

Shift-Variant Blind Deconvolution Using a Field of Kernels*

Motoharu SONOGASHIRA^{†a)}, Nonmember, Masaaki IIYAMA^{†b)}, Member, and Michihiko MINOH^{†c)}, Fellow

SUMMARY Blind deconvolution (BD) is the problem of restoring sharp images from blurry images when convolution kernels are unknown. While it has a wide range of applications and has been extensively studied, traditional shift-invariant (SI) BD focuses on uniform blur caused by kernels that do not spatially vary. However, real blur caused by factors such as motion and defocus is often nonuniform and thus beyond the ability of SI BD. Although specialized methods exist for nonuniform blur, they can only handle specific blur types. Consequently, the applicability of BD for general blur remains limited. This paper proposes a shift-variant (SV) BD method that models nonuniform blur using a field of kernels that assigns a local kernel to each pixel, thereby allowing pixelwise variation. This concept is realized as a Bayesian model that involves SV convolution with the field of kernels and smoothing of the field for regularization. A variational-Bayesian inference algorithm is derived to jointly estimate a sharp latent image and a field of kernels from a blurry observed image. Owing to the flexibility of the field-of-kernels model, the proposed method can deal with a wider range of blur than previous approaches. Experiments using images with nonuniform blur demonstrate the effectiveness of the proposed SV BD method in comparison with previous SI and SV approaches.

key words: blind deconvolution, deblurring, shift-variant, variational Bayes

1. Introduction

Blind deconvolution (BD) for deblurring is one of the most extensively studied topics in image processing [1], [2]. Essentially, blurring of images is modeled as convolution of images and kernels. Deconvolution is the inverse of this process, which effectively restores sharp images from blurry ones. The objective of BD is to perform deconvolution when kernels are unknown, which is often the case in practice [1]. Since undesirable blur is quite common in the real world, BD has a wide range of applications, e.g., mobile photography [3], computational photography [4], computer vision [5], astronomical imaging [6], and biomedical imaging [7].

Traditional *shift-invariant* (SI) BD assumes uniform blur produced by kernels that do not vary spatially across images, such as the example in Fig. 1 (b). However, in reality blur kernels often spatially vary, producing nonuni-

form blur [1]. For example, independently moving objects produce different effects of motion blur, and objects distant from a camera are more subject to defocus blur than close objects. Blur is often more complex since both motion and defocus blur can simultaneously occur, as shown in Fig. 1 (c). In such cases, SI BD methods fail in kernel estimation and also in deconvolution [1]. Although specialized BD methods for nonuniform blur have been recently developed, they can only handle certain types of blur, e.g., motion, defocus, or locally uniform blur [2]. Consequently, the applicability of BD to general blurry images remains limited.

In this paper, we propose a *shift-variant* (SV) BD method that can handle nonuniform blur regardless of type. The idea is to model a spatially variant kernel as a *field of kernels* that assigns a local blur kernel to each image pixel. By allowing different kernels among pixels, we can flexibly represent nonuniform blur without making assumptions about its type. To alleviate ill-posedness of the SV BD problem, we also introduce smoothing of the field, which sufficiently regularizes kernel estimation without losing flexibility. Under this model, we estimate both a sharp latent image and the field of kernels from a blurry image using techniques of *variational Bayes* (VB). As demonstrated through experiments, the flexibility of the proposed field-of-kernels model enables BD to be applied to complex blur beyond the ability of previous SI and SV approaches.

The rest of this paper is organized as follows. First, we review previous work related to BD in Sect. 2. Next, we describe the construction of a Bayesian model with a field of kernels in Sect. 3, and then develop an SV BD algorithm using VB techniques. Numerical implementation of the algorithm is presented in Sect. 5. In Sect. 6, we show experimental results to demonstrate the effectiveness of our

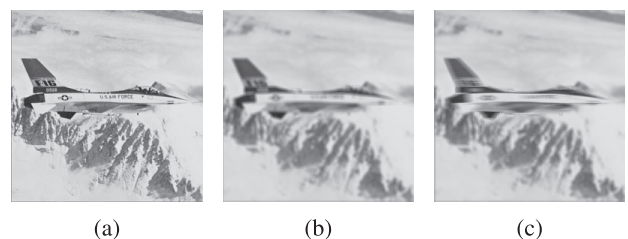


Fig. 1 (a) A sharp image as a reference. (b) Uniform blur assumed by SI BD, where both the foreground and the background are affected by the same blur. (c) Nonuniform blur assumed by SV BD, where the foreground and the background are affected by motion and defocus blur, respectively.

Manuscript received December 16, 2016.

Manuscript revised April 14, 2017.

Manuscript publicized June 14, 2017.

[†]The authors are with Kyoto University, Kyoto-shi, 606–8501 Japan.

*This work was supported by JSPS KAKENHI Grant Number JP16K12463.

a) E-mail: sonogashira@mm.media.kyoto-u.ac.jp

b) E-mail: iiyama@mm.media.kyoto-u.ac.jp

c) E-mail: minoh@mm.media.kyoto-u.ac.jp

DOI: 10.1587/transinf.2016PCP0013

method. Finally, we conclude this paper with comments on future work in Sect. 7.

