based on the transformation, the theoretical analysis on the choice of image priors is made, showing that the Jeffreys prior is optimal *to a certain degree* in terms of blur-kernel estimation precision. However, the VB-based blur-kernel estimation methods in [15] are found not rooted in a rigorous optimization perspective. On one hand, the kernel is modeled as a random variable but practically inferred as a deterministic parameter. On the other hand, noise variance updating is full of heuristics. A decreasing sequence of noise variance is pre-specified in [1, 8, 16, 17], while as for [15], although it achieves an adaptive update of the noise variance, the updating formula is purely empirical. Compared to VB, the MAP principle is practiced more commonly. The reason lies in that (i) it is intuitive; (ii) it provides a simple problem formulation; (iii) it is flexible in choosing regularization terms; and (iv) it typically leads to an efficient numerical implementation. More importantly, the state-of-the-art MAP approaches achieve comparable or even higher accuracy of blur-kernel estimation than the VB ones, via performing kinds of heuristic tricks for salient edge prediction, e.g., using highly nonconvex image priors [18–20], enhancing edges by smoothing and shock filtering [21–23], and so on. Nevertheless, it is generally difficult to conduct a profound analysis on success of those empirical prior-based MAP methods. In addition, there are many parameters that influence the results

considerably [2], having to be tuned frequently for generating good deblurring results.

According to the literature survey, it is seen that unnatural image priors are more popular and preferred, particularly as practicing MAP-based single-shot blur-kernel estimation. We also note that the priors are either the off-the-shelf ones, e.g., Jeffreys prior [15, 16], normalized sparsity measure [18], or the empirically designed ones, e.g., [20–23]. Therefore, it is natural to raise a question that whether there is a possibility of adaptively learning a parametric prior for the latent image along with the blur-kernel estimation, so as to achieve more accurate blur-kernel estimation while avoiding the annoying parameter adjustment.

### 1.1.2 Multi-shot Case

The ill-posed nature of blur-kernel estimation can be remedied to a great extent when capturing multi-observation of the same scene [2]. The possible first multi-shot blur-kernel estimation method, proposed by Rav-Acha and Peleg [24], uses two-directional motion blurred images, demonstrating a superior deblurring quality than the case of single image. The benefits of multi-observation deblurring, both nonblind and blind, are also empirically analyzed in [25]. Due to the considered better posedness of multi-shot blur-kernel estimation, the plain  $\ell_\alpha$ -norm-based natural image statistics are frequently utilized in existing blur-kernel estimation approaches, where  $\alpha$  is around 1. For example, [2, 26] exploit an  $\ell_1$ -norm of image gradients; the  $\ell_{1,2}$ -norm and  $\ell_{0,8}$ -norm are, respectively, utilized in [24, 27]; and [25] harnesses the  $\ell_1$ -norm of the tight frame coefficients of images. The reason of MAP-based multi-shot kernel estimation methods relieving from unnatural image priors as required in the single-shot case is that they place more emphasis on exploring the intrinsic properties of two blur-kernels.

A well-known one is the commutative property of convolution, i.e.,  $\mathbf{k}_2 * \mathbf{y}_1 - \mathbf{k}_1 * \mathbf{y}_2 = 0$ , originally proposed in [28], where  $\mathbf{y}_1, \mathbf{y}_2$  are the two blurred versions of a clear image  $\mathbf{u}$ , and  $\mathbf{k}_1, \mathbf{k}_2$  are the blur-kernels corresponding to  $\mathbf{y}_1, \mathbf{y}_2$ . In [29], Sroubek and Flusser reformulate the idea of [28] into a multi-shot regularization term on the blur-kernels. However, there are at least two aspects of modeling flaws as using the commutative property. On one hand, it builds on a noiseless observation model; on the other hand, there is a solution ambiguity in that, for an optimal solution pair  $\{\mathbf{k}_1, \mathbf{k}_2\}$ , a family of solutions  $\{\mathbf{k}_1 * \mathbf{h}, \mathbf{k}_2 * \mathbf{h}\}$  also satisfy the commutative property of convolution. To overcome the inherent problems of the property, several works extend the idea of [28]. For example, in [2] the multi-shot regularization term is improved for robustness against noise; in [27] a sparse prior and a continuity prior are designed for the kernel to mitigate the ambiguity issue; in [26], the estimated blur-kernels  $\mathbf{k}_1 * \mathbf{h}$  and  $\mathbf{k}_2 * \mathbf{h}$  are viewed as blurred images and deblurred

using an  $\ell_\alpha$ -norm-based prior on the kernels  $\{k_1, k_2\}$ . One more limitation of the commutative property of convolution is that, it just applies to the blurred image pairs. As the number of observations grows, it would expand combinatorially hence leading to higher computational burden. In addition, the commutative property of convolution does not apply to a blurred/noisy image pair [30,31].

Recently, another multi-observation blur-kernel estimation method is proposed by Zhang et al. [29], free of the known commutative property. It roughly extends their results on the modeling equivalence [15] between the VB and MAP-based single-shot blur-kernel estimation methods to the multi-shot case, which, however, still imposes the Jeffreys prior on the sharp image. Although the Jeffreys prior is shown to be optimal *to a certain degree* in the single-shot case, in [29], there is no theoretical analysis made on the optimality of the Jeffreys prior for the multi-shot scenario. Note that Zhang et al. [29] cannot be reduced to the single-shot blur-kernel estimation [15], either, without carefully modifying their updating equation of the noise variance. Our another observation from [29] is that it requires the size of the blurred images smaller than the latent sharp image, which is obviously irrational. As for the practical implementation, the minimizing objective function in [29] is derived from the empirical Bayes analysis scheme, or type-II maximum likelihood (ML), and the majorization- minimization (MM) technique [32]. MM has been also extensively used in [29] to derive a tractable optimization scheme. The detailed algorithmic derivation can be referred to Sect. 6 in [29]. Interestingly, the final algorithm of [29] is just a degenerated version of the proposed Algorithm 1 in Sect. 4 of the present paper, despite their difference in modeling motivation, problem formulation, and algorithmic derivation between the two approaches. And, what is more important, the proposed method has been demonstrated to achieve fairly better deblurring performance than [29].

## 1.2 Applications

One natural application of single/multi-shot blur-kernel estimation is the camera-shake deblurring [1], i.e., restoring a clear image from the acquired single- or multiple-motion-blurred photographs. Another application of the proposed method to be considered in this paper is the nonparametric blind super-resolution. For readers better understanding the importance of blur-kernel estimation for super-resolution, we introduce a brief background on nonparametric blind super-resolution. Since the seminal work is proposed by Freeman and Pasztor [33] and Baker and Kanade [34], single-image SR has earned a considerable attention. A careful inspection of the literature in this arena finds that, existing approaches, either learning- or reconstruction-based, concentrate on developing advanced image priors, while mostly

ignoring the need to estimate the blur-kernel. A recent comprehensive survey on SR in [35] (covering work up to 2012) testifies that SR methods generally resort to the assumption of a known kernel, both in the single image and the multi-frame SR regimes. More specifically, in the context of multi-frame SR, most methods assume a squared Gaussian kernel with a suitable standard deviation  $\delta$ , e.g.,  $3 \times 3$  with  $\delta = 0.4$  [36],  $4 \times 4$  with  $\delta = 1$  [37],  $5 \times 5$  with  $\delta = 1$  [38], or a simple pixel averaging, e.g.,  $3 \times 3$  [39,40]. In [41], the authors have considered not only a  $3 \times 3$  Gaussian blur but also a delta blur, i.e., no blur is involved in the modeling of low-res imaging. In single-image nonblind SR, we list a few commonly used options: bicubic low-pass filter (implemented by the Matlab default function *imresize*) [42–49],  $7 \times 7$  Gaussian kernel with  $\delta = 1.6$  [49],  $3 \times 3$  Gaussian kernel with  $\delta = 0.55$  [50], and a pixel averaging [51]. Interestingly, a critical study on single-image SR is recently presented in [52]. The authors conduct an analysis on the effects of two components in single-image SR, i.e., the choice of the image prior and the availability of an accurate blur model. Their conclusion, based on both empirical and theoretical analysis, is that the influence of an accurate blur-kernel is significantly larger than that of an advanced image prior. Furthermore, [52] validates that “an accurate reconstruction constraint<sup>1</sup> combined with a simple gradient regularization achieves SR results almost as good as those of state-of-the-art algorithms with sophisticated image priors.” However, very limited works have addressed the estimation of an accurate nonparametric blur model in the context of image SR reconstruction, either with a single image or multiple frames. As a matter of fact, existing blind SR methods usually assume a parametric model, and the Gaussian kernel is a common choice, e.g., [53,54]. The work reported in [55] is an exception, as it does present a nonparametric kernel estimation approach for blind SR and blind deblurring in a unified framework. However, it is restricting its treatment to single-mode blur-kernels. In addition, [55] is not rooted in a rigorous optimization principle, but rather building on the prediction of sharp edges as an important clue to blur-kernel estimation. Another noteworthy and highly related work is the one recently proposed by Michaeli and Irani [56], who exploit an inherent recurrence property of small natural image patches across different scales and propose a MAP<sub>k</sub>-based estimation method for recovering the blur-kernel (nonparametrically) [17]. The effectiveness of [56] largely builds on the searched nearest neighbors to the query low-res patches in the input blurred and low-res image. We should note that in both [55,56], the kernel estimation includes an  $\ell_2$ -norm-based kernel gradient regularization for promoting smoothness. In this paper, we harness the state-of-the-art nonblind dictionary-based fast single-image SR methods [44–46], which assume

<sup>1</sup> I.e., knowing the blur-kernel.

the simplest bicubic interpolation kernel for the observation model. With the super-resolved but blurred high-res image generated by those methods, the proposed approach (single-shot blur-kernel estimation) is found able to estimate a highly reasonable blur-kernel for practicing blind super-resolution, showing the effectiveness and robustness of the proposed method.

### 1.3 Our Contributions

The core contribution in this paper is the proposal of a unified minimization approach to single/multi-observation blur-kernel estimation. Specifically, our contributions include as follows:

Firstly, the unified approach bases on a directed graphical model, in which the latent sharp image and the inverse noise variance associated with each observation are viewed as random variables, while each blur-kernel, different from existing VB methods, is just modeled as a deterministic parameter. Utilizing a universal, three-level hierarchical prior on the latent sharp image and a Gamma hyper-prior on each inverse noise variance, the single- and multi-shot blur-kernel estimations are uniformly formulated into a negative log-marginal-likelihood minimization problem, which is regularized by an  $\ell_{0.5}$ -norm of blur-kernels.

Secondly, by borrowing ideas of expectation-maximization, majorization-minimization, and mean field approximation, as well as iteratively reweighted least squares, all the unknowns of interest, i.e., the latent sharp image, each blur-kernel, and inverse noise variance, and other related parameters are automatically estimated. Compared with most existing single/multi-shot blur-kernel estimation methods, the proposed approach is not only more flexible in processing multiple observations under distinct imaging scenarios due to its independence of the commutative property of convolution, but also more adaptive in image prior modeling while in the meanwhile with much less implementational heuristics.

Thirdly, the proposed blur-kernel estimation approach is applied to single-image nonparametric blind super-resolution (SR). It utilizes the state-of-the-art nonblind dictionary-based SR methods which often produce a super-resolved but blurred high-res image. With the blur-kernel estimated from the blurred, super-resolved image, the final super-resolved image can be recovered by a relatively simple nonblind SR method that uses a natural image prior. Compared against Michaeli and Irani’s [56] nonparametric blind SR method, the proposed approach is demonstrated to achieve quite comparable or even more accurate blur-kernel estimation, and hence, better super-resolved high-res images. Experimental results on Levin et al.’s [8] benchmark real-world motion blurred images as well as simulated multiple-blurred observations also validate the effectiveness and superiority of the proposed approach for camera-shake deblurring.

The rest of the paper is organized as follows. Section 2 sets up the problem of single/multi-shot blur-kernel estimation. In Sect. 3, Bayesian models of the latent sharp image, each blurred image and inverse noise variance are discussed and depicted using a directed graphical model, based on which a unified minimizing objective function is constructed for the nonparametric single/multi-shot blur-kernel estimation. Section 4 presents a numerical scheme of estimating the posterior distributions of hidden random variables as well as the involved deterministic model parameters. Experiment results on nonparametric blind SR and camera-shake deblurring are provided in Sect. 5, with comparison against the state-of-the-art approaches. The paper is finally concluded in Sect. 6.

## 2 Problem Setup

The single/multi-shot observation process is described as a uniform convolution blur model [2, 24–29]

$$y_i = k_i * u + n_i, \forall i \in \mathcal{U} = \{1, 2, \dots, I\}, \tag{1}$$

where  $I$  is the number of captured corrupted observations,  $y_i$  is the  $i$ th observation,  $k_i$  represents the  $i$ th blur-kernel, and  $n_i$  is supposed to be a Gaussian noise. To be noted that, a registration step is required to align the input observations prior to running the proposed blur-kernel estimation algorithm, just similar to the state-of-the-art multi-shot methods, e.g., [2, 29]. The core task in this paper is hence to estimate the blur-kernels  $\{k_i\}_{i \in \mathcal{U}}$  from the provided blurry images  $\{y_i\}_{i \in \mathcal{U}}$ .

Similar to many existing blur-kernel estimation methods [1, 8, 15–17, 20, 21, 29], we address the blur-kernel estimation in the derivative domain of images for better performance. In specific, Eq. (1) is formulated in the derivative domain using the first-order horizontal and vertical derivative filters  $\nabla^h = [1, -1]$ ,  $\nabla^v = [1; -1]$ , given as

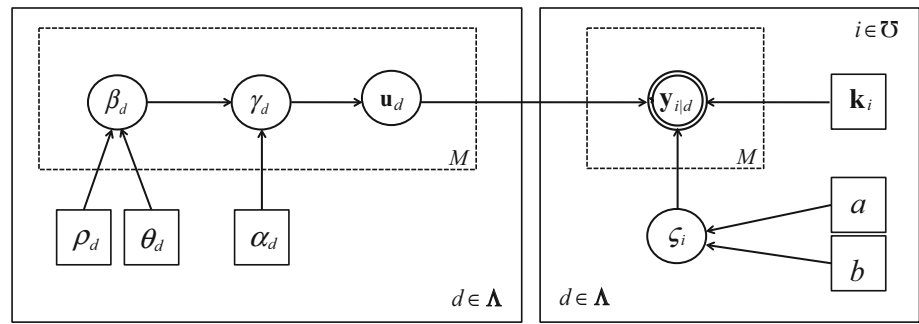
$$y_{i|d} = k_i * u_d + n_{i|d}, \forall d \in \Lambda = \{h, v\}, \tag{2}$$

where  $y_{i|d} = \nabla^d * y_i$ ,  $u_d = \nabla^d * u$ ,  $n_{i|d} = \nabla^d * n_i$ . For the convenience of presentation, Eq. (2) can be also rewritten in a matrix-vector form as

$$y_{i|d} = \mathbf{K}_i \mathbf{u}_d + \mathbf{n}_{i|d}, \tag{3}$$

where  $\{y_{i|d}, \mathbf{u}_d, \mathbf{n}_{i|d}\}_{i \in \mathcal{U}, d \in \Lambda} \in \mathbf{R}^{M \times 1}$  are the vectorized representations of  $\{y_i, u, n_i\}_{i \in \mathcal{U}, d \in \Lambda}$ ,  $M$  is the pixel number in the sharp image  $u$ , and  $\mathbf{K}_i \in \mathbf{R}_0^{+M \times M}$  denotes the convolutional matrix corresponding to the vectorized blur-kernel  $k_i$ . For simplicity, we still assume that  $\mathbf{n}_{i|d}$  is a zero-mean Gaussian noise, i.e.,  $\mathbf{n}_{i|d} \sim \mathcal{N}(0, \zeta_i \mathbf{I}_M)$ , where  $\zeta_i$  is the noise variance called the precision hyper-parameter.