2. Related Work

2.1 Shift-Invariant and Variant Blind Deconvolution

Throughout the history of image processing, a large body of literature has been devoted to BD. Generally, SI BD assumes globally uniform blur caused by a kernel that does not vary between pixels. While this assumption has facilitated development of various BD methods by simplifying convolution models, it only holds for limited instances of blur in reality [1]. When the assumption is violated, SI BD cannot accurately estimate kernels, and often produces ringing artifacts that even degrade the quality of blurry images [8].

Recently, SV BD has been investigated to deal with nonuniform blur. One approach to the SV BD problem is segmentation-based BD [9]–[13], which performs SI BD in each image segment, assuming locally uniform blur. Although this assumption is more realistic than that of SI BD, blur is often more complex and nonuniform even in each segment. For example, motion involving rotation and defocus on surfaces nonparallel to an image plane are typical sources of smoothly varying blur [1], [14], [15]. Such blur is difficult to handle for segmentation-based BD since smooth variation cannot be approximated well with the combination of a small number of kernels. While interpolation can be used to describe smooth variation [16], its accuracy still depends on the selection of kernels as samples. Another problem with segmentation-based BD is that segmentation itself becomes inaccurate when edges are corrupted by blur [14]. Although soft-segmentation alleviates this limitation by ambiguously assigning multiple kernels to each pixel [8], [17], [18], it still has difficulty in approximating blur due to non-rigid motion and varying defocus [14]. In addition, the number of segments, which is generally not known, is usually required to be specified for each image as a parameter in segmentation-based BD [14]. Compared with segmentation-based models that allow variation of kernels only between segments, the proposed field-based model is more fine-grained and can describe kernels that vary between pixels. Therefore, the proposed method can deal with a wider range of nonuniform blur, including smoothly varying blur while avoiding difficulties in segmentation.

Another approach to SV BD is to parameterize kernels with a small number of parameters, assuming specific blur types, e.g., motion blur [6], [11], [19]–[23] or defocus blur [15], [24]–[27]. A drawback of such methods is that they can only handle targeted blur types. For example, methods designed for motion blur due to camera shake [20] cannot deal with defocus blur, and motion blur due to moving objects. Generally, when we try to deblur an image, we do not know what type of blur is present in the image; thus, we cannot use parametric methods unless we identify the type in advance, which is not a trivial task. Furthermore, multiple types of blur often occur in a sin-

gle image. Although a few methods have been proposed to handle different blur types simultaneously by decomposing blur [28]–[30], their performance still depends on the flexibility of parameterization for each type and the accuracy of decomposition. In contrast, the proposed method is non-parametric and can handle general blur even when blur types are unknown. Note that the proposed field-based methodology does not fully conflict with the parametric approach, but can be integrated with it by considering a field of parametric kernels instead of nonparametric kernels. Such a hybrid approach, which seems to be effective when blur types are known but not globally parameterizable, is one of the possible future extensions of this work.

2.2 Other Topics

(1) Variational Bayes

Recent studies on BD have shown promising results for BD using VB techniques, which enable robust joint estimation of kernels and images. Unlike conventional techniques such as maximum a posteriori, VB can avoid trivial solutions in BD [1]. In fact, it has been reported that VB methods often outperform non-VB methods [1], [31]. Furthermore, VB algorithms can automatically adjust control parameters [31], which usually require image-dependent tuning in non-VB algorithms, via joint estimation under a single Bayesian model. In this paper, we also employ VB techniques to robustly estimate a field of kernels without difficult parameter tuning.

(2) Image and Kernel Priors

As estimation of blur-free images is inherently an ill-posed problem, successful BD depends on the choice of image priors. Various image priors that exploit properties of natural images have been proposed for BD [31]–[34]. As our focus is to enable flexible SV BD rather than seeking for better image priors, we use an image prior based on total variation (TV) that is known to be effective for VB-based BD [31]. Meanwhile, priors for kernels have been also proposed in SI BD [2], [35]. We discuss our choice of a prior for the proposed field of kernels during model construction in Sect. 3.

(3) Acceleration

Efficient implementation of BD algorithms has also been studied. However, acceleration of SV BD is more difficult than that of SI BD since frequency-domain techniques cannot be directly applied [6]. As we aim at effectiveness rather than efficiency, we leave accelerated implementation of the proposed SV BD algorithm for future work.

(4) Multiframe

Using multiple blurry images is known to be beneficial in BD [36]–[39]. In multiframe BD, an approach similar to ours was taken in [7], where a SV kernel is considered as a set of SI kernels for each pixel. However, assuming a common scenario in biomedical imaging and taking a non-VB approach, this method requires multiple aligned images

and parameters to be tuned, as opposed to the proposed VB method for general images. Recently, BD has been incorporated into superresolution (SR) [40]–[44], which primarily aims to recover high-resolution images from one or more low-resolution images. Although extending the proposed SV BD methodology to multiframe BD and SR is an interesting future direction, it is beyond the scope of this paper.