**Fig. 1** Directed graphical model describing the dependencies among the stochastic variables and the deterministic model parameters. The observed stochastic variables are represented by the double circled nodes; the hidden random variables by the circular nodes, and the deterministic parameters by the square nodes



The problem now turns to estimating  $\{\mathbf{u}_d\}_{d \in \Lambda}$ ,  $\{\mathbf{k}_i\}_{i \in \mathcal{U}}$  and  $\{\zeta_i\}_{i \in \mathcal{U}}$  from the images  $\{\mathbf{y}_{i|d}\}_{i \in \mathcal{U}, d \in \Lambda}$ .

### 3 Single/Multi-shot Blur-Kernel Estimation: A Unified Formulation

In this section, single/multi-shot blur-kernel estimation is formulated into a unified framework, incorporating prior assumptions on both the latent sharp images and the blur-kernels via a composite VB and MAP modeling strategy. Simply speaking, the inverse problem (3) is addressed using a directed graphical model, where the observations  $\{\mathbf{y}_{i|d}\}_{i \in \mathcal{U}, d \in \Lambda}$  are treated as the observed random variables, the sharp images  $\{\mathbf{u}_d\}_{d \in \Lambda}$  and the inverse noise variances  $\{\zeta_i\}_{i \in \mathcal{U}}$  are treated as the hidden random variables with assigned proper probability densities, and the blur-kernels  $\{\mathbf{k}_i\}_{i \in \mathcal{U}}$  are just modeled as deterministic parameters imposed by certain regularization constraints. To be specific,  $\mathbf{y}_{i|d}$  obeys a conditional density  $p(\mathbf{y}_{i|d} | \mathbf{u}_d; \zeta_i, \mathbf{k}_i)$ , where a hyper-prior density  $p(\zeta_i; a, b)$  is imposed on the hyper-parameter  $\zeta_i$ . The sharp image  $\mathbf{u}_d$  is modeled by a general sparseness-promoting density represented in a three-layer hierarchical integral form, i.e.,

$$p(\mathbf{u}_d; \alpha_d, \rho_d, \theta_d) = \int p(\mathbf{u}_d | \boldsymbol{\gamma}_d) p(\boldsymbol{\gamma}_d | \boldsymbol{\beta}_d; \alpha_d) p(\boldsymbol{\beta}_d; \rho_d, \theta_d) d\boldsymbol{\gamma}_d d\boldsymbol{\beta}_d, \tag{4}$$

where  $\boldsymbol{\gamma}_d$  and  $\boldsymbol{\beta}_d$  are the stochastic hyper-parameters assigned, respectively, the hyper-prior densities  $p(\boldsymbol{\gamma}_d | \boldsymbol{\beta}_d; \alpha_d)$  and  $p(\boldsymbol{\beta}_d; \rho_d, \theta_d)$ . As for  $\alpha_d, \rho_d, \theta_d$ , the same as blur-kernels  $\{\mathbf{k}_i\}_{i \in \mathcal{U}}$ , they are just modeled as deterministic parameters. The directed graphical model as shown in Fig. 1 plots the dependencies among the stochastic variables and the deterministic model parameters.

#### 3.1 Modeling of Observations and Sharp Images

In the following, Bayesian models of the multi-observations and the latent sharp images are elaborated and formulated in more details. As  $\mathbf{n}_{i|d}$  is assumed a zero-mean Gaussian

noise, the likelihood function  $p(\mathbf{y}_{i|d} | \mathbf{u}_d, \zeta_i; \mathbf{k}_i)$  can be given as

$$p(\mathbf{y}_{i|d} | \mathbf{u}_d, \zeta_i; \mathbf{k}_i) = \mathcal{N}(\mathbf{y}_{i|d} | \mathbf{K}_i \mathbf{u}_d, \zeta_i^{-1} \mathbf{I}_M), \tag{5}$$

where  $\zeta_i = \tilde{\zeta}_i^{-1}$ , and a Gamma hyper-prior is imposed on  $\zeta_i$  to simplify posterior inference because it is conjugate to the Gaussian probability density function (PDF), i.e.,

$$p(\zeta_i; a, b) = \mathcal{G}a(\zeta_i | a, b) = \frac{b^a}{\Gamma(a)} \zeta_i^{a-1} \exp(-b\zeta_i), \tag{6}$$

where  $a$  is the shape parameter,  $b$  is the rate parameter, and  $\Gamma(\cdot)$  denotes the gamma function. In this paper,  $a$  and  $b$  are not automatically estimated but provided empirically and uniformly as  $a = 5 \times 10^{-5}M$  and  $b = 3 \times 10^{-3}M$  for both single- and multi-shot blur-kernel estimation. In this case,  $p(\zeta_i; a, b)$  is simply denoted as  $p(\zeta_i) \triangleq p(\zeta_i; a, b)$ . In the experimental part, the above settings on  $a$  and  $b$  are found work very well for all the blurred images in either camera-shake deblurring or nonparametric blind super-resolution. We should note that, a similar strategy for choosing  $a$  and  $b$  by trial-and-error is also exploited in one of our previous works on single-image blind motion deblurring [57], which is, however, based on a simple nonstationary Gaussian image prior, therefore different from the general sparseness-promoting prior in (4).

According to the literature review in Introduction, image priors are indispensable to blind image deconvolution. The three-layer hierarchical prior (4) is of our particular interest, in that, on one hand, a sparse image prior for blur-kernel estimation should be as general as possible in the form so as to adapt the proposed approach to diverse degradation scenarios; on the other hand, adaptive prior learning becomes computationally tractable as the image prior (4) is represented in the hierarchical form and only determined by three deterministic parameters. With automatic parameter estimation for every blur-kernel estimation problem, it not only relieves us from the parameter adjustment which is unavoidable particularly in the state-of-the-art MAP methods, but also leads to great possibility of achieving higher accurate blur-kernel estimation.

In the first layer of the Bayesian hierarchical modeling,  $p(\mathbf{u}_d|\boldsymbol{\gamma}_d)$  is defined as a product of spatially variant Gaussian PDFs, formulated as

$$p(\mathbf{u}_d|\boldsymbol{\gamma}_d) = \prod_{l \in \Omega} p(\mathbf{u}_d^l|\boldsymbol{\gamma}_d^l) = \prod_{l \in \Omega} \mathcal{N}(\mathbf{u}_d^l|0, \boldsymbol{\gamma}_d^l), \tag{7}$$

where  $\boldsymbol{\gamma}_d^l = (\tilde{\boldsymbol{\gamma}}_d^l)^{-1}$ . In the second layer,  $p(\boldsymbol{\gamma}_d|\boldsymbol{\beta}_d; \alpha_d)$  is selected to be a product of Gamma PDFs with a common shape parameter  $\alpha_d$  for all components of  $\boldsymbol{\gamma}_d$  and an individual rate hyper-parameter  $\boldsymbol{\beta}_d^l$ , given as

$$p(\boldsymbol{\gamma}_d|\boldsymbol{\beta}_d; \alpha_d) = \prod_{l \in \Omega} p(\boldsymbol{\gamma}_d^l|\boldsymbol{\beta}_d^l; \alpha_d) = \prod_{l \in \Omega} \mathcal{Ga}(\boldsymbol{\gamma}_d^l|\alpha_d, \boldsymbol{\beta}_d^l). \tag{8}$$

As a matter of fact, a two-layer hierarchical model is more frequently used in the literature with common shape and rate parameters. To achieve more universal and adaptive sparse image modeling, a three-layer one is preferred in this paper in spite of the possible cost for estimating more parameters. In the third layer, a hyper-prior  $p(\boldsymbol{\beta}_d; \rho_d, \theta_d)$  is imposed on the rate parameter  $\boldsymbol{\beta}_d$  and also selected as a product of Gamma PDFs

$$p(\boldsymbol{\beta}_d; \rho_d, \theta_d) = \prod_{l \in \Omega} p(\boldsymbol{\beta}_d^l; \rho_d, \theta_d) = \prod_{l \in \Omega} \mathcal{Ga}(\boldsymbol{\beta}_d^l|\rho_d, \theta_d), \tag{9}$$

where  $\rho_d, \theta_d$  are the deterministic model parameters. With (7), (8), (9), it is not hard to derive that  $p(\mathbf{u}_d; \alpha_d, \rho_d, \theta_d)$  will reduce to the Jeffreys prior as  $\alpha_d, \rho_d, \theta_d$  are all set as 0.

We should note that the three-layer hierarchical prior was used previously in the fields of compressed sensing [58] and variable selection [59]. However, the three-layer models in [58, 59] are, respectively, Bayesian adaptive versions of the Laplacian distribution and the LASSO estimator. Their optimization schemes are also different from the proposed method in this paper. Specifically, in [58], the type-II maximum likelihood (ML) method has been used to estimate all the random hyper-parameters and deterministic model parameters through extending the relevance vector machine (RVM) in [60]; and [59] exploits the expectation-maximization (EM) method.

### 3.2 Minimizing Objective Function

Now, we come to the task of single/multi-shot blur-kernel estimation, which is to be formulated by taking advantage of both the theoretical robustness of VB-based approaches and the modeling flexibility of MAP-based approaches. The sole prior knowledge about a spatially invariant blur-kernel that we use is that, it can be very sparse when assuming a proper

size for blur-kernels. Therefore, a sparse regularization term is incorporated in our framework to promote the sparseness of blur-kernels. In specific, with Eqs. (9), (10), (11), (12), and (13), the proposed blur-kernel estimation method is formulated into an  $\ell_{0.5}$ -norm-regularized minimizing objective function

$$\begin{aligned} \mathcal{Q}(\{\alpha_d, \rho_d, \theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}}) \\ = \frac{1}{2} \sum_{i \in \mathcal{U}} \|\mathbf{k}_i\|_{0.5}^{0.5} - \log \left\{ \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} p(\mathbf{y}_{i|d}; \alpha_d, \rho_d, \theta_d, \mathbf{k}_i) \right\}. \end{aligned} \tag{10}$$

To be noted that, the  $\ell_{0.5}$ -norm used in (10) is actually an empirical choice which is found to perform better in most of image blurring degradation scenarios than the  $\ell_1$ -norm or the naive  $\ell_0$ -norm. In fact, another property of blur-kernels is also very important, i.e., continuity of the kernel support [27]. In spite of that, the proposed method needs not to impose a continuity prior on the kernels; as a matter of fact, the quality of latent sharp images determines the precision of recovered kernels to a great degree [20], and the proposed approach is capable of producing more accurate edge images  $\{\mathbf{u}_d\}_{d \in \Lambda}$  as core clues to kernel estimation due to the adaptive learning of the image priors  $\{p(\mathbf{u}_d; \alpha_d, \rho_d, \theta_d)\}_{d \in \Lambda}$ . As for  $p(\mathbf{y}_{i|d}; \alpha_d, \rho_d, \theta_d, \mathbf{k}_i)$ , we get it by marginalizing out all the hidden random variables including  $\mathbf{u}_d, \boldsymbol{\gamma}_d, \boldsymbol{\beta}_d, \zeta_i$ , given as

$$\begin{aligned} p(\mathbf{y}_{i|d}; \alpha_d, \rho_d, \theta_d, \mathbf{k}_i) \\ = \int p(\mathbf{y}_{i|d}|\mathbf{u}_d, \zeta_i; \mathbf{k}_i) p(\mathbf{u}_d; \alpha_d, \rho_d, \theta_d) p(\zeta_i) d\mathbf{u}_d d\zeta_i, \end{aligned} \tag{11}$$

just similar to the state-of-the-art VB blind deblurring methods [1, 8, 15–17]. However, compared with those approaches, our method has provided a completely rigorous and feasible optimization framework for estimating the posterior distributions of all the hidden random variables  $\{\mathbf{u}_d, \boldsymbol{\gamma}_d, \boldsymbol{\beta}_d\}_{d \in \Lambda}, \{\zeta_i\}_{i \in \mathcal{U}}$ , as well as the deterministic model parameters including the blur-kernels  $\{\mathbf{k}_i\}_{i \in \mathcal{U}}$  and those involved in the hyper-priors, i.e.,  $\{\alpha_d, \rho_d, \theta_d\}_{d \in \Lambda}$ .

With (11), single/multi-shot blur-kernel estimation is finally formulated as the following minimization problem:

$$\begin{aligned} \min_{\{\alpha_d, \rho_d, \theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}}} \mathcal{Q}(\{\alpha_d, \rho_d, \theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}}) \\ \text{subject to } \alpha_d \geq 0, \rho_d \geq 0, \theta_d \geq 0, \mathbf{k}_i^j \geq 0, \\ \sum_{j \in \mathbf{Z}} \mathbf{k}_i^j = 1. \end{aligned} \tag{12}$$

where  $\mathbf{Z} = \{1, 2, \dots, P\}$ , and  $P$  is the length of the vectorized kernel  $\mathbf{k}_i$ . Note that,  $\alpha_d, \rho_d, \theta_d$  are allowed to be 0 in

(12) just for including the noninformative Jeffreys prior. Nevertheless, it is hard to solve the minimization problem (12) in a direct way. In the following, a computationally tractable upper bound of  $Q(\{\alpha_d, \rho_d, \theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}})$  is to be pursued by borrowing ideas of EM [12] and MM [32].

Firstly, let  $\Upsilon_d = \{\mathbf{u}_d, \boldsymbol{\gamma}_d, \boldsymbol{\beta}_d\}$  and  $\Theta_d = \{\alpha_d, \rho_d, \theta_d\}$ . Then, the objective function (12) can be rewritten as

$$\begin{aligned}
 & Q(\{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}}) \\
 &= \frac{1}{2} \sum_{i \in \mathcal{U}} \|\mathbf{k}_i\|_{0.5}^{0.5} \\
 &+ \{-F(\tilde{q}, \{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}}) - \text{KL}(\tilde{q} \|\tilde{p})\}, \tag{13}
 \end{aligned}$$

where  $\tilde{q}$  and  $\tilde{p}$  are defined as

$$\tilde{q} = \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\Upsilon_d, \zeta_i), \tag{14}$$

$$\tilde{p} = \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} p(\Upsilon_d, \zeta_i | \mathbf{y}_{i|d}; \Theta_d, \mathbf{k}_i), \tag{15}$$

with  $q$  being an any joint PDF of the hidden random variables  $\Upsilon_d, \zeta_i$ , and  $\text{KL}(\tilde{q} \|\tilde{p}) \geq 0$  is the Kullback–Leibler (KL) divergence between PDFs  $\tilde{q}$  and  $\tilde{p}$

$$\text{KL}(\tilde{q} \|\tilde{p}) = - \int \tilde{q} \log \left( \frac{\tilde{p}}{\tilde{q}} \right) d\{\Upsilon_d\}_{d \in \Lambda} d\{\zeta_i\}_{i \in \mathcal{U}}. \tag{16}$$

As for  $F(\tilde{q}, \{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}})$ , it can be formulated as

$$\begin{aligned}
 & F(\tilde{q}, \{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}}) \\
 &= \int \tilde{q} \log \left( \frac{\prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} p(\mathbf{y}_{i|d}, \Upsilon_d, \zeta_i; \Theta_d, \mathbf{k}_i)}{\tilde{q}} \right) \\
 & d\{\Upsilon_d\}_{d \in \Lambda} d\{\zeta_i\}_{i \in \mathcal{U}}. \tag{17}
 \end{aligned}$$

Due to the nonnegativity of  $\text{KL}(\tilde{q} \|\tilde{p}) \geq 0$ , an upper bound of  $Q(\{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}})$  can be thus obtained as

$$\begin{aligned}
 & Q(\{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}}) \leq \tilde{Q}(\tilde{q}, \{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}}) \\
 &= \frac{1}{2} \|\mathbf{k}\|_{0.5}^{0.5} - F(\tilde{q}, \{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}}), \tag{18}
 \end{aligned}$$

where the equality holds only as  $\text{KL}(\tilde{q} \|\tilde{p}) = 0$  implying that  $q(\Upsilon_d, \zeta_i) = p(\Upsilon_d, \zeta_i | \mathbf{y}_{i|d}; \Theta_d, \mathbf{k}_i)$ . With Eqs. (18) and (12), single/multi-shot blur-kernel estimation can now be reformulated with the upper bound  $\tilde{Q}$ , i.e.,

$$\begin{aligned}
 & \min_{\tilde{q}, \{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}}} \tilde{Q}(\tilde{q}, \{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}}) \\
 & \text{subject to } \alpha_d \geq 0, \rho_d \geq 0, \theta_d \geq 0, \mathbf{k}_i^j \geq 0, \\
 & \sum_{j \in \mathcal{Z}} \mathbf{k}_i^j = 1. \tag{19}
 \end{aligned}$$

In other words, the blur-kernel estimation problem reduces to alternating minimization of  $\tilde{Q}$  with respect to the model parameters  $\{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}}$  and the approximate posterior distributions  $\tilde{q}$ . Note that, the proposed approach rigorously relies on Eq. (19) and works without any extra pre-processing or post-processing operations, such as the smoothing or edge enhancement [20,21,23].

## 4 Numerical Implementation

### 4.1 Estimating Posterior Distributions

In order to perform minimization of  $\tilde{Q}$  with respect to the approximating posterior distributions  $\tilde{q}$ , the mean field approximation (MFA) method [12] is used, assuming that the posterior distributions of the hidden random variables  $\{\Upsilon_d\}_{d \in \Lambda}, \{\zeta_i\}_{i \in \mathcal{U}}$  are independent to each other, i.e.,

$$\begin{aligned}
 \tilde{q} &= \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\Upsilon_d, \zeta_i) = \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\zeta_i) q(\Upsilon_d) \\
 &= \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\zeta_i) q(\mathbf{u}_d) q(\boldsymbol{\gamma}_d) q(\boldsymbol{\beta}_d). \tag{20}
 \end{aligned}$$

First of all, we discuss the pursuit of the posterior distribution for the hyper-parameter  $\zeta_i$ . After some straightforward computations,  $-F(\tilde{q}, \{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}})$  can be decomposed with respect to  $q(\zeta_i)$  as follows:

$$\begin{aligned}
 & -F(\tilde{q}, \{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}}) \\
 &= \text{KL} \left( \prod_{d \in \Lambda} q(\zeta_i) \|\tilde{p}(\{\mathbf{y}_{i|d}\}_{d \in \Lambda}, \zeta_i; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i) \right) \\
 &+ \int \prod_{i' \in \mathcal{U} \setminus i} \prod_{d \in \Lambda} q(\zeta_{i'}) \log \prod_{i' \in \mathcal{U} \setminus i} \prod_{d \in \Lambda} q(\zeta_{i'}) d\zeta_{i'} \\
 &+ \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\mathbf{u}_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\mathbf{u}_d) d\{\mathbf{u}_d\}_{d \in \Lambda} \\
 &+ \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\boldsymbol{\gamma}_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\boldsymbol{\gamma}_d) d\{\boldsymbol{\gamma}_d\}_{d \in \Lambda} \\
 &+ \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\boldsymbol{\beta}_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\boldsymbol{\beta}_d) d\{\boldsymbol{\beta}_d\}_{d \in \Lambda}, \tag{21}
 \end{aligned}$$

where  $\tilde{p}(\{\mathbf{y}_{i|d}\}_{d \in \Lambda}, \zeta_i; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i)$  is defined as

$$\begin{aligned}
 & \log \tilde{p}(\{\mathbf{y}_{i|d}\}_{d \in \Lambda}, \zeta_i; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i) \\
 &= \int \prod_{i' \in \mathcal{U} \setminus i} \prod_{d \in \Lambda} q(\zeta_{i'}) q(\mathbf{u}_d) q(\boldsymbol{\gamma}_d) q(\boldsymbol{\beta}_d) \\
 & \log \left( \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} p(\mathbf{y}_{i|d}, \Upsilon_d, \zeta_i; \Theta_d, \mathbf{k}_i) \right) \\
 & d\{\Upsilon_d\}_{d \in \Lambda} d\{\zeta_{i'}\}_{i' \in \mathcal{U} \setminus i}. \tag{22}
 \end{aligned}$$

It is seen that,  $-F(\tilde{q}, \{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}})$  is minimized as the KL divergence  $KL(\prod_{d \in \Lambda} q(\zeta_i) \|\tilde{p}(\{\mathbf{y}_{i|d}\}_{d \in \Lambda}, \zeta_i; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i)$  equals zero, i.e.,  $\prod_{d \in \Lambda} q(\zeta_i) = \tilde{p}(\{\mathbf{y}_{i|d}\}_{d \in \Lambda}, \zeta_i; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i)$ , leading to

$$\begin{aligned} \sum_{d \in \Lambda} \log q(\zeta_i) &= |\Lambda| \log q(\zeta_i) \\ &= \left\langle \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} p(\mathbf{y}_{i|d}, \Upsilon_d, \zeta_i; \Theta_d, \mathbf{k}_i) \right\rangle_{\{q(\zeta_i)\}_{i \in \mathcal{U}} \{q(\Upsilon_d)\}_{d \in \Lambda}} \\ &= \left[ \begin{aligned} &\frac{M}{2} \sum_{d \in \Lambda} \log \zeta_i - \frac{\zeta_i}{2} \sum_{d \in \Lambda} \left\langle \|\mathbf{y}_{i|d} - \mathbf{K}_i \mathbf{u}_d\|_2^2 \right\rangle_{q(\mathbf{u}_d)} \\ &+ \sum_{d \in \Lambda} \log((a-1) \log \zeta_i - b_{\zeta_i}) \end{aligned} \right] \\ &+ const. \end{aligned} \tag{23}$$

Hence,  $q(\zeta_i)$  is actually a Gamma PDF  $\mathcal{G}a(\zeta_i | a_{\zeta_i}, b_{\zeta_i})$  where the shape and rate parameters are defined as

$$\begin{aligned} a_{\zeta_i} &= \frac{M}{2} + a, \\ b_{\zeta_i} &= \frac{1}{2|\Lambda|} \sum_{d \in \Lambda} \left\langle \|\mathbf{y}_{i|d} - \mathbf{K}_i \mathbf{u}_d\|_2^2 \right\rangle_{q(\mathbf{u}_d)} + b, \end{aligned} \tag{24}$$

and its mean is given by  $\langle \zeta_i \rangle_{q(\zeta_i)} = a_{\zeta_i} / b_{\zeta_i}$ .

In Appendix 1, the posterior distributions of  $\zeta_i$  and  $\mathbf{u}_d$ , i.e.,  $q(\zeta_i)$  and  $q(\mathbf{u}_d)$ , have been deduced in every detail for clearness, according to which,  $q(\mathbf{u}_d)$  is a multivariate Gaussian PDF given as

$$q(\mathbf{u}_d) = \mathcal{N}(\mathbf{u}_d | \boldsymbol{\mu}_d, \mathbf{C}_d), \tag{25}$$

with its mean  $\boldsymbol{\mu}_d$  and covariance matrix  $\mathbf{C}_d$  defined as

$$\boldsymbol{\mu}_d = \mathbf{C}_d \sum_{i \in \mathcal{U}} \langle \zeta_i \rangle_{q(\zeta_i)} \mathbf{K}_i^T \mathbf{y}_{i|d}, \tag{26}$$

$$\mathbf{C}_d = \left[ |\mathcal{U}| \text{diag}\{\langle \Upsilon_d \rangle_{q(\Upsilon_d)}\} + \sum_{i \in \mathcal{U}} \langle \zeta_i \rangle_{q(\zeta_i)} \mathbf{K}_i^T \mathbf{K}_i \right]^{-1}. \tag{27}$$

Since it is computationally demanding to calculate Eq. (26) directly, the conjugate gradient (CG) method is used to compute (26) in practice. As for the posterior distributions  $q(\Upsilon_d)$  and  $q(\boldsymbol{\beta}_d)$ , it is easy to derive them in a very similar fashion to  $q(\mathbf{u}_d)$ . Specifically, both  $q(\Upsilon_d)$  and  $q(\boldsymbol{\beta}_d)$  are also Gamma PDFs, given as

$$q(\Upsilon_d) = \prod_{l \in \Omega} \mathcal{G}a(\Upsilon_d^l | a_{\Upsilon_d^l}, b_{\Upsilon_d^l}), \tag{28}$$

$$q(\boldsymbol{\beta}_d) = \prod_{l \in \Omega} \mathcal{G}a(\boldsymbol{\beta}_d^l | a_{\boldsymbol{\beta}_d^l}, b_{\boldsymbol{\beta}_d^l}), \tag{29}$$

with their means, respectively, given by

$$\langle \Upsilon_d^l \rangle_{q(\Upsilon_d^l)} = a_{\Upsilon_d^l} / b_{\Upsilon_d^l}, \quad \langle \boldsymbol{\beta}_d^l \rangle_{q(\boldsymbol{\beta}_d^l)} = a_{\boldsymbol{\beta}_d^l} / b_{\boldsymbol{\beta}_d^l}, \tag{30}$$

where

$$a_{\Upsilon_d^l} = \frac{1}{2} + \alpha_d, \quad b_{\Upsilon_d^l} = \frac{1}{2} \left\langle (\mathbf{u}_d^l)^2 \right\rangle_{q(\mathbf{u}_d^l)} + \langle \boldsymbol{\beta}_d^l \rangle_{q(\boldsymbol{\beta}_d^l)}, \tag{31}$$

with  $\langle (\mathbf{u}_d^l)^2 \rangle_{q(\mathbf{u}_d^l)} = (\boldsymbol{\mu}_d^l)^2 + \mathbf{C}_d^{l,l}$ , and

$$a_{\boldsymbol{\beta}_d^l} = \alpha_d + \rho_d, \quad b_{\boldsymbol{\beta}_d^l} = \langle \Upsilon_d^l \rangle_{q(\Upsilon_d)} + \theta_d. \tag{32}$$

For computational efficiency of  $\langle (\mathbf{u}_d^l)^2 \rangle_{q(\mathbf{u}_d^l)}$ , the covariance matrix  $\mathbf{C}_d$  in Eq. (27) is approximated here by using only its diagonal entries.

### 4.2 Estimating Image Prior Parameters

With (19), (24), (26), (27), (31), (32), estimating  $\{\Theta_d\}_{d \in \Lambda}$  can be achieved by maximizing the objective function

$$\begin{aligned} \mathcal{G}(\{\Theta_d\}_{d \in \Lambda}) &= \left\langle \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} p(\mathbf{y}_{i|d}, \Upsilon_d, \zeta_i; \Theta_d, \mathbf{k}_i) \right\rangle_{\{q(\Upsilon_d)\}_{d \in \Lambda}} \\ &= \sum_{l \in \Omega} \left\{ \begin{aligned} &\alpha_d \langle \log \boldsymbol{\beta}_d^l \rangle_{q(\boldsymbol{\beta}_d^l)} + (\alpha_d - 1) \langle \log \Upsilon_d^l \rangle_{q(\Upsilon_d^l)} \\ &- \log \Gamma(\alpha_d) + \rho_d \log \theta_d - \log \Gamma(\rho_d) \\ &+ (\rho_d - 1) \langle \log \boldsymbol{\beta}_d^l \rangle_{q(\boldsymbol{\beta}_d^l)} - \theta_d \langle \boldsymbol{\beta}_d^l \rangle_{q(\boldsymbol{\beta}_d^l)} \end{aligned} \right\}, \end{aligned} \tag{33}$$

where  $\psi(x) = \Gamma'(x) / \Gamma(x)$  is called a digamma function, and  $\langle \log \Upsilon_d^l \rangle = \psi(\alpha_{\Upsilon_d^l}) - \log b_{\Upsilon_d^l}$ ,  $\langle \log \boldsymbol{\beta}_d^l \rangle = \psi(\alpha_{\boldsymbol{\beta}_d^l}) - \log b_{\boldsymbol{\beta}_d^l}$ . It is impossible to derive the analytical solutions to  $\alpha_d$ ,  $\rho_d$ , and hence the Matlab function `fminbnd` is utilized to estimate the two parameters with projection onto the set of nonnegativity constraints.

### 4.3 Estimating Blur-Kernels

Similarly, blur-kernels can be estimated by solving the following minimization problem:

$$\min_{\{\mathbf{k}_i\}_{i \in \mathcal{U}}} H(\{\mathbf{k}_i\}_{i \in \mathcal{U}}) \text{ subject to } \mathbf{k}_i^j \geq 0, \quad \sum_{j \in \mathcal{Z}} \mathbf{k}_i^j = 1, \tag{34}$$

where

$$\begin{aligned} H(\{\mathbf{k}_i\}_{i \in \mathcal{U}}) &= \frac{1}{2} \sum_{i \in \mathcal{U}} \|\mathbf{k}_i\|_{0.5}^{0.5} \\ &+ \frac{1}{2} \sum_{i \in \mathcal{U}} \sum_{d \in \Lambda} \langle \zeta_i \rangle_{q(\zeta_i)} \left\langle \|\mathbf{y}_{i|d} - \mathbf{K}_i \mathbf{u}_d\|_2^2 \right\rangle_{q(\mathbf{u}_d)}. \end{aligned} \tag{35}$$



Then, using iteratively reweighted least squares (IRLS), estimating  $\mathbf{k}_i$  can be approximately achieved by minimizing

$$\begin{aligned} H_i(\mathbf{k}_i) = & \left( -\langle \zeta_i \rangle_{q(\zeta_i)} \sum_{d \in \Lambda} \mathbf{b}_{i|d}^T \right) \mathbf{k}_i \\ & + \frac{1}{2} \mathbf{k}_i^T \left( \text{diag} \left( \left\{ \frac{1}{\max((\mathbf{k}_i^j)^{1.5}, \varepsilon)} \right\} \right) \right) \\ & + \left( \langle \zeta_i \rangle_{q(\zeta_i)} \sum_{d \in \Lambda} \mathbf{A}_d \right) \mathbf{k}_i, \end{aligned} \quad (36)$$

where  $\varepsilon$  is a very small positive constant, chosen as  $1e6$  for all the experiments in this paper, and

$$\begin{aligned} \mathbf{A}_d(r, s) &= \sum_{l \in \Omega} (\mu_d^{l+r} \mu_d^{l+s} + \mathbf{C}_d^{l+r, l+s}), \\ \mathbf{b}_{i|d}(r) &= \sum_{l \in \Omega} \mu_d^{l+r} \mathbf{y}_{i|d}^l. \end{aligned} \quad (37)$$

where  $r, s$  are the blur-kernel indexes. Note that, Equation (36) is a classic quadratic programming problem, which can be solved using the Matlab function `quadprog` along with projections onto the constraint set.