3. Model

We assume that we have a blurry observed image of n pixels, and denote it by a vector $\mathbf{y} = [y_1 \dots y_n]^T \in \mathbb{R}^n$. Our goal is to recover a sharp latent image, i.e., a blur-free version of \mathbf{y} , denoted by $\mathbf{x} = [x_1 \dots x_n]^T \in \mathbb{R}^n$. To achieve this, we assume that each observed pixel y_i is the convolution of the latent image \mathbf{x} and a *local* blur kernel with m coefficients $\mathbf{w}_i = [w_{i1} \dots w_{im}]^T \in \mathbb{R}^m$:

$$y_i = \sum_{j=1}^m w_{ij} x_{i \oplus j}, \quad (1)$$

where $i \oplus j$ denotes the j -th coefficient in the spatial support of the kernel at the i -th pixel. In practice, we assume that the support is square and centered at the i -th pixel. Then, we gather the n local kernels from all the pixels into a single vector $\mathbf{w} = [\mathbf{w}_1^T \dots \mathbf{w}_n^T]^T \in \mathbb{R}^{mn}$. We refer to this vector as a *field of kernels* since it has one kernel at each pixel in space. This model is a generalization of SI BD and segmentation-based BD, in which the same kernel is assumed for the pixels in the whole image and in each segment, respectively. For example, by omitting i for \mathbf{w}_i in (1), we could assume the same kernel at all pixels and obtain the traditional SI BD problem. Here, we do not make such a restrictive assumption; thus, our field-of-kernels model is more flexible, allowing an arbitrary kernel at each pixel.

In our BD problem, both the field of kernels the latent image are unknown. Thus, we need to estimate both of them to perform deconvolution. For this purpose, taking a Bayesian approach, we treat all observed and latent quantities as random variables and describe their relationships as probability distributions. In the following, we construct a Bayesian model by assigning a distribution to each variable. The graphical representation in Fig. 2 provides an overview of this model.

(1) Observed Image

First, we rephrase the SV convolution model in (1) in Bayesian terms. Here, we assume that dominant noise in observation is zero-mean additive Gaussian, which is a standard assumption in image restoration and BD [2], and define an elementwise Gaussian distribution on \mathbf{y} as follows:

$$p(\mathbf{y}|\mathbf{x}, \mathbf{w}, b) \propto \prod_{i=1}^n \sqrt{b} \exp\left(-\frac{b}{2}(\mathbf{y} - \mathbf{H}_w \mathbf{x})_i^2\right), \quad (2)$$

where $b \in \mathbb{R}$ is a fidelity parameter that determines the reliability of the observation, and $\mathbf{H}_w \in \mathbb{R}^{n \times n}$ is the SV convolution matrix with respect to \mathbf{w} such that $(\mathbf{H}_w \mathbf{x})_i$ equals the

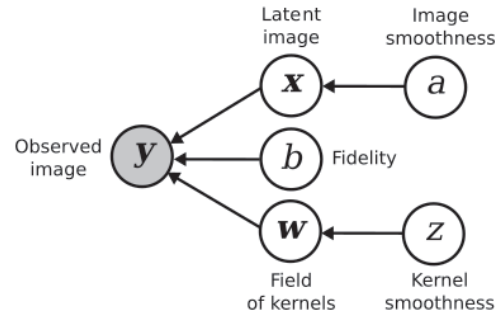


Fig. 2 Our Bayesian model for shift-variant blind deconvolution.

right-hand side of (1).

(2) Latent Image

Estimation of clean images from degraded images is a well-known ill-posed problem, i.e., we cannot identify \mathbf{x} from \mathbf{y} only; thus, we need to supplement observation with additional information. Exploiting a common property of natural images, we assume that the latent image is smooth except for edges, and minimize its variation while allowing for large variation at a small number of pixels. We achieve this by defining an elementwise zero-mean Laplacian distribution on the gradient magnitude of \mathbf{x} as follows:

$$p(\mathbf{x}|a) \propto \prod_{i=1}^n a \exp\left(-a \sqrt{\sum_{k=1}^2 (\mathbf{G}_k \mathbf{x})_i^2}\right), \quad (3)$$

where $a \in \mathbb{R}$ is a smoothness parameter that determines the global smoothness of \mathbf{x} , and $\mathbf{G}_1, \mathbf{G}_2 \in \mathbb{R}^{m \times m}$ are vertical and horizontal differentiation matrices for \mathbf{x} , respectively. This is basically a Bayesian version of the classic TV image prior [31], and its usage and effectiveness have been well-studied in the context of VB BD [31].

(3) Field of Kernels

In SI BD, it is known that kernels can be estimated even without priors [2]. This is because, given a reasonable estimate of a latent image, estimating a single kernel of m elements from an observed image of n elements is a well-posed problem when $m \ll n$. In our SV case, however, we have a kernel of m coefficients at each of n pixels and thus mn unknowns for a field of kernels, which makes the problem highly ill-posed. To overcome this ill-posedness, we also impose a prior on the field of kernels \mathbf{w} . While several kernel priors have been proposed for SI BD and we could use them for each local kernel, such priors lead to the same problem as explicitly parameterized methods, i.e., a limited range of tractable blur types [2]. Instead, we assume that the field of kernels is smooth, i.e., local kernels at adjacent pixels are similar. This assumption effectively regularizes the ill-posed SV BD problem but still allows for kernel variation between pixels, while SI and segmentation-based SV BD assume no variation at least locally. We realize such smoothing by defining an elementwise zero-mean Gaussian distribution on the gradient magnitudes of the field

as follows:

$$p(\mathbf{w}|z) \propto \prod_{i=1}^{mn} \sqrt{z} \exp\left(-\frac{z}{2} \sum_{k=1}^2 (\mathbf{F}_k \mathbf{w})_i^2\right), \quad (4)$$

where $z \in \mathbb{R}$ is a smoothness parameter for \mathbf{w} , and $\mathbf{F}_1, \mathbf{F}_2 \in \mathbb{R}^{mn \times mn}$ are horizontal and vertical differentiation matrices for \mathbf{w} , respectively, which take the differences of corresponding coefficients between adjacent pixels; thereby, the Gaussian reduces the variations in local kernels between pixels, effectively promoting smoothness of the field. While this smoothing prior is simple and not edge-preserving, it is sufficient to enable SV BD with the field of kernels, which is the primary purpose of this paper. Note that the field of kernels is not necessarily smooth in the same manner as natural images; for example, camera shake affects a whole image with smoothly varying blur, where the corresponding field has no distinct edges. In fact, edge-preserving priors such as Laplacians make a similar assumption to segmentation-based methods, i.e., that local kernels in each segments are constant, which can limit the range of tractable blur types as we discussed in Sect. 2. In addition, using such priors generally requires complicated approximations in inference or optimization, which have been well-studied only for natural image priors [32]. Since the main focus of this paper is on flexible BD via the field of kernels, we just use the simple yet effective Gaussian prior for the field of kernels, leaving the use of more elaborate priors as a possible future extension.

(4) Weights

For the weight-like parameters in the proposed model, i.e., b, a , and z , we do not have reasonable prior information; thus, we assume noninformative uniform distributions for these parameters, i.e., $p(b)$, $p(a)$, and $p(z)$ are constant. While such parameters require careful tuning for each image in non-VB BD [31], the proposed method can automatically adjust them by joint estimation via VB inference, as described in Sect. 4.

4. Algorithm

In Bayesian BD, our objective is to find the most probable latent image given an observed image, i.e., under the model defined in Sect. 3, we maximize the posterior probability of \mathbf{x} given \mathbf{y} to obtain a restored image $\hat{\mathbf{x}}$ as follows:

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}). \quad (5)$$

Using Bayes' theorem and the law of total probability, we obtain the posterior distribution of \mathbf{x} by marginalizing out latent variables other than \mathbf{x} from the joint distribution of all variables:

$$p(\mathbf{x}|\mathbf{y}) \propto \int p(\mathbf{y}, \mathbf{x}, \mathbf{w}, b, a, z) d\mathbf{w} db da dz, \quad (6)$$

where the joint distribution is the product of all the distributions in our model, i.e.,

$$p(\mathbf{y}, \mathbf{x}, \mathbf{w}, b, a, z) = p(\mathbf{y}|\mathbf{x}, \mathbf{w}, b) p(\mathbf{x}|a) p(\mathbf{w}|z) p(b) p(a) p(z). \quad (7)$$

Since exact marginalization is difficult due to the mutual dependencies between variables, we invoke a VB technique called *mean-field approximation* [45], [46]. Specifically, we approximate the exact joint posterior by the product of independent posteriors of individual variables as follows:

$$p(\mathbf{x}, \mathbf{w}, b, a, z|\mathbf{y}) \approx q(\mathbf{x}, \mathbf{w}, b, a, z) = q(\mathbf{x}) q(\mathbf{w}) q(b) q(a) q(z), \quad (8)$$

where exact and approximate posteriors are denoted by p and q , respectively. Here, $q(\mathbf{x})$ approximates $p(\mathbf{x}|\mathbf{y})$, which is required to evaluate (5). Following previous work on BD via VB [2], we also assume that $q(\mathbf{x})$ is Gaussian and $q(\mathbf{w})$, $q(b)$, $q(a)$, and $q(z)$ are degenerate. This assumption makes all approximate posteriors well-parameterized, thereby simplifying optimization [45], [47]. Let $\boldsymbol{\mu}_{\mathbf{x}} \in \mathbb{R}^n$ and $\boldsymbol{\Sigma}_{\mathbf{x}} \in \mathbb{R}^{n \times n}$ be the mean and covariance of $q(\mathbf{x})$, respectively, and let \hat{b} , \hat{a} , \hat{z} , and $\hat{\mathbf{w}}$ be the modes of $q(b)$, $q(a)$, $q(z)$, and $q(\mathbf{w})$, respectively. Then, we can obtain optimal approximate posteriors in terms of Kullback-Leibler divergence by maximizing the lower bound of the log-evidence of the model, which is denoted by f and defined as follows [45], [48]:

$$\begin{aligned} f[q(\mathbf{x}, \mathbf{w}, b, a, z)] &= \int q(\mathbf{x}, \mathbf{w}, b, a, z) \\ &\quad \ln \frac{p(\mathbf{y}, \mathbf{x}, \mathbf{w}, b, a, z)}{q(\mathbf{x}, \mathbf{w}, b, a, z)} d\mathbf{x} d\mathbf{w} db da dz. \end{aligned} \quad (9)$$

The non-Gaussian image prior defined in (3) prevents us from optimizing f directly; therefore, we employ another VB technique called local approximation [45], [49], which is common in VB-based BD [2]. Specifically, we approximate the non-Gaussian $p(\mathbf{x}|a)$ in (7) with a Gaussian-like $p'(\mathbf{x}|a, \mathbf{v})$ with an auxiliary parameter $\mathbf{v} \in \mathbb{R}^n$, which is derived from lower-bounding of $p(\mathbf{x}|a)$ [31] and defined as follows:

$$p'(\mathbf{x}|a, \mathbf{v}) \propto \prod_{i=1}^n \sqrt{a} \exp\left(-\frac{av_i}{2} \left(\sum_{k=1}^2 (\mathbf{G}_k \mathbf{x})_i^2 + v_i^{-2}\right)\right). \quad (10)$$

Intuitively, v_i is a local smoothing weight at the i -th pixel, which is adapted to the latent image automatically through VB inference, thereby enabling edge-preserving smoothing.