#### Algorithm 1. Unified single/multi-shot blur-kernel estimation

- 1: **Input.** Blurred images  $\{\mathbf{y}_i\}_{i \in \mathcal{O}}$ .
- 2: **Initialization.** Blur-kernels  $\{\mathbf{k}_i\}_{i \in \mathcal{O}}$ , hyper-parameters  $\{\zeta_i\}_{i \in \mathcal{O}}$ ,  $\{\boldsymbol{\gamma}_d, \boldsymbol{\beta}_d\}_{d \in \Lambda}$ , model parameters  $\{\alpha_d, \rho_d, \theta_d\}_{d \in \Lambda}$ .
- 3: **While** outer iteration number  $L$  not reached, do
  - 4: **While** inner iteration number  $J$  not reached, do
    - 5: **Update**  $q(\mathbf{u}_d)$  using (25), (26), (27);
    - 6: **Update**  $q(\boldsymbol{\gamma}_d)$  and  $q(\boldsymbol{\beta}_d)$  using (28), (29), (30);
    - 7: **Update**  $q(\zeta_i)$  using (23), (24);
    - 8: **Update**  $\alpha_d, \rho_d, \theta_d$  by maximizing (33);
  - 9: **End While**
  - 10: **Update**  $\mathbf{k}_i$  by minimizing (36);
- 11: **End While**
- 12: **Output.** Estimated blur-kernels  $\{\hat{\mathbf{k}}_i\}_{i \in \mathcal{O}}$ .

Based on the above discussion, single/multi-shot blur-kernel estimation is summarized into Algorithm 1. Interestingly, Algorithm 1 will be degenerated to the algorithm in [29] as we set the model parameters  $\{\alpha_d, \rho_d, \theta_d\}_{d \in \Lambda}$  as 0 and omit the  $\ell_{0.5}$ -norm-based kernel regularization in (10). Therefore, the proposed approach is more adaptive in problem modeling than [29] in this sense. In order to account for large-scale blur-kernel estimation and further reduce the risk of getting stuck in poor local minima, a multi-scale scheme of Algorithm 1 is used, similar to many existing VB and

MAP methods in the literature. The number of scales is determined by the input kernel size at the finest scale. The size ratio between consecutive scales in each dimension is  $\sqrt{2}$  such that the kernel size at the coarsest scale is  $3 \times 3$ . In each scale, the input blurred images are the correspondingly downsampled versions of the finest scale, i.e.,  $\{\mathbf{y}_i\}_{i \in \mathcal{O}}$ , and initial blur-kernels are set as the upsampled blur-kernels estimated from the coarser scale. Note that as for the coarsest scale, initial blur-kernels are chosen as a Dirac pulse of size  $3 \times 3$ . The initializations of other parameters are set the same for all the scales:  $\zeta_i$  is set as  $1e2$ , entries of  $\boldsymbol{\gamma}_d$  are all set as  $1e4$ , those of  $\boldsymbol{\beta}_d$  are 0, and the initial  $\alpha_d, \rho_d, \theta_d$  are 0, too. The outer iteration number  $L$  and the inner iteration number  $J$  are, respectively, chosen as 10 and 5.

## 5 Validation and Discussion

In this section, the proposed blur-kernel estimation method is applied to both camera-shake deblurring and single-image nonparametric blind super-resolution, so as to validate its effectiveness and robustness in practical applications. All the

experiments reported in this paper are performed with MATLAB v7.0 on a computer with an Intel i7-4600M CPU (2.90 GHz) and 8 GB memory.

### 5.1 Camera-shake Deblurring

In camera-shake deblurring, deblurred images restored from the provided single/multiple-blurry observations are of our final interest. Hence, with Algorithm 1 producing the estimated blur-kernels, the hyper-Laplacian prior-based non-

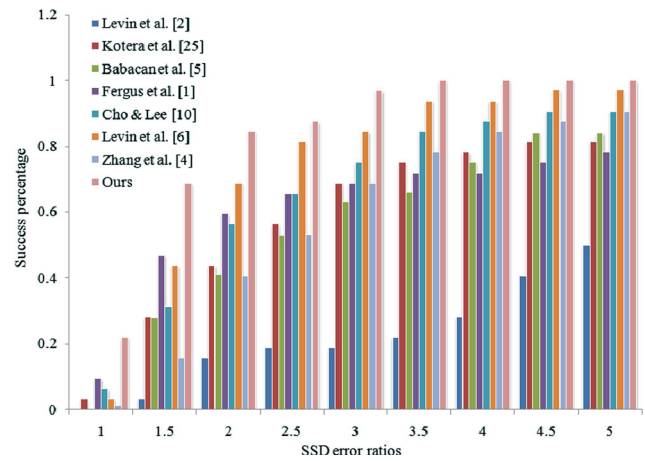
blind deblurring method [61] is utilized to perform nonblind deconvolution for all the camera-shake deblurring experiments in this subsection. Note that, many other more advanced nonblind deconvolution algorithms can be tried, too, e.g., [62–66]. The main reason for choosing [61] in this paper is that it achieves a good compromise between the deblurring quality and the computational complexity. For a fair comparison, [61] is also used for other single/multi-shot blur-kernel estimation methods to perform final nonblind deblurring.

### 5.1.1 Single-shot Case

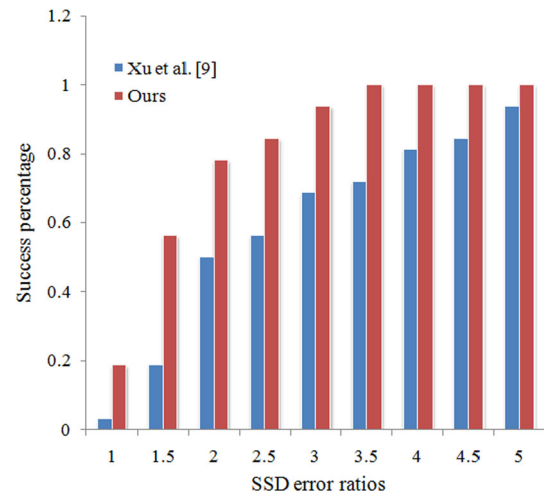
The proposed approach for single-shot blur-kernel estimation is tested on the benchmark blurred dataset provided by Levin et al. in [8], downloaded from [www.wisdom.weizmann.ac.il/~levina/papers/LevinEtalCVPR2011Code.zip](http://www.wisdom.weizmann.ac.il/~levina/papers/LevinEtalCVPR2011Code.zip). In this dataset, 4 sharp images ( $\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \mathcal{I}_4$ ) of size  $255 \times 255$  are blurred by 8 blur-kernels ( $\mathcal{K}_1, \mathcal{K}_2, \mathcal{K}_3, \mathcal{K}_4, \mathcal{K}_5, \mathcal{K}_6, \mathcal{K}_7, \mathcal{K}_8$ ) of sizes ranging from  $13 \times 13$  to  $27 \times 27$ , producing 32 real-captured blurred images. The SSD (sum of squared difference) error ratio defined in [8] is used as the final evaluation measure on different methods. Figure 2 provides the cumulative histogram of SSD error ratios for the proposed and existing seven methods [1, 8, 15–17, 21, 67].<sup>2</sup> As noted by Levin et al. [8], the deblurred images are visually plausible in general as the SSD error ratios are below 3, and in this case, the blind deblurring is considered successful. It is clear that the proposed kernel estimation method has achieved the best performance in terms of the success percentage (97%), as well as uniformly better performance throughout all the bins. The average SSD error ratios of different methods are also provided here: 3.74 [1], 7.05 [8], 2.97 [15], 2.94 [16], 2.02 [17], 2.42 [21], 2.77 [67], and 1.44 (Ours). To show the superiority of adaptively learning image priors, the Jeffreys prior is tried in the proposed framework; the model parameters  $\{\alpha_d, \rho_d, \theta_d\}_{d \in \Lambda}$  in Algorithm 1 are just set as 0 without altering other configurations, which leads to an increased average SSD error ratio 1.82 and a decreased success percentage 91%. Recall that, methods [15, 16] build on the Jeffreys prior but perform far worse than our framework, demonstrating the benefit of adaptively estimating the inverse noise variance as well as additionally incorporating the  $\ell_{0.5}$ -norm-based kernel regularization. To be noted that the size information on the ground truth blur-kernels is assumed to be known in the above experiments.

Another approach to be compared is Xu et al. [20] without any size information on the blur-kernels, which is one of the best-performing MAP methods in the current literature. Since the sizes of blur-kernels in the benchmark blurred

<sup>2</sup> Blur-kernels corresponding to [1, 8, 15–17, 21, 67] are estimated by the Matlab codes provided by the authors.



**Fig. 2** The cumulative histogram of the SSD deconvolution error ratios as assuming an accurate kernel size. The success percentages (SSD error ratio below 3) of different methods are 69% [1], 19% [8], 69% [15], 63% [16], 84% [17], 75% [21], 69% [67], and 97% (Ours) (Color figure online)



**Fig. 3** The cumulative histogram of the SSD error ratios as without accurate kernel size information (a size of  $31 \times 31$  is set uniformly). The success percentages achieved by the two methods are, respectively, 69% [20] and 94% (Ours) (Color figure online)

dataset are not greater than  $31 \times 31$  (a configuration of the software released by Xu et al.), the size  $31 \times 31$  is used for both [20] and the proposed method in this group of experiments. In general, the larger the size of the blur-kernel, the harder the blind image deblurring. In spite of that, the proposed kernel estimation method still achieves a higher success percentage (94%) and a lower average SSD error ratio (1.60) than those of Xu et al. [20] (69%, 2.57). It is also observed from the cumulative histogram of SSD error ratios in Fig. 3 that, the proposed kernel estimation method performs better than Xu et al. [20] throughout all the bins. In Fig. 4, the blur-kernels estimated, respectively, by Xu et al. [20] and the proposed kernel estimation approach are provided for visual perception. It is observed that those of the proposed

**Fig. 4** Ground truth blur-kernels (*top block*) and the kernels estimated by Xu et al. [20] and the proposed method for the 32 benchmark blurred images, uniformly assuming the kernel size as  $31 \times 31$ . The figure shown in each kernel is the calculated kernel SSD error (the lower the better)

	$\mathcal{J}1$	$\mathcal{J}2$	$\mathcal{J}3$	$\mathcal{J}4$	$\mathcal{J}1$	$\mathcal{J}2$	$\mathcal{J}3$	$\mathcal{J}4$	
$\mathcal{K}1$									$\mathcal{K}1$
	0.0214	0.0217	0.0214	0.0230	0.0109	0.0119	0.0164	0.0178	
$\mathcal{K}2$									$\mathcal{K}2$
	0.0160	0.0201	0.0154	0.0193	0.0093	0.0111	0.0103	0.0162	
$\mathcal{K}3$									$\mathcal{K}3$
	0.0118	0.0107	0.0097	0.0116	0.0037	0.0079	0.0064	0.0090	
$\mathcal{K}4$									$\mathcal{K}4$
	0.0011	0.0012	0.0011	0.0014	0.0008	0.0011	0.0011	0.0014	
$\mathcal{K}5$									$\mathcal{K}5$
	0.0180	0.0155	0.0170	0.0211	0.0062	0.0060	0.0104	0.0137	
$\mathcal{K}6$									$\mathcal{K}6$
	0.0065	0.0066	0.0054	0.0078	0.0018	0.0020	0.0029	0.0039	
$\mathcal{K}7$									$\mathcal{K}7$
	0.0026	0.0022	0.0024	0.0029	0.0009	0.0012	0.0015	0.0015	
$\mathcal{K}8$									$\mathcal{K}8$
	0.0053	0.0052	0.0052	0.0062	0.0040	0.0043	0.0038	0.0047	
	Xu et al. [9]				Proposed Approach				

**Fig. 5** Three benchmark test images for constructing multiple-blurred observations. *Left* Cameraman (C), *Middle*: Lena (L), *Right* Pepper (P)



kernel estimation method have less false-motion trajectories and isolated points. To make direct comparison between the two blur-kernel estimation methods, the kernel SSD errors corresponding to each of the 32 benchmark blurred images are also provided in Fig. 4 for both methods. It is not surprising that the proposed kernel estimation method also performs better than Xu et al. [20] in terms of the kernel SSD error.

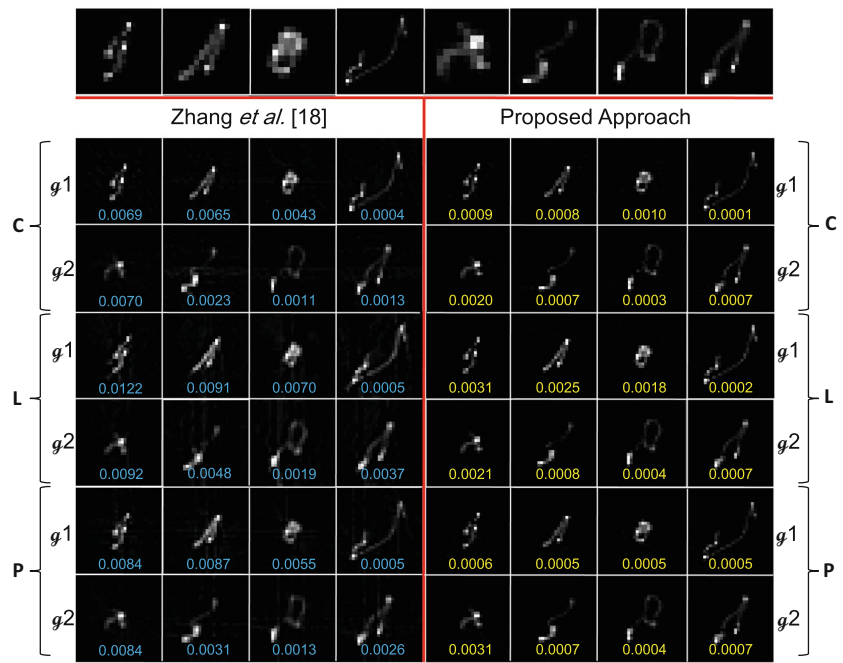
### 5.1.2 Multi-shot Case

For a quantitative comparison, 6 groups of multiple-blurred observations are constructed, based on 3 commonly utilized benchmark test images shown in Fig. 5: Cameraman (C), Lena (L), and Pepper (P). In each group, a test image is blurred by either  $g1 = \{\mathcal{K}1 \dots \mathcal{K}4\}$  or  $g2 = \{\mathcal{K}5 \dots \mathcal{K}8\}$ ,

producing 4 blurred observations. The proposed kernel estimation method is applied to each group of multi-shot observations and compared with the state-of-the-art method [29] which is shown to outperform [2] significantly. Figure 6 provides the blur-kernels (also the kernel SSDs) estimated by the proposed method and [29] for each group of multi-shot observations.<sup>3</sup> It is clearly seen that the proposed kernel estimation method performs better than [29] in terms of the kernel SSD. For visual perception of the final deblurred images using the blur-kernels estimated by [29] and the proposed approach, Figs. 7 and 8 provide the 4 deblurred images for each group of multi-shot observations constructed from

<sup>3</sup> The authors of [29] integrate both the single- and multi-shot blur-kernel estimations in their Matlab p-codes.

**Fig. 6** Ground truth blur-kernels and the kernels estimated by [29] [with the Matlab code (see footnote 3) released by the authors] and the proposed kernel estimation approach for 6 groups of multi-shot images, constructed from three benchmark test images Cameraman (C), Lena (L), and Pepper (P), and two sets of kernels  $g1 = \{\mathcal{K}1 \dots \mathcal{K}4\}$  and  $g2 = \{\mathcal{K}5 \dots \mathcal{K}8\}$ . The figure shown in each blur-kernel is the computed kernel SSD (the lower the better)



**Fig. 7** Deblurred images using the blur-kernels estimated by [29] (*top row*) and the proposed kernel estimation method (*bottom row*) for the multiple-blurred observations simulated from Cameraman (C) and  $g1 = \{\mathcal{K}1 \dots \mathcal{K}4\}$ . The figure in each image is the calculated PSNR score

Cameraman (C), along with the PSNR scores; Figs. 9 and 10 correspond to those of Lena (L); and Figs. 11 and 12 correspond to those of Pepper (P). It is not surprising that the proposed kernel estimation approach achieves not only higher PSNR but also better visual quality, thus validating the modeling superiority of the proposed method to [29], indicating that on average lower kernel SSDs amount to better final deconvolution quality.