Combining (7) and (9) and replacing $p(\mathbf{x}|a)$ with $p'(\mathbf{x}|a, \mathbf{v})$, we obtain a modified lower bound, denoted by f' , which depends on both the approximate posteriors and the auxiliary \mathbf{v} . We can evaluate it using the parameterizations of approximate posteriors and the definition of the distributions in our model, i.e., (2), (4), and (10) as follows:

$$f'[q(\mathbf{x}), q(\mathbf{w}), q(b), q(a), q(z), \mathbf{v}]$$

$$\begin{aligned}
&= \frac{n}{2} \ln \hat{b} - \frac{\hat{b}}{2} (\|\mathbf{y} - \mathbf{H}_{\hat{\mathbf{w}}} \boldsymbol{\mu}_x\|_2^2 + \text{tr}(\mathbf{H}_{\hat{\mathbf{w}}} \boldsymbol{\Sigma}_x \mathbf{H}_{\hat{\mathbf{w}}}^T)) \\
&+ \frac{n}{2} \ln \hat{a} \\
&- \frac{\hat{a}}{2} \left(\sum_{k=1}^2 (\|V^{\frac{1}{2}} \mathbf{G}_k \boldsymbol{\mu}_x\|_2^2 + \text{tr}(\mathbf{V} \mathbf{G}_k \boldsymbol{\Sigma}_x \mathbf{G}_k^T)) + \text{tr}(\mathbf{V}^{-1}) \right) \\
&+ \frac{mn}{2} \ln \hat{z} - \frac{\hat{z}}{2} \sum_{k=1}^2 \|\mathbf{F}_k \hat{\mathbf{w}}\|_2^2 \\
&+ \frac{1}{2} \ln |\boldsymbol{\Sigma}_x| + \text{const.}, \tag{11}
\end{aligned}$$

where $\mathbf{H}_{\hat{\mathbf{w}}}$ is the version of $\mathbf{H}_{\mathbf{w}}$ constructed with $\hat{\mathbf{w}}$ and \mathbf{V} is the diagonal matrix of \mathbf{v} ; here, $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_x$ are the mean and covariance of $q(\mathbf{x})$, respectively, as defined above. Then, we maximize $\ln f'$ with respect to each parameter, as described in the following. The derivation of each equation is detailed in Appendix.

(1) Auxiliary Parameter

Taking the derivative of (11) with respect to v_i and setting it to zero, we obtain each element v_i of the auxiliary parameter \mathbf{v} as follows:

$$v_i = \frac{1}{\sqrt{\sum_{k=1}^2 ((\mathbf{G}_k \boldsymbol{\mu}_x)_i)^2 + (\mathbf{G}_k \boldsymbol{\Sigma}_x \mathbf{G}_k^T)_{ii}}}. \tag{12}$$

Here, v_i is basically the inverse of the local image variation at the i -th pixel captured by high-pass filtering via $\mathbf{G}_1, \mathbf{G}_2$; therefore, v_i becomes small around an edge, thereby weakening local smoothing to preserve the edge.

(2) Weights

Taking the derivative of (11) with respect to each of \hat{z} , \hat{a} , and \hat{b} and setting it to zero, we obtain the parameters of $q(\mathbf{z})$, $q(\mathbf{a})$, and $q(\mathbf{b})$, respectively, as follows:

$$\hat{z} = \frac{mn}{\sum_{i=1}^{mn} \sum_{k=1}^2 (\mathbf{F}_k \hat{\mathbf{w}})_i^2}, \tag{13}$$

$$\hat{a} = \frac{n}{\sum_{i=1}^n (v_i \sum_{k=1}^2 ((\mathbf{G}_k \boldsymbol{\mu}_x)_i)^2 + (\mathbf{G}_k \boldsymbol{\Sigma}_x \mathbf{G}_k^T)_{ii}) + v_i^{-1}}, \tag{14}$$

$$\hat{b} = \frac{n}{\sum_{i=1}^n ((\mathbf{y} - \mathbf{H}_{\hat{\mathbf{w}}} \boldsymbol{\mu}_x)_i)^2 + (\mathbf{H}_{\hat{\mathbf{w}}} \boldsymbol{\Sigma}_x \mathbf{H}_{\hat{\mathbf{w}}}^T)_{ii}}. \tag{15}$$