The blur-kernel size for the above experiments is uniformly specified set as  $31 \times 31$ . With the same size, the proposed blur-kernel estimation method is applied to each

of the 24 blurred images simulated in the multi-shot case (3 benchmark images are blurred by 8 kernels), just to validate the superiority of the multi-shot blur-kernel estimation to the single-shot scenario. The estimated blur-kernels (along with their kernel SSDs) by the proposed approach are provided in Fig. 13. Compared with the results obtained in the multi-shot case as provided in Fig. 6, it is found that the proposed blur-kernel estimation approach in the single-shot case achieves comparable or sometimes even superior performance than [29] but inferior performance on average than the proposed approach in the multi-shot case. In particular, Fig. 14 pro-



**Fig. 8** Deblurred images using the blur-kernels estimated by [29] (*top row*) and the proposed kernel estimation method (*bottom row*) for the multiple-blurred observations simulated from Cameraman (C) and  $g_2 = \{\mathcal{K}5 \dots \mathcal{K}8\}$



**Fig. 9** Deblurred images using the blur-kernels estimated by [29] (*top row*) and the proposed kernel estimation method (*bottom row*) for the multiple-blurred observations simulated from Lena (L) and  $g_1 = \{\mathcal{K}1 \dots \mathcal{K}4\}$

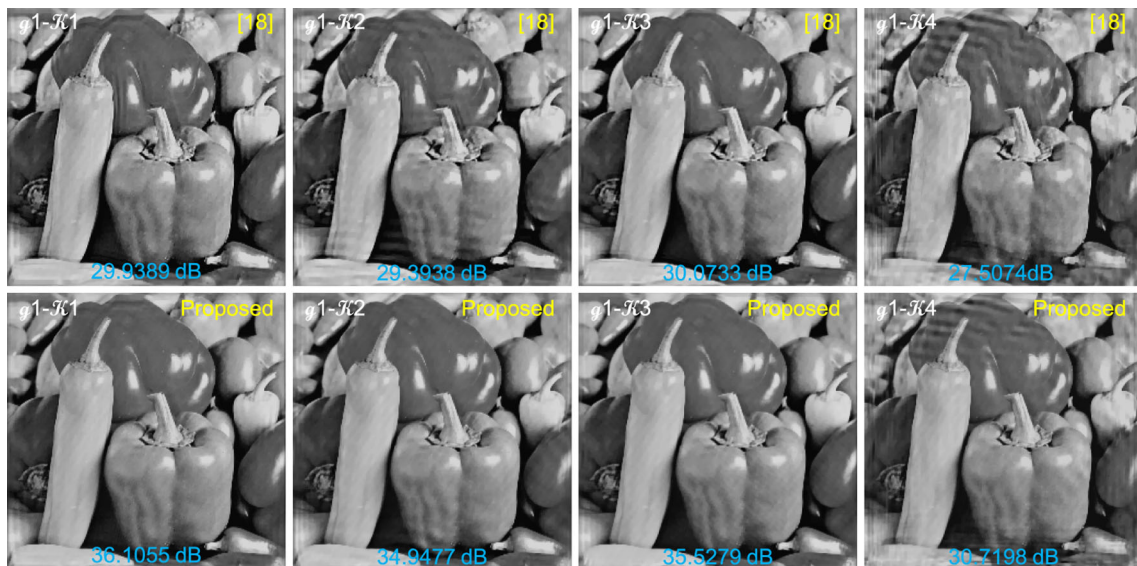
vides the three final deblurred images corresponding to the estimated kernel  $\mathcal{K}8$ , which are obviously with worse visual perception and PSNR scores than the ones in the case of multi-shot kernel estimation. But it should be noted that the multi-shot blur-kernel estimation is computationally more expensive than the single-shot case. We take the experiments with Cameraman (C) and  $g_2 = \{\mathcal{K}5 \dots \mathcal{K}8\}$ , for example. The running time of the proposed kernel estimation approach in the multi-shot case is 908s, while the running time in the single-shot case is, respectively, 348s, 351s, 368s, and 353s. It is also seen that the proposed blur-kernel estimation approach has a much high computation burden even in the single-shot

case, which is actually a common characteristic of VB-like methods [1, 8, 9, 11, 15–17, 32]. One strategy for reducing the running time is to perform blur-kernel estimation using only part of the blurred images. How to improve the computational efficiency of the proposed method in a fundamental way is actually another challenge to work on in the future.

In the following, we discuss the convergence of the proposed blur-kernel estimation approach using multiple-blurred images in an empirical way. Considering the limited space, we still take Cameraman (C) and  $g_2 = \{\mathcal{K}5 \dots \mathcal{K}8\}$ , for example. In Fig. 15, the intermediately estimated blur-kernels during the 10 outer iterations of Algorithm 1 (the



**Fig. 10** Deblurred images using the blur-kernels estimated by [29] (*top row*) and the proposed kernel estimation method (*bottom row*) for the multiple-blurred observations simulated from Lena (L) and  $g2 = \{K5 \dots K8\}$



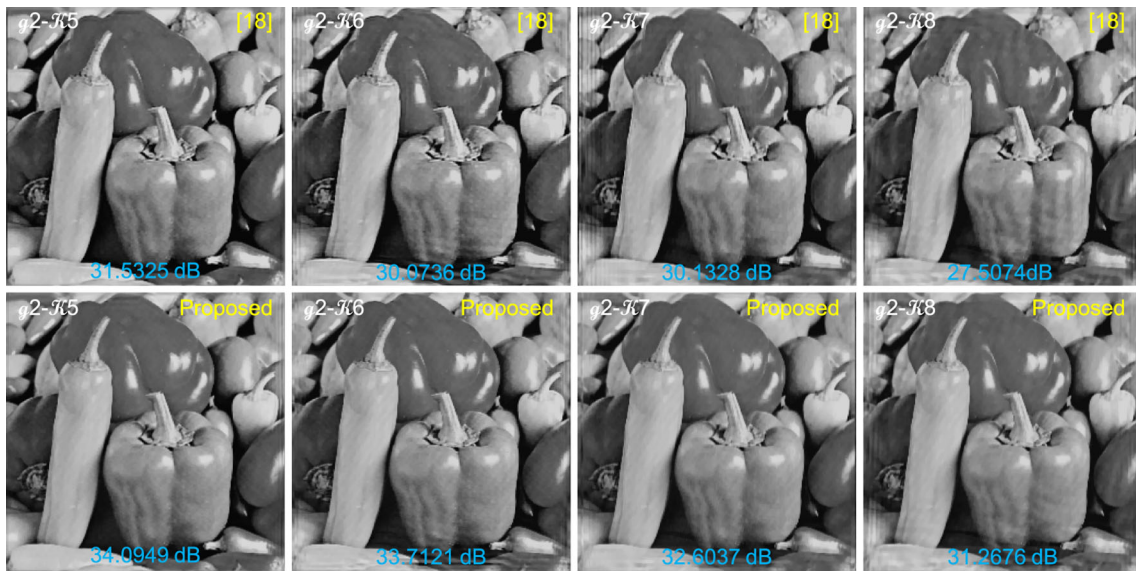
**Fig. 11** Deblurred images using the blur-kernels estimated by [29] (*top row*) and the proposed kernel estimation method (*bottom row*) for the multiple-blurred observations simulated from Pepper (P) and  $g1 = \{K1 \dots K4\}$

finest scale) are provided for each of the 4 blurry observations; in Fig. 16, the intermediate results of gradient magnitude image<sup>4</sup> are also shown for visual perception. The intermediate estimates of both the gradient magnitude image and the blur-kernel clearly demonstrate that the proposed kernel estimation approach achieves a good convergence performance. It is more obviously observed from the kernel SSD listed in each blur-kernel that, the proposed method has pro-

duced a reasonable blur-kernel with 6 outer iterations for each blurry observation.

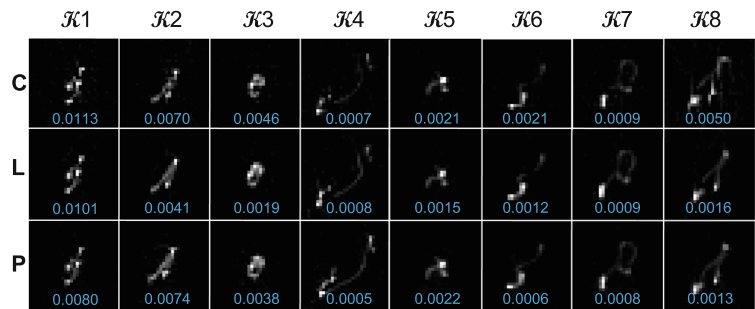
As suggested by one of the reviewers, we also discuss the sensitivity of the proposed kernel estimation approach to the initializations of hyper-parameters  $\{\zeta_i\}_{i \in \mathcal{S}}$ ,  $\{\gamma_d, \beta_d\}_{d \in \Lambda}$  and model parameters  $\{\alpha_d, \rho_d, \theta_d\}_{d \in \Lambda}$ . Recall that, the above experimental results are generated with the following initializations: (O)  $\zeta_i = 1e2$ , entries of  $\gamma_d$  are all set as  $1e4$ , those of  $\beta_d$  are set as 0, and  $\alpha_d, \rho_d, \theta_d$  are set as 0. In the following, eight groups of experiments are conducted via altering those initializations, still using the 4 blurry observations simulated from Cameraman (C) and  $g2 = \{K5 \dots K8\}$ . The initializa-

<sup>4</sup> The gradient magnitude image refers to the magnitude image calculated based on the intermediate estimates of horizontal and vertical gradients.



**Fig. 12** Deblurred images using the blur-kernels estimated by [29] (top row) and the proposed kernel estimation method (bottom row) for the multiple-blurred observations simulated from Pepper (P) and  $g_2 = \{K_5 \dots K_8\}$

**Fig. 13** Blur-kernels estimated by the proposed kernel estimation method applied to each of the 24 blurred images simulated in the multi-shot case (3 benchmark images in Fig. 5 are blurred by 8 kernels provided by Levin et al. [8])



**Fig. 14** Final deconvolved images with the motion blur-kernel  $K_8$  estimated by the proposed kernel estimation approach in the scenario of single-shot blur-kernel estimation

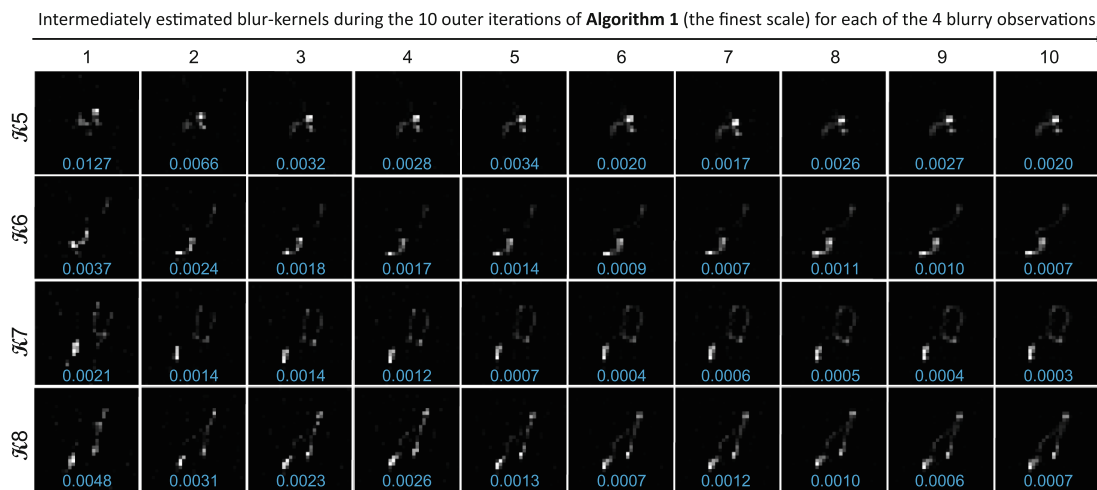


tion settings for the eight groups of experiments are given as follows:

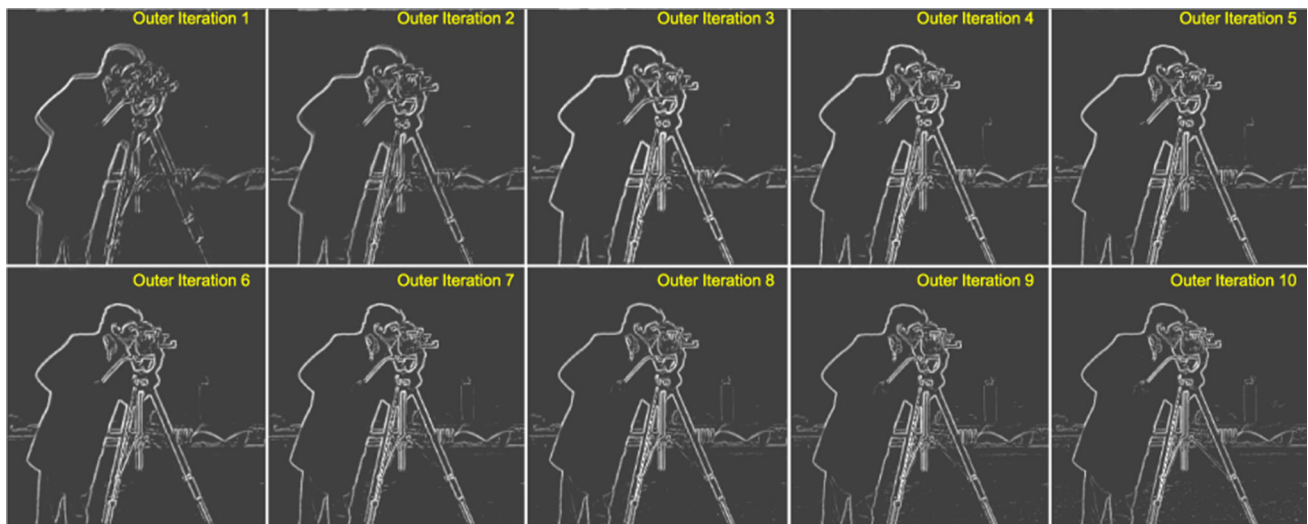
- (1-1)  $\varsigma_i = 2e1$  with other initializations in (O) unchanged;
- (1-2)  $\varsigma_i = 2e2$  with other initializations in (O) unchanged;
- (2-1)  $\alpha_d = \rho_d = \theta_d = 1$  with other initializations in (O) unchanged;
- (2-2)  $\alpha_d = \rho_d = \theta_d = 1e1$  with other initializations in (O) unchanged;
- (3-1) entries of  $\beta_d$  are set as  $1e-6$  with other initializations in (O) unchanged;

- (3-2) entries of  $\beta_d$  are set as  $1e-5$  with other initializations in (O) unchanged;
- (4-1) entries of  $\gamma_d$  are set as  $5e3$  with other initializations in (O) unchanged;
- (4-2) entries of  $\gamma_d$  are set as  $5e4$  with other initializations in (O) unchanged.

Using the proposed multi-shot kernel estimation method, Tables 1 and 2, respectively, provide the kernel SSD and the PSNR score of the final deblurred image for each of the 4 blurry observations. The results show that our proposed



**Fig. 15** Intermediately estimated blur-kernels during the 10 outer iterations of Algorithm 1 (the finest scale) for each of the 4 blurry observations simulated utilizing the benchmark test image Cameraman (C) and the ground truth kernels  $\mathbf{g}2 = \{\mathcal{K}5 \dots \mathcal{K}8\}$



**Fig. 16** Intermediate results of gradient magnitude image during the 10 outer iterations of Algorithm 1 (the finest scale). The magnitude images are shown with a linear transform for better visual perception

**Table 1** SSD of the blur-kernel estimated by the proposed multi-shot kernel estimation method for each of the 4 blurry observations simulated from Cameraman (C) and  $\mathbf{g}2 = \{\mathcal{K}5 \dots \mathcal{K}8\}$

Initialization setting	$\mathbf{g}2 - \mathcal{K}5$	$\mathbf{g}2 - \mathcal{K}6$	$\mathbf{g}2 - \mathcal{K}7$	$\mathbf{g}2 - \mathcal{K}8$
(O)	0.0020	0.0007	0.0003	0.0007
(1-1)	0.0025	0.0009	0.0004	0.0008
(1-2)	0.0028	0.0008	0.0004	0.0007
(2-1)	0.0018	0.0006	0.0003	0.0005
(2-2)	0.0018	0.0007	0.0003	0.0006
(3-1)	0.0014	0.0005	0.0003	0.0004
(3-2)	0.0027	0.0008	0.0004	0.0007
(4-1)	0.0026	0.0006	0.0003	0.0004
(4-2)	0.0036	0.0015	0.0005	0.0010

**Table 2** PSNR score (dB) of the final deblurred image based on the proposed multi-shot kernel estimation method for each of the 4 blurry observations simulated from Cameraman (C) and  $\mathbf{g}2 = \{\mathcal{K}5 \dots \mathcal{K}8\}$

Initialization setting	$\mathbf{g}2 - \mathcal{K}5$	$\mathbf{g}2 - \mathcal{K}6$	$\mathbf{g}2 - \mathcal{K}7$	$\mathbf{g}2 - \mathcal{K}8$
(O)	31.9723	32.0125	31.4351	30.7235
(1-1)	31.6642	31.9495	30.8894	30.2004
(1-2)	31.9264	32.0263	31.2532	30.6720
(2-1)	32.4856	32.5502	31.6454	31.0186
(2-2)	32.2637	32.3758	31.3750	30.7695
(3-1)	33.2348	33.3459	32.1010	31.4282
(3-2)	32.1017	32.2195	31.3313	30.7037
(4-1)	32.1381	32.4358	31.5551	30.9264
(4-2)	30.9525	31.6289	30.3770	29.6792





**Fig. 17** Nonparametric blind SR ( $\text{SR} \times 2$ ) with low-res *Hollywood* using the proposed kernel estimation method (-NE [44], -JSC [45], -ANR [46]), Michaeli and Irani [56], and nonblind SR methods [44–46].

kernel estimation method achieves stability to a great extent in different initialization settings. In the meanwhile, we also see that the selected initialization setting (O) for Algorithm 1 is absolutely not the best, and hence, we believe that the comparison between Algorithm 1 and other kernel estimation methods is convincing.

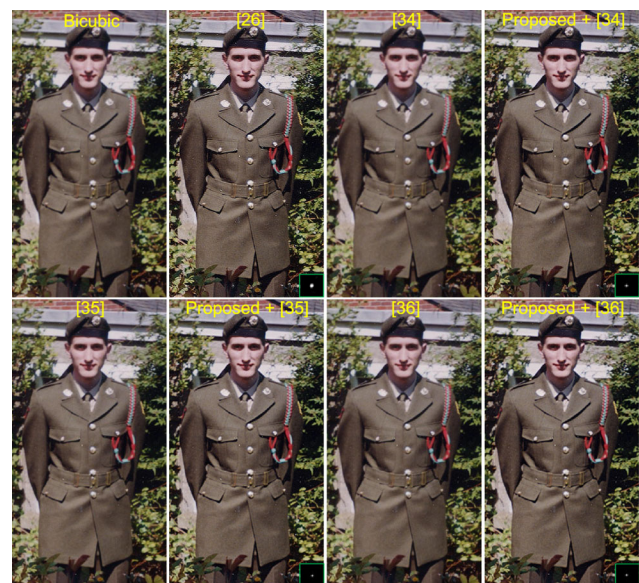
## 5.2 Nonparametric Blind Super-Resolution

When super-resolving a clear, high-res image from its corresponding blurred, low-res version in this paper, the size of the blur-kernel in the degradation model is also uniformly set as  $31 \times 31$ . Intuitively, for the estimated blur-kernel reliable to its ground truth as much as possible, we should harness those nonblind SR methods assuming a blur-kernel as simple as possible, e.g., bicubic interpolation kernel. As such, three candidate state-of-the-art nonblind dictionary-based fast SR algorithms [44–46] are chosen for practicing nonparametric blind SR, including neighborhood embedding (NE) [44], joint sparse coding [45] (JSC), and anchored neighbor regression (ANR) [46] (A brief review is provided as Appendix 2). With a pre-estimated blur-kernel, we generate the final super-resolved image using a simple reconstruction-based nonblind SR method [52], which is regularized by the natural hyper-Laplacian image prior [68] (Fig. 17).

Using low-res blurred images, we make comparisons between the proposed kernel estimation method (-NE [44], -JSC [45], -ANR [46]) and state-of-the-art nonparametric blind SR method proposed by Michaeli and Irani [56].<sup>5</sup> The

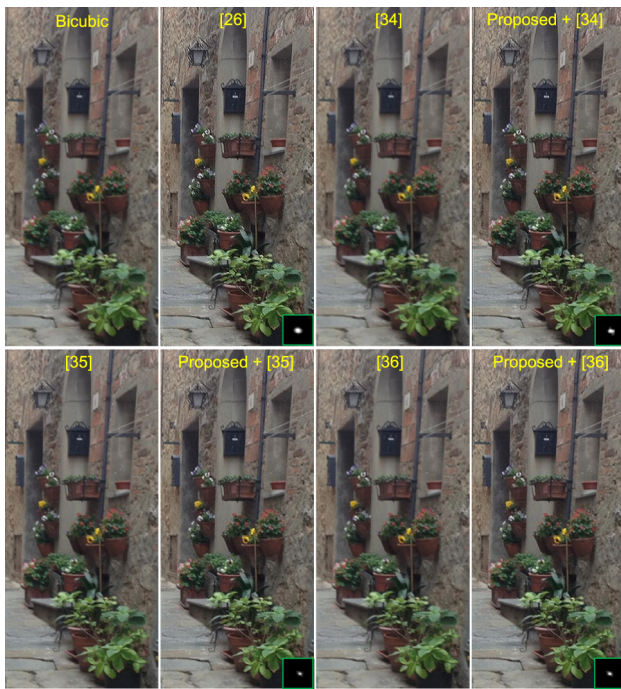
<sup>5</sup> The estimated blur-kernels corresponding to [56] were kindly provided to us by Tomer Michaeli who is the first author of [56].

The first image in the *top row* is the bicubic interpolated version of the low-res image using the Matlab function `imresize`

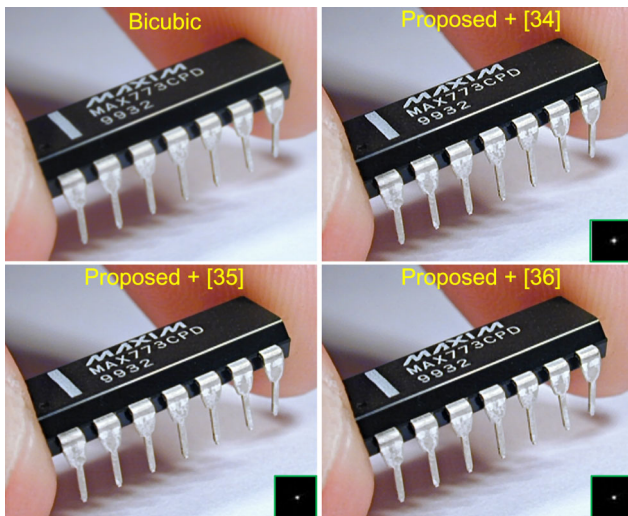


**Fig. 18** Nonparametric blind SR ( $\text{SR} \times 2$ ) with low-res *Soldier* using the proposed kernel estimation method (-NE [44], -JSC [45], -ANR [46]), Michaeli and Irani [56], and nonblind SR methods [44–46]. The first image in the *top row* is the bicubic interpolated version of the low-res image using the Matlab function `imresize`

synthetic low-res image in Fig. 15 ( $\text{SR} \times 2$ ) is of severe motion blur. We see that the SR results (Proposed + [44], Proposed + [45], Proposed + [46]) corresponding to our estimated kernels are of better visual quality than that of [56] and the state-of-the-art nonblind SR methods [44–46], showing the necessity of blur-kernel estimation for achieving high performance of SR reconstruction. Actually, [56] fails in this

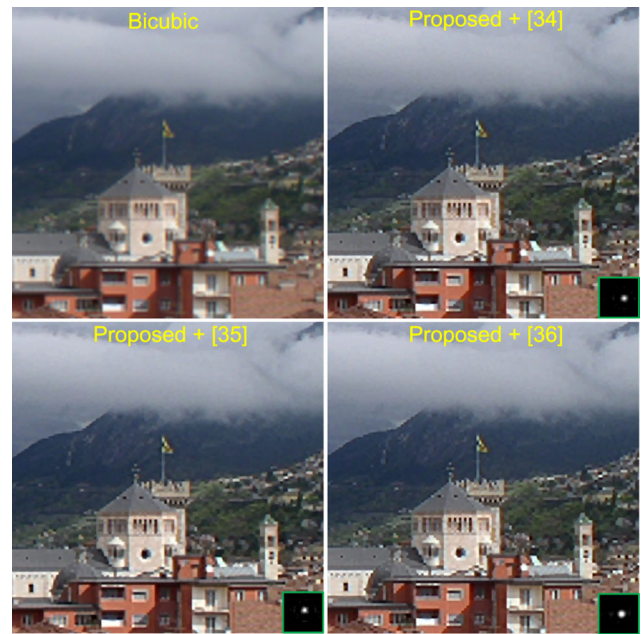


**Fig. 19** Nonparametric blind SR ( $SR \times 3$ ) with low-res *Flower* using the proposed kernel estimation method (-NE [44], -JSC [45], -ANR [46]), Michaeli and Irani [56], and nonblind SR methods [44–46]. The first image in the *top row* is the bicubic interpolated version of the low-res image using the Matlab function `imresize`



**Fig. 20** Nonparametric blind SR ( $SR \times 2$ ) with low-res *Chip* using the proposed kernel estimation method (-NE [44], -JSC [45], -ANR [46]). The first image in the *top row* is the bicubic interpolated version of the low-res image using the Matlab function `imresize`

example, producing a smooth Gaussian-like blur-kernel. The low-res images corresponding to Figs. 18, 19, 20 and 21 are real. In Fig. 18 ( $SR \times 2$ ), the proposed kernel estimation method (-NE [44], -JSC [45], -ANR [46]) still performs better, producing clearer images without notable ringing arti-



**Fig. 21** Nonparametric blind SR ( $SR \times 4$ ) with low-res *University* using the proposed kernel estimation method (-NE [44], -JSC [45], -ANR [46]). The first image in the *top row* is the bicubic interpolated version of the low-res image using the Matlab function `imresize`

facts. However, in the super-resolved image corresponding to [56], remarkable ringing artifacts can be observed, in that the produced blur-kernel by [56] has a larger support than the true one [52]. We also see that the proposed kernel estimation method is robust to nonblind SR methods to some degree, although it is empirically shown in [46] that ANR performs comparatively or better than JSC on average while NE is worse than JSC in the term of reconstruction quality. In Fig. 19 ( $SR \times 3$ ), the proposed kernel estimation method achieves comparable performance to [56], in spite of that they produce visually different kernels, and their SR images are naturally better than those of nonblind SR methods. In Fig. 20 ( $SR \times 2$ ) and Fig. 21 ( $SR \times 4$ ), another two groups of nonparametric blind SR results are provided for the proposed kernel estimation method (-NE [44], -JSC [45], -ANR [46]), validating its effectiveness, robustness, and reasonableness. In the above mentioned experiments, the bicubic interpolation images are also provided for comparison.

## 6 Conclusion

This paper provides a unified optimization perspective to single/multi-observation blur-kernel estimation, and therefore bridges the gap between the two highly related problems. In specific, blur-kernel estimation, either single-shot or multi-observation, is formulated into an  $\ell_{0.5}$ -norm-regularized negative log-marginal-likelihood minimization problem, which

couples the VB and MAP principles, and imposes a universal, three-layer hierarchical prior on the latent sharp image and a Gamma hyper-prior on each inverse noise variance. By borrowing ideas of EM, MM, MFA, and IRLS, all the unknowns of interest, including the sharp image, each inverse noise variance and blur-kernel, and other relevant model parameters are estimated automatically. The experimental results on Levin et al.'s [8] benchmark real-world motion-blurred images as well as the simulated multiple-blurred observations validate the effectiveness and superiority of the proposed kernel estimation approach for camera-shake deblurring. Our kernel estimation approach is also applied to nonparametric single-image blind SR, achieving comparable or even better performance than the state-of-the-art method [56] recently proposed by Michaeli and Irani.

**Acknowledgments** Many thanks are given to the anonymous reviewers for their insightful comments which help to significantly strengthen the paper. The first author Wen-Ze Shao is very grateful to Prof. Michael Elad for the financial support allowing him to conduct Postdoc research at the Computer Science Department, Technion-Israel Institute of Technology. He also thanks Prof. Michael Zibulevsky from Technion for his help with the numerical implementation of the paper. In addition, many thanks are given to Prof. Yi-Zhong Ma and Dr. Min Wu for their encouraging supports in the past years, as well as Mr. Ya-Tao Zhang and other kind people for helping him through his lost and sad years. This work is supported partially by the National Natural Science Foundation (NSF) of China (Grant No. 61402239), the NSF of Jiangsu Province (Grant No. BK20130868), the NSF for Jiangsu Higher Education Institutions (Grant Nos. 13KJB510022 and 13KJB120005), and the Jiangsu Key Laboratory of Image and Video Understanding for Social Safety (Nanjing University of Science and Technology, Grant No. 30920140122007).