(3) Latent Image

Taking the derivative of (11) with respect to $\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x$ and setting it to zero, we obtain the parameters of $q(\mathbf{x})$ as follows:

$$\boldsymbol{\mu}_x = \left(\mathbf{H}_{\hat{\mathbf{w}}}^T \mathbf{H}_{\hat{\mathbf{w}}} + \frac{\hat{a}}{\hat{b}} \sum_{k=1}^2 \mathbf{G}_k^T \mathbf{V} \mathbf{G}_k \right)^{-1} \mathbf{H}_{\hat{\mathbf{w}}}^T \mathbf{y}, \tag{16}$$

$$\boldsymbol{\Sigma}_x = \left(\hat{b} \mathbf{H}_{\hat{\mathbf{w}}}^T \mathbf{H}_{\hat{\mathbf{w}}} + \hat{a} \sum_{k=1}^2 \mathbf{G}_k^T \mathbf{V} \mathbf{G}_k \right)^{-1}. \tag{17}$$

The covariance $\boldsymbol{\Sigma}_x$ represents uncertainty in estimates of

-
- 1: Input \mathbf{y} .
 - 2: Initialize $q(\mathbf{x})$ and $q(\mathbf{w})$.
 - 3: **repeat**
 - 4: Update \mathbf{v} with (12).
 - 5: Update $q(\mathbf{z})$, $q(\mathbf{a})$, $q(\mathbf{b})$ with (13), (14), (15).
 - 6: Update $q(\mathbf{x})$ with (16), (17).
 - 7: Update $q(\mathbf{w})$ with (18).
 - 8: **until** convergence.
 - 9: Output $\hat{\mathbf{x}} = \boldsymbol{\mu}_x$.
-

Fig. 3 The algorithm of VB inference for SV BD.

\mathbf{x} [45], [47], [50], working as regularizers in the denominators of (12), (14), (15). In BD, this property of VB inference helps avoid trivial solutions and improve deblurring [1], [31], [33]. By replacing the exact posterior $p(\mathbf{x}|\mathbf{y})$ with the obtained approximate posterior $q(\mathbf{x})$ in (5), we can obtain an optimal clean image estimate $\hat{\mathbf{x}}$ as the mode of $q(\mathbf{x})$, which coincides with $\boldsymbol{\mu}_x$ since $q(\mathbf{x})$ is Gaussian.

(4) Field of Kernels

Taking the derivative of (11) with respect to $\hat{\mathbf{w}}$ and setting it to zero, we obtain the parameter of $q(\mathbf{w})$ as follows:

$$\hat{\mathbf{w}} = \left(\mathbf{H}_{\boldsymbol{\mu}_x}^T \mathbf{H}_{\boldsymbol{\mu}_x} + \mathbf{H}_{\boldsymbol{\Sigma}_x}^2 + \frac{\hat{z}}{\hat{b}} \sum_{k=1}^2 \mathbf{F}_k^T \mathbf{F}_k \right)^{-1} \mathbf{H}_{\boldsymbol{\mu}_x}^T \mathbf{y}, \tag{18}$$

and $\mathbf{H}_{\boldsymbol{\mu}_x}$ and $\mathbf{H}_{\boldsymbol{\Sigma}_x}^2$ are the matrices such that $\mathbf{H}_{\boldsymbol{\mu}_x} \hat{\mathbf{w}} = \mathbf{H}_{\hat{\mathbf{w}}} \boldsymbol{\mu}_x$ and $\hat{\mathbf{w}}^T \mathbf{H}_{\boldsymbol{\Sigma}_x}^2 \hat{\mathbf{w}} = \text{tr}(\mathbf{H}_{\hat{\mathbf{w}}} \boldsymbol{\Sigma}_x \mathbf{H}_{\hat{\mathbf{w}}}^T)$, respectively; here, to reflect the physical properties of real kernels [2], we also constrain each local kernel $\hat{\mathbf{w}}_i$ so that $0 \leq \hat{w}_{ij} \leq 1$ and $\sum_{j=1}^m \hat{w}_{ij} = 1$.

Since each approximate posterior depends on others through parameters, we iteratively update them one by one, starting with some initial estimates, which we discuss in more detail in Sect. 5. Such a VB algorithm is guaranteed to converge [45], [51]. The resulting algorithm is summarized in Fig. 3.

5. Implementation

The algorithm in Fig. 3 can be numerically implemented using linear algebra routines on dense vectors and sparse matrices. While we sometimes need to invert large sparse matrices, exact inversion is computationally intractable. Thus, we use conjugate gradient in (18) and (16) and diagonal covariance approximation in (12), (14), and (15), as in previous VB-based methods [2], [31].

At each iteration, we normalize the estimate of each local kernel in the field to enforce the constraint described at (18). Besides, sometimes very large values occur in estimates of other parameters, particularly at the first iteration. For example, according to (13), \hat{z} approaches to infinity if $\hat{\mathbf{w}}$ is initialized by a smooth kernel field with no variation. To ensure numerical stability in such cases, we perform elementwise value clamping on each estimate after updating it so that the absolute value of each element is no greater than 10^4 .