**Appendix 1: Derivation of the posterior distributions  $q(\zeta_i)$  and  $q(\mathbf{u}_d)$**

**Estimating Posterior Distribution  $q(\zeta_i)$**

In order to derive the posterior distribution  $q(\zeta_i)$ ,  $-F(\tilde{q}, \{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}})$  is decomposed with respect to  $q(\zeta_i)$  as follows:

$$\begin{aligned}
 & -F(\tilde{q}, \{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}}) \\
 &= -\int \tilde{q} \log \left( \frac{\prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} p(\mathbf{y}_{i|d}, \Upsilon_d, \zeta_i; \Theta_d, \mathbf{k}_i)}{\tilde{q}} \right) \\
 & \quad d\{\Upsilon_d\}_{d \in \Lambda} d\{\zeta_i\}_{i \in \mathcal{U}} \\
 &= -\int \tilde{q} \log \left( \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} p(\mathbf{y}_{i|d}, \Upsilon_d, \zeta_i; \Theta_d, \mathbf{k}_i) \right) \\
 & \quad d\{\Upsilon_d\}_{d \in \Lambda} d\{\zeta_i\}_{i \in \mathcal{U}} + \int \tilde{q} \log \tilde{q} d\{\Upsilon_d\}_{d \in \Lambda} d\{\zeta_i\}_{i \in \mathcal{U}}
 \end{aligned}$$

$$\begin{aligned}
 & \triangleq -\int \prod_{d \in \Lambda} q(\zeta_i) \log \tilde{p}(\{\mathbf{y}_{i|d}\}_{d \in \Lambda}, \zeta_i; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i) d\zeta_i \\
 & + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\zeta_i) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\zeta_i) d\zeta_i \\
 & + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\mathbf{u}_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\mathbf{u}_d) d\{\mathbf{u}_d\}_{d \in \Lambda} \\
 & + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\Upsilon_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\Upsilon_d) d\{\Upsilon_d\}_{d \in \Lambda} \\
 & + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\beta_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\beta_d) d\{\beta_d\}_{d \in \Lambda} \\
 & = -\int \prod_{d \in \Lambda} q(\zeta_i) \log \tilde{p}(\{\mathbf{y}_{i|d}\}_{d \in \Lambda}, \zeta_i; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i) d\zeta_i \\
 & + \int \prod_{d \in \Lambda} q(\zeta_i) \log \prod_{d \in \Lambda} q(\zeta_i) d\zeta_i \\
 & + \int \prod_{i' \in \mathcal{U} \setminus i} \prod_{d \in \Lambda} q(\zeta_{i'}) \log \prod_{i' \in \mathcal{U} \setminus i} \prod_{d \in \Lambda} q(\zeta_{i'}) d\zeta_{i'} \\
 & + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\mathbf{u}_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\mathbf{u}_d) d\{\mathbf{u}_d\}_{d \in \Lambda} \\
 & + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\Upsilon_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\Upsilon_d) d\{\Upsilon_d\}_{d \in \Lambda} \\
 & + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\beta_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\beta_d) d\{\beta_d\}_{d \in \Lambda} \\
 & = \text{KL} \left( \prod_{d \in \Lambda} q(\zeta_i) \parallel \tilde{p}(\{\mathbf{y}_{i|d}\}_{d \in \Lambda}, \zeta_i; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i) \right) \\
 & + \int \prod_{i' \in \mathcal{U} \setminus i} \prod_{d \in \Lambda} q(\zeta_{i'}) \log \prod_{i' \in \mathcal{U} \setminus i} \prod_{d \in \Lambda} q(\zeta_{i'}) d\zeta_{i'} \\
 & + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\mathbf{u}_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\mathbf{u}_d) d\{\mathbf{u}_d\}_{d \in \Lambda} \\
 & + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\Upsilon_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\Upsilon_d) d\{\Upsilon_d\}_{d \in \Lambda} \\
 & + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\beta_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\beta_d) d\{\beta_d\}_{d \in \Lambda}, \tag{38}
 \end{aligned}$$

where

$$\begin{aligned}
 & \log \tilde{p}(\{\mathbf{y}_{i|d}\}_{d \in \Lambda}, \zeta_i; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i) \\
 &= \int \prod_{i' \in \mathcal{U} \setminus i} \prod_{d \in \Lambda} q(\zeta_{i'}) q(\mathbf{u}_d) q(\Upsilon_d) q(\beta_d) \\
 & \quad \log \left( \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} p(\mathbf{y}_{i|d}, \Upsilon_d, \zeta_i; \Theta_d, \mathbf{k}_i) \right) \\
 & \quad d\{\Upsilon_d\}_{d \in \Lambda} d\{\zeta_{i'}\}_{i' \in \mathcal{U} \setminus i}. \tag{39}
 \end{aligned}$$

It is seen that,  $-F(\tilde{q}, \{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}})$  is minimized as the KL divergence  $KL(\prod_{d \in \Lambda} q(\zeta_i) || \tilde{p}(\{y_{i|d}\}_{d \in \Lambda}, \zeta_i; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i))$  equals zero, i.e.,  $\prod_{d \in \Lambda} q(\zeta_i) = \tilde{p}(\{y_{i|d}\}_{d \in \Lambda}, \zeta_i; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i)$ , leading to

$$\begin{aligned}
 |\Lambda| \log q(\zeta_i) &= \sum_{d \in \Lambda} \log q(\zeta_i) \\
 &= \log \tilde{p}(\{y_{i|d}\}_{d \in \Lambda}, \zeta_i; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i) \\
 &= \left\langle \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} p(y_{i|d}, \Upsilon_d, \zeta_i; \Theta_d, \mathbf{k}_i) \right\rangle_{\{q(\zeta_{i'})\}_{i' \in \mathcal{U}}, \{q(\mathbf{u}_d)\}_{d \in \Lambda}, \{q(\gamma_d)\}_{d \in \Lambda}, \{q(\beta_d)\}_{d \in \Lambda}} \\
 &= \left\{ \frac{M}{2} \sum_{d \in \Lambda} \log \zeta_i - \frac{\zeta_i}{2} \sum_{d \in \Lambda} \langle \|\mathbf{y}_{i|d} - \mathbf{K}_i \mathbf{u}_d\|_2^2 \rangle_{q(\mathbf{u}_d)} + \sum_{d \in \Lambda} \log((a-1) \log \zeta_i - b_{\zeta_i}) \right\} + const. \tag{40}
 \end{aligned}$$

Therefore, the posterior distribution  $q(\zeta_i)$  is actually a Gamma PDF  $\mathcal{G}a(\zeta_i | a_{\zeta_i}, b_{\zeta_i})$ , where the shape and rate parameters are defined as

$$a_{\zeta_i} = \frac{M}{2} + a, \tag{41}$$

$$b_{\zeta_i} = \frac{1}{2|\Lambda|} \sum_{d \in \Lambda} \langle \|\mathbf{y}_{i|d} - \mathbf{K}_i \mathbf{u}_d\|_2^2 \rangle_{q(\mathbf{u}_d)} + b, \tag{42}$$

and its mean is given by  $\langle \zeta_i \rangle_{q(\zeta_i)} = \frac{a_{\zeta_i}}{b_{\zeta_i}}$ .

**Estimating Posterior Distribution  $q(\mathbf{u}_d)$**

In order to derive  $q(\mathbf{u}_d)$ ,  $-F(\tilde{q}, \{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}})$  is decomposed with respect to  $q(\mathbf{u}_d)$  as follows:

$$\begin{aligned}
 &-F(\tilde{q}, \{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}}) \\
 &= - \int \tilde{q} \log \left( \frac{\prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} p(y_{i|d}, \Upsilon_d, \zeta_i; \Theta_d, \mathbf{k}_i)}{\tilde{q}} \right) \\
 &\quad d\{\Upsilon_d\}_{d \in \Lambda} d\{\zeta_i\}_{i \in \mathcal{U}} \\
 &= - \int \tilde{q} \log \left( \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} p(y_{i|d}, \Upsilon_d, \zeta_i; \Theta_d, \mathbf{k}_i) \right) \\
 &\quad d\{\Upsilon_d\}_{d \in \Lambda} d\{\zeta_i\}_{i \in \mathcal{U}} + \int \tilde{q} \log \tilde{q} d\{\Upsilon_d\}_{d \in \Lambda} d\{\zeta_i\}_{i \in \mathcal{U}} \\
 &\triangleq - \int \prod_{i \in \mathcal{U}} q(\mathbf{u}_d) \log \tilde{p}(\{y_{i|d}\}_{d \in \Lambda}, \mathbf{u}_d; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i) d\mathbf{u}_d \\
 &\quad + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\mathbf{u}_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\mathbf{u}_d) d\{\mathbf{u}_d\}_{d \in \Lambda} \\
 &\quad + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\zeta_i) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\zeta_i) d\zeta_i
 \end{aligned}$$

$$\begin{aligned}
 &+ \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\gamma_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\gamma_d) d\{\gamma_d\}_{d \in \Lambda} \\
 &+ \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\beta_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\beta_d) d\{\beta_d\}_{d \in \Lambda}
 \end{aligned}$$

$$\begin{aligned}
 &= - \int \prod_{i \in \mathcal{U}} q(\mathbf{u}_d) \log \tilde{p}(\{y_{i|d}\}_{d \in \Lambda}, \mathbf{u}_d; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i) d\mathbf{u}_d \\
 &\quad + \int \prod_{i \in \mathcal{U}} q(\mathbf{u}_d) \log \prod_{i \in \mathcal{U}} q(\mathbf{u}_d) d\mathbf{u}_d \\
 &\quad + \int \prod_{i \in \mathcal{U}} \prod_{d' \in \Lambda \setminus d} q(\mathbf{u}_{d'}) \log \prod_{i \in \mathcal{U}} \prod_{i \in \mathcal{U}} q(\mathbf{u}_{d'}) d\mathbf{u}_{d'} \\
 &\quad + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\zeta_i) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\zeta_i) d\zeta_i \\
 &\quad + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\gamma_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\gamma_d) d\{\gamma_d\}_{d \in \Lambda} \\
 &\quad + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\beta_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\beta_d) d\{\beta_d\}_{d \in \Lambda} \\
 &= KL \left( \prod_{i \in \mathcal{U}} q(\mathbf{u}_d) || \tilde{p}(\{y_{i|d}\}_{d \in \Lambda}, \mathbf{u}_d; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i) \right) \\
 &\quad + \int \prod_{i \in \mathcal{U}} \prod_{d' \in \Lambda \setminus d} q(\mathbf{u}_{d'}) \log \prod_{i \in \mathcal{U}} \prod_{i \in \mathcal{U}} q(\mathbf{u}_{d'}) d\mathbf{u}_{d'} \\
 &\quad + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\zeta_i) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\zeta_i) d\zeta_i \\
 &\quad + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\gamma_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\gamma_d) d\{\gamma_d\}_{d \in \Lambda} \\
 &\quad + \int \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\beta_d) \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\beta_d) d\{\beta_d\}_{d \in \Lambda} \tag{43}
 \end{aligned}$$

where

$$\begin{aligned}
 &\log \tilde{p}(\{y_{i|d}\}_{d \in \Lambda}, \mathbf{u}_d; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i) \\
 &= \int \prod_{i \in \mathcal{U}} \prod_{d' \in \Lambda \setminus d} q(\mathbf{u}_{d'}) \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} q(\zeta_i) q(\gamma_d) q(\beta_d)
 \end{aligned}$$

$$\log \left( \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} p(\mathbf{y}_{i|d}, \Upsilon_d, \zeta_i; \Theta_d, \mathbf{k}_i) \right) d\{\mathbf{u}_{d'}\}_{d' \in \Lambda \setminus d} d\{\boldsymbol{\gamma}_d\}_{d \in \Lambda} d\{\boldsymbol{\beta}_d\}_{d \in \Lambda} d\{\zeta_i\}_{i \in \mathcal{U}}. \quad (44)$$

It is seen that,  $-F(\tilde{q}, \{\Theta_d\}_{d \in \Lambda}, \{\mathbf{k}_i\}_{i \in \mathcal{U}})$  is minimized as the KL divergence  $\text{KL} \left( \prod_{i \in \mathcal{U}} q(\mathbf{u}_d) \parallel \tilde{p}(\{\mathbf{y}_{i|d}\}_{d \in \Lambda}, \mathbf{u}_d; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i) \right)$  equals zero, i.e.,  $\prod_{i \in \mathcal{U}} q(\mathbf{u}_d) = \tilde{p}(\{\mathbf{y}_{i|d}\}_{d \in \Lambda}, \mathbf{u}_d; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i)$ , leading to

$$\begin{aligned} |\mathcal{U}| \log q(\mathbf{u}_d) &= \sum_{i \in \mathcal{U}} \log q(\mathbf{u}_d) \\ &= \log \tilde{p}(\{\mathbf{y}_{i|d}\}_{d \in \Lambda}, \mathbf{u}_d; \{\Theta_d\}_{d \in \Lambda}, \mathbf{k}_i) \\ &= \left\langle \log \prod_{i \in \mathcal{U}} \prod_{d \in \Lambda} p(\mathbf{y}_{i|d}, \Upsilon_d, \zeta_i; \Theta_d, \mathbf{k}_i) \right\rangle_{\{q(\zeta_i)\}_{i \in \mathcal{U}} \{q(\mathbf{u}_{d'})\}_{d' \in \Lambda \setminus d} \{q(\boldsymbol{\gamma}_d)\}_{d \in \Lambda} \{q(\boldsymbol{\beta}_d)\}_{d \in \Lambda}} \\ &= -\frac{1}{2} \left\{ \sum_{i \in \mathcal{U}} \langle \zeta_i \rangle_{q(\zeta_i)} \|\mathbf{y}_{i|d} - \mathbf{K}_i \mathbf{u}_d\|_2^2 + \sum_{i \in \mathcal{U}} \mathbf{u}_d^T \text{diag}\{\langle \boldsymbol{\gamma}_d \rangle_{q(\boldsymbol{\gamma}_d)}\} \mathbf{u}_d \right\} + \text{const}. \end{aligned} \quad (45)$$

Therefore, the posterior distribution  $q(\mathbf{u}_d)$  is a multivariate Gaussian PDF  $\mathcal{N}(\mathbf{u}_d | \boldsymbol{\mu}_d, \mathbf{C}_d)$ , and its mean  $\boldsymbol{\mu}_d = \langle \mathbf{u}_d \rangle_{q(\mathbf{u}_d)}$  and covariance matrix  $\mathbf{C}_d$  are, respectively, defined as

$$\boldsymbol{\mu}_d = \mathbf{C}_d \sum_{i \in \mathcal{U}} \langle \zeta_i \rangle_{q(\zeta_i)} \mathbf{K}_i^T \mathbf{y}_{i|d}, \quad (46)$$

$$\mathbf{C}_d = \left[ |\mathcal{U}| \text{diag}\{\langle \boldsymbol{\gamma}_d \rangle_{q(\boldsymbol{\gamma}_d)}\} + \sum_{i \in \mathcal{U}} \langle \zeta_i \rangle_{q(\zeta_i)} \mathbf{K}_i^T \mathbf{K}_i \right]^{-1}. \quad (47)$$

### Appendix 2: Representative Dictionary-Based Super-Resolution (SR) Methods

Three representative approaches of dictionary-based single frame fast SR are introduced for reference in Sect. 5.2, i.e., neighborhood embedding (NE) of Chang et al. [44], joint sparse coding (JSC) of Zeyde et al. [45], and anchored neighbor regression (ANR) of Timofte et al. [46]. These three nonblind SR methods all assume the simplest bicubic interpolation kernel in the observation model and are harnessed to generate a super-resolved but blurred image, used as the input of the proposed blur-kernel estimation approach for practicing nonparametric blind SR.

### Neighborhood Embedding

The NE algorithm [44] is a manifold learning approach assuming that image patches of a low-res image and its high-res counterpart form manifolds with similar local geometry in two distinct feature spaces. It roughly means that, as long as enough sampled patches are available, patches in the high-res feature space can be re-constructed as a weighted average of local neighbors using the same weights as in the low-res feature space. In practice, weights are computed by solving a constrained least squares problem.

### Joint Sparse Coding

The JSC algorithm originates from Yang et al. [42], which is improved by Zeyde et al. [45] in both SR restoration quality and execution speed. The core idea of JSC is akin to that of NE. What is distinct to NE is that, the low-res and high-res manifolds in the training step are not the sampled patches but the jointly learned compact dictionaries for the sampled low-res and high-res patches, in order to achieve the same sparse coding for low-res patches as their corresponding high-res patches. Here, the entries of sparse coding play a similar role to the weights in NE. The test high-res patch can be reconstructed by exactly the same sparse coding of its counterpart low-res patch and the learnt high-res dictionary.

### Anchored Neighbor Regression

Combining ideas of NE and JSC, the ANR algorithm [46] achieves comparative or higher quality and one or two orders of magnitude efficiency improvements over the state-of-the-art methods. The ANR is essentially an intimate approximation of the NE based on the JSC, lying in that the low-res and high-res manifolds in the training step are the jointly learned compact dictionaries for the sampled low-res and high-res image patches, just the same as the JSC, but the weights for a test low-res patch are no longer the immedi-

ate goals, instead a separate projection matrix to be stored is computed for each atom in the low-res dictionary by Ridge Regression and the neighborhood of the atom. The high-res patch can then be reconstructed using its low-res counterpart and the stored projection matrix corresponding to the atom computed as the nearest neighbor to the test low-res patch.

## References

- Fergus, R., Singh, B., Hertzmann, A., Roweis, S.T., Freeman, W.T.: Removing camera shake from a single photograph. *ACM Trans. Graph.* **25**(3), 787–794 (2006)
- Sroubek, F., Milanfar, P.: Robust multichannel blind deconvolution via fast alternating minimization. *IEEE Trans. Image Process.* **21**(4), 1687–1700 (2012)
- Zhu, X., Milanfar, P.: Restoration for weakly blurred and strongly noisy images. In: *IEEE Workshop on Applications of Computer Vision*, pp. 103–109 (2011)
- Takeda, H., Milanfar, P.: Removing motion blur with space-time processing. *IEEE Trans. Image Process.* **20**(10), 2990–3000 (2011)
- Zhu, X., Cohen, S., Schiller, S., Milanfar, P.: Estimating spatially varying defocus blur from a single image. *IEEE Trans. Image Process.* **22**(12), 4879–4891 (2013)
- Tai, Y.-W., Tan, P., Brown, M.S.: Richardson-Lucy deblurring for scenes under a projective motion path. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(8), 1603–1618 (2011)
- Tai, Y.-W.: Motion-aware noise filtering for deblurring of noisy and blurry images. *IEEE CVPR*, pp. 17–24 (2012)
- Levin, A., Weiss, Y., Durand, F., Freeman, W.T.: Understanding blind deconvolution algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(12), 2354–2367 (2011)
- Molina, R., Mateos, J., Katsaggelos, A.K.: Blind deconvolution using a variational approach to parameter, image, and blur estimation. *IEEE Trans. Image Process.* **15**(12), 3715–3727 (2006)
- Babacan, S.D., Molina, R., Katsaggelos, A.K.: Variational Bayesian blind deconvolution using a total variation prior. *IEEE Trans. Image Process.* **18**(1), 12–26 (2009)
- Amizic, B., Molina, R., Katsaggelos, A.K.: Sparse Bayesian blind image deconvolution with parameter estimation. *EURASIP J. Image Video Process.* **2012–20**, 1–15 (2012)
- Tzikas, D.G., Likas, A.C., Galatsanos, N.P.: The variational approximation for Bayesian inference: life after the EM algorithm. *IEEE Signal Process. Mag.* **131**, 131–146 (2008)
- Mohammad-Djafari, A.: Bayesian approach with prior models which enforce sparsity in signal and image processing. *EURASIP J. Adv. Signal Process. (Special issue on Sparse Signal Processing)* **2012–52**, 1–19 (2012)
- Mohammad-Djafari, A., Knuth, K.H.: Bayesian approaches. In: Comon, P., Jutten, C., et al. (eds.) *Handbook of Blind Source Separation: ICA and Applications*, pp. 467–513. Elsevier, New York (2010)
- Wipf, D.P., Zhang, H.: Analysis of Bayesian blind deconvolution. *EMMCVPR. LNCS*, vol. 8081, pp. 40–53. Springer, Heidelberg (2013)
- Babacan, S.D., Molina, R., Do, M.N., Katsaggelos, A.K.: Bayesian blind deconvolution with general sparse image priors. *ECCV. LNCS*, vol. 7577, pp. 341–355. Springer, Heidelberg (2012)
- Levin, A., Weiss, Y., Durand, F., Freeman, W.T.: Efficient marginal likelihood optimization in blind deconvolution. In: *CVPR*, pp. 2657–2664 (2011)
- Krishnan, D., Tay, T., Fergus, R.: Blind deconvolution using a normalized sparsity measure. In: *CVPR* pp. 233–240 (2011)
- Almeida, M., Almeida, L.: Blind and semi-blind deblurring of natural images. *IEEE Trans. Image Process.* **19**(1), 36–52 (2010)
- Xu, L., Zheng, S., Jia, J.: Unnatural  $L_0$  sparse representation for natural image deblurring. In: *CVPR*, pp. 1107–1114 (2013)
- Cho, S., Lee, S.: Fast motion deblurring. *ACM Trans. Graph.* **28**(5): article no. 145 (2009)
- Shan, Q., Jia, J., Agarwala, A.: High-quality motion deblurring from a single image. In: *SIGGRAPH, ACM, New York, NY, USA*, pp. 1–10 (2008)
- Xu, L., Jia, J.: Two-phase kernel estimation for robust motion deblurring. *ECCV. LNCS*, vol. 6311, pp. 157–170. Springer, Heidelberg (2010)
- Rav-Acha, A., Peleg, S.: Two motion blurred images are better than one. *Pattern Recog. Lett.* **26**, 311–317 (2005)
- Cai, J.-F., Ji, H., Liu, C., Shen, Z.: Blind motion deblurring using multiple images. *J. Comput. Phys.* **228**, 5057–5071 (2009)
- Zhu, X., Sroubek, F., Milanfar, P.: Deconvolving PSFs for a better motion deblurring using multiple images. *ECCV. LNCS*, pp. 636–647. Springer, Heidelberg (2012)
- Chen, J., Yuan, L., Tang, C.-K., Quan, L.: Robust dual motion deblurring. In: *IEEE CVPR*, pp. 1–8 (2008)
- Harikumar, G., Bresler, Y.: Perfect blind restoration of images blurred by multiple filters: theory and efficient algorithms. *IEEE Trans. Image Process.* **8**(2), 202–219 (1999)
- Zhang, H., Wipf, D., Zhang, Y.: Multi-observation blind deconvolution with an adaptive sparse prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(8), 1628–1643 (2014)
- Yuan, L., Sun, J., Quan, L., Shum, H.-Y.: Image deblurring with blurred/noisy image pairs. *ACM Trans. Graph.* **26**(3), 1–10 (2007)
- Babacan, S.D., Wang, J., Molina, R., Katsaggelos, A.K.: Bayesian blind deconvolution from differently exposed image pairs. *IEEE Trans. Image Process.* **19**(11), 2874–2888 (2009)
- Figueiredo, M., Nowak, R.: A bound optimization approach to wavelet-based image deconvolution. In: *ICIP* vol. 2, pp. 782–785 (2005)
- Freeman, W.T., Pasztor, E.C.: Learning to estimate scenes from images. In: Kearns, M.S., Solla, S.A., Cohn, D.A. (eds.) *Adv. Neural Information Processing Systems*, Vol. 11, Cambridge University Press, Cambridge. See also <http://www.merl.com/reports/TR99-05/> (1999)
- Baker, S., Kanade, T.: Hallucinating faces. In: *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 83–88 (2000)
- Nasrollahi, K., Moeslund, T.B.: Super-resolution: a comprehensive survey. *Mach. Vis. Appl.* **25**(6), 1423–1468 (2014)
- Mudenagudi, U., Singla, R., Kalra, P., Banerjee.: Super-resolution using graph-cut. In: *ACCV*, pp. 385–394 (2006)
- Mansouri, M., Mohammad-Djafari, A.: Joint super-resolution and segmentation from a set of low resolution images using a Bayesian approach with a Gauss-Markov-Potts prior. *Int. J. Signal and Imaging. Syst. Eng.* **3**(4), 211–221 (2010)
- Farsiu, S., Robinson, D., Elad, M., Milanfar, P.: Advances and challenges in super-resolution. *Int. J. Imaging Syst. Technol.* **14**(2), 47–57 (2004)
- Takeda, H., Milanfar, P., Protter, M., Elad, M.: Super-resolution without explicit subpixel motion estimation. *IEEE Trans. Image Process.* **18**(9), 1958–1975 (2009)
- Protter, M., Elad, M., Takeda, H., Milanfar, P.: Generalizing the nonlocal-means to super-resolution reconstruction. *IEEE Trans. Image Process.* **18**(1), 36–51 (2009)
- Shen, H., Zhang, L., Huang, B., Li, P.: A MAP approach for joint motion estimation, segmentation, and super resolution. *IEEE Trans. Image Process.* **16**(2), 479–490 (2007)
- Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution via sparse representation. *IEEE TIP* **19**(11), 2861–2873 (2010)

43. Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: ICCV, pp. 349–356 (2009)
44. Chang, H., Yeung, D.-Y., Xiong, Y.: Super-resolution through neighbor embedding. In: CVPR, pp. 275–282 (2004)
45. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. Curves and Surfaces. LNCS, vol. 6920, pp. 711–730. Springer, Heidelberg (2012)
46. Timofte, R., Smet, V.D., Gool, L.V.: Anchored neighborhood regression for fast example-based super-resolution. In: ICCV, pp. 1920–1927 (2013)
47. Timofte, R., Smet, V.D., Gool, L.V.: A+: Adjusted anchored neighborhood regression for fast super-resolution. In: ECCV, pp. 1–5 (2014)
48. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: ECCV (2014)
49. Peleg, T., Elad, M.: A statistical prediction model based on sparse representations for single image super-resolution. IEEE Trans. Image Process. **23**(6), 2569–2582 (2014)
50. Yang, J., Lin, Z., Cohen, S.: Fast image super-resolution based on in-place example regression. In: CVPR, pp. 1059–1066 (2013)
51. Fattal, R.: Image upsampling via imposed edge statistics. ACM Trans. Graph. (SIGGRAPH) **26**(3), 95–102 (2007)
52. Efrat, N., Glasner, D., Apartsin, A., Nadler, B., Levin, A.: Accurate blur models vs. image priors in single image super-resolution. In: ICCV (2013)
53. Begin, I., Ferrie, F.R.: PSF recovery from examples for blind super-resolution. In: ICIP vol. 5, pp. 421–424 (2007)
54. Wang, Q., Tang, X., Shum, H.: Patch based blind image super resolution. In: ICCV, pp. 709–716 (2005)
55. Joshi, N., Szeliski, R., Kriegman, D.J.: PSF estimation using sharp edge prediction. In: CVPR, pp. 1–8 (2008)
56. Michaeli, T., Irani, M.: Nonparametric blind super-resolution. In: ICCV, pp. 945–952 (2013)
57. Shao, W.-Z., Ge, Q., Deng, H.-S., Wei, Z.-H., Li, H.-B.: Motion deblurring using non-stationary image modeling. J. Math. Imaging Vis. (2014). doi:[10.1007/s10851-014-0537-9](https://doi.org/10.1007/s10851-014-0537-9)
58. Babacan, S.D., Molina, R., Katsaggelos, A.K.: Bayesian compressive sensing using Laplace priors. IEEE Trans. Image Process. **19**(1), 53–63 (2010)
59. Griffin, J.E., Brown, P.J.: Bayesian adaptive Lassos with non-convex penalization. Aust. NZ. J. Stat. **53**, 423–442 (2012)
60. Tipping, M.E.: Sparse Bayesian learning and the relevance vector machine. J. Mach. Learn. Res. **1**, 211–244 (2001)
61. Krishnan, D., Fergus, R.: Fast image deconvolution using hyper-Laplacian priors. In: NIPS vol. 22, pp. 1033–1041 (2009)
62. Kheradmand, A., Milanfar, P.: A general framework for regularized, similarity-based image restoration. IEEE Trans. Image Process. **23**(12), 5136–5151 (2014)
63. Takeda, H., Farsiu, S., Milanfar, P.: Deblurring using regularized locally-adaptive kernel regression. IEEE Trans. Image Process. **17**(4), 550–563 (2008)
64. Schmidt, U., Rother, C., Nowozin, S., Jancsary, J.: Discriminative non-blind deblurring. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 23–28 (2013)
65. Schuler, C. J., Burger, H. C., Harmeling, S., Schölkopf, B.: A machine learning approach for non-blind image deconvolution. In: IEEE Conference Computer Vision and Pattern Recognition, pp. 1067–1074 (2013)
66. Danielyan, A., Katkovnik, V., Egiazarian, K.: BM3D frames and variational image deblurring. IEEE Trans. Image Process. **21**(4), 1715–1728 (2012)
67. Kotera, J., Sroubek, F., Milanfar, P.: Blind deconvolution using alternating maximum a posteriori estimation with heavy-tailed priors. CAIP, Part II, LNCS, vol. 8048, pp. 59–66. Springer, Heidelberg (2013)
68. Levin, A., Fergus, R., Durand, F., Freeman, W.T.: Image and depth from a conventional camera with a coded aperture. ACM Trans. Graph. (SIGGRAPH) **26**(3), 701–709 (2007)



**Wen-Ze Shao** was born in Ganyu, Jiangsu Province, China. He completed the B.S. degree in Science of Information and Computation in June 2003, and the Ph.D. degree in Pattern Recognition and Intelligent System in July 2008, both from Nanjing University of Science and Technology (NUST), Nanjing, China. From June 2003 to December 2011, he served as an officer in the People's Liberation Army (PLA) of China. In January 2012, he joined Nanjing

University of Posts and Telecommunications (NUPT) as an assistant professor. Since May 2014, he also worked as a visiting scientist (Post-doc) at the Department of Computer Science in Technion-Israel Institute of Technology. He is currently interested in the fields of overcomplete sparse representation, statistical learning, and variational PDE methods with applications to innovative and interactive imaging and vision problems.



**Qi Ge** was born in Nantong, Jiangsu Province, China. She completed the B.S. degree in Science of Information and Computation in 2006 and the M.S. degree in Applied Mathematics in 2009, both from Nanjing University of Information and Engineering, Nanjing, China, and the Ph.D. degree in Pattern Recognition and Intelligent System in 2013 from NUST, Nanjing, China. Now, she serves as an assistant professor at College of Telecommunications and Information

Engineering at NUPT. She is particularly interested in the fields of variational PDE approaches, probabilistic graphical models, sparse representation and their applications to image restoration, segmentation, and so on.



**Hai-Song Deng** was born in Xuzhou, Jiangsu Province, China. She completed the B.S. degree in Mathematics in 2003 from Xuzhou Normal University, Xuzhou, China, the M.S. degree in Statistics in 2006, and the Ph.D. degree in Management Science and Engineering in 2011 both from NUST, Nanjing, China. Since 2011, she served as an assistant professor at School of Science in Nanjing Audit University. Her research

fields include Bayesian variable selection, sparse optimization, surrogate modeling, design and analysis of computer experiments, and so on.



**Zhi-Hui Wei** was born in Huai'an, Jiangsu Province, China. He completed the B.S. degree in 1983 and the M.S. degree in 1986, both in Mathematics, and the Ph.D. degree in Electrical Engineering in 2003, all from Southeast University, Nanjing, China. Now, he is a Professor of Image Processing and Pattern Recognition, and Mathematics in NUST, Nanjing, China. He has been working on the fields of statistical image modeling, geometrical multi-scale analysis,

variational PDE, sparse representation, compressed sensing, super-resolution, image compression, and so on.



**Hai-Bo Li** was born in Xuzhou, Jiangsu Province, China. He completed the B.E. degree in Wireless Engineering in 1985 and the M.S. degree in Communication and Electronic Systems in 1988, both from NUPT, Nanjing, China, and the Ph.D. degree in Information Theory in 1993 from Lin ping University, Sweden. In 1997, he became docent (UK equivalent senior lecturer, US equivalent associate professor) in image coding, and in 1999 he became the youngest lifetime

professor of signal processing in Umea University, Sweden. Now, he is a Professor of innovative media technology in KTH Royal Institute of Technology. His research interest is mainly media signal processing, including image coding, video compression, motion estimation, facial and hand gesture recognition for the next generation of mobile phones, invisible interaction technology, and so on.