In SI BD, estimation in filter domains tends to yield better kernel estimates [2]. To improve kernel estimation, we also apply this technique to our SV BD. Specifically, in our model we replace \mathbf{y} and \mathbf{x} in (2) with prefiltered versions $\mathbf{G}\mathbf{y}$ and $\mathbf{G}\mathbf{x}$, respectively, where $\mathbf{G} = [\mathbf{G}_1^T \mathbf{G}_2^T]^T$. Then, we perform lower bound maximization with $q(\mathbf{G}\mathbf{x})$ instead of $q(\mathbf{x})$. If we consider \mathbf{y} and \mathbf{x} in the original algorithm as $\mathbf{G}\mathbf{y}$ and $\mathbf{G}\mathbf{x}$, the resulting algorithm is nearly the same as the original one, except that \mathbf{G}_1 and \mathbf{G}_2 disappear from (12), (14) (16), and (17) since we have prefiltered images, and the matrices multiplied with \mathbf{y} and \mathbf{x} become two-channel operators, i.e., separately operate on horizontal and vertical components of $\mathbf{G}\mathbf{y}$ and $\mathbf{G}\mathbf{x}$. This modified algorithm only yields the latent image estimate in the filter domain; thus, we perform a final non-blind deconvolution step to obtain $\hat{\mathbf{x}}$ in the image domain using $\hat{\mathbf{w}}$ estimated in the filter domain. To achieve this, we run the original image-domain algorithm with $\hat{\mathbf{w}}$ fixed to the estimate obtained in the filter domain.

To initialize $\hat{\mathbf{w}}$ in the filter domain, we set each local kernel to the delta kernel, i.e., the no-blur solution. Regarding $q(\mathbf{x})$, we initialize $\mu_{\mathbf{x}}$ using observations, i.e., with $\mathbf{G}\mathbf{y}$ and \mathbf{y} in the filter and image domains, respectively. Meanwhile, in both domains we initialize the diagonal of $\Sigma_{\mathbf{x}}$ with constant 10^4 , which means large initial uncertainty and drives optimization away from initial no-deblurring state. For each run of the algorithm, we stop when variation in $\mathbf{G}\hat{\mathbf{x}}$ or $\hat{\mathbf{x}}$ between iterations is less than 0.1 in mean squared error (MSE) or after 16 iterations, which are empirically sufficient for convergence.

6. Experiments

We conducted experiments to evaluate the effectiveness of the proposed method. To enable quantitative evaluation in reference to ground truth, we prepared observed images as inputs by convolving standard test images as latent images with synthesized SV kernels, which imitate typical types of nonuniform blur in reality. For each test image, a SV kernel was synthesized as a field of kernels, where a local kernel is assigned to each pixel in the manually-extracted foreground and background, thereby allowing different blur types between them. In the following, the sizes of each image and local kernel were $n = 256 \times 256$ pixels and $m = 5 \times 5$ coefficients per pixel, respectively.

For a fair comparison, we modified the proposed SV BD method into a SI version with nearly the same model, assuming a single local kernel common to all pixels in the convolution model (1) and disabling kernel smoothing. Note that, ignoring minor differences in optimization, this SI variant is in principle equivalent to a previous method of VB-based BD [31], which in turn is a special case of another method for general image priors [32] that was reported to achieve state-of-the-art performance for SI BD [33]. We also compared the proposed field-of-kernels-based method with a segmentation-based variant that assumes the same kernel in each segment in (1). Here, we reused the ground-truth segmentation result of foreground and background

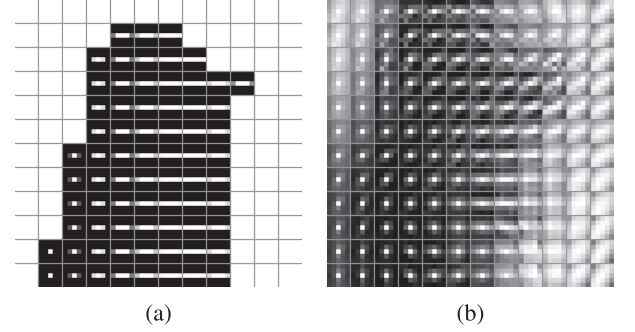


Fig. 4 Fields of kernels for *Lena*. From the 256×256 pixels of each field, 12×12 pixels were sampled at an equal interval. The value range of the local kernel at each pixel were maximized for visualization. (a) Ground truth. (b) Estimated by the proposed method. Note that the white blocks in (a) indicate uniform kernels.

used in blurring; thus, this is basically a SV BD method based on perfect segmentation with the state-of-the-art image prior [32]. These three BD methods were applied to each observed image. The image quality of each restored image was assessed in terms of peak-signal-to-noise ratio (PSNR) with respect to the corresponding latent image.

First, we blurred the *Lena* image with horizontal line kernels of variable lengths for the foreground and a box kernel for the background, which mimic motion blur due to rotation around the vertical axis and uniform defocus blur, respectively. Several samples of local kernels from this SV kernel are shown in Fig. 4 (a), where we can see that the rotation caused smooth variation of kernels in the foreground. The results are shown in Fig. 5 with PSNR values. Here, the SI method could not fully remove blur and even degraded image quality in terms of PSNR, attempting deconvolution based on SI kernel estimation against nonuniform blur. Meanwhile, the segmentation-based SV BD could make only a slight PSNR improvement despite the use of the perfect segmentation result, since the smoothly varying blur in the foreground is difficult to approximate by segmentation. Note that such composite blur is out of the domain of most parametric BD methods that globally assume a single type of blur, e.g., motion or defocus. In contrast, the proposed field-based SV method successfully recovered a visually sharper image (e.g., around the edge of the face) with a higher PSNR, dealing with both the smooth variation and the complexity of the blur. Furthermore, the field of kernels estimated by the proposed method is shown in Fig. 4 (b). Here, we can see that the proposed method could capture the outline of the field, i.e., the difference between the foreground and the background, and also the smooth variation in the foreground along the horizontal axis. Since we used the non-edge-preserving Gaussian prior for smoothing of the field, the discontinuities around the object boundaries were smoothed out. Still, such a kernel prior was sufficient for achieving PSNR improvement in the image domain, which is the ultimate objective of BD. Using edge-preserving priors for the field of kernels can incur additional complexities and limitations as we discussed in Sect. 3; thus, we leave

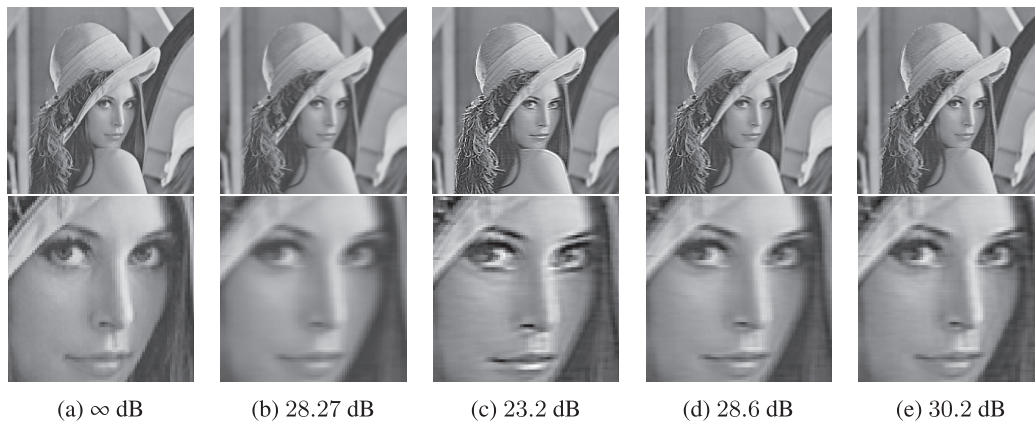


Fig. 5 Images of Lena (top) and their closeups (bottom) with PSNR values: (a) Latent. (b) Observed. (c) Restored by SI BD. (d) Restored by segmentation-based SV BD. (e) Restored by field-of-kernels-based SV BD (proposed).

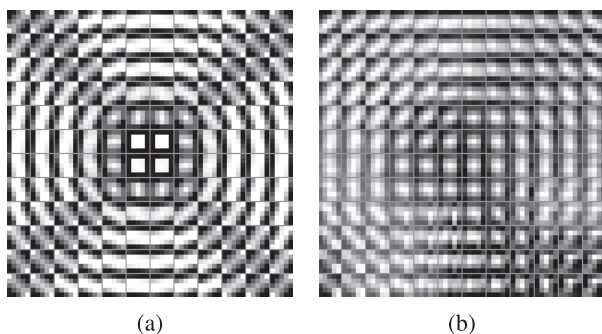


Fig. 6 Fields of kernels for *Cameraman*. From the 256×256 pixels of each field, 12×12 pixels were sampled at an equal interval. The value range of the local kernel at each pixel were maximized for visualization. (a) Ground truth. (b) Estimated by the proposed method.

this topic for future work.

Next, we applied another realistic instance of blur to the *Cameraman* image, where global in-plane rotation (which often occurs in camera shake [11]) is mixed with a out-of-focus effect; thus, both motion and defocus blur simultaneously affected each pixel, which resulted in very complex local kernels as shown in Fig. 6(a). Moreover, the smooth variation of rotational motion blur cannot be described by a finite number of kernels assumed by segmentation-based methods, while parametric methods specialized for motion blur cannot deal with the mixed defocus blur. As can be seen from the results shown in Fig. 7, the proposed SV method outperformed the SI and segmentation-based SV methods also for this image, both in appearance and in PSNR. From the estimated field of kernel shown in Fig. 6(b), we can see that the proposed method could stably capture smooth variations in the field, except for the bottom-right region of the image, where the estimated local kernels were relatively close to initial delta kernels. This could be because this background region had only textures of rather random patterns but no strong edges; therefore, there were no distinct image features available for kernel estimation. Note that BD in such regions is in

principle highly ill-posed, and to address this issue, we might need to resort to more radical approaches such as multi-observation [37]. It is noteworthy that the proposed method achieved these results without image-dependent tuning of parameters, e.g., numbers of segments, nor prior knowledge on blur types. In practical applications, this is a huge advantage over previous approaches.

Table 1 summarizes the computational time for the *Lena* and *Cameraman* images with our current implementation on an *Intel Xeon E5-2660* CPU. In these experiments, the execution time of the SI method was longer than the segmentation-based SV method, possibly because the violation of the invariance assumption slowed down its convergence. Basically, the proposed field-of-kernels-based SV method took longer than the SI and segmentation-based SV methods due to the increased number of kernel coefficients to be estimated.

Figure 8 shows the convergence of the proposed algorithm in terms of the MSE in filter-domain image estimates between iterations when we varied the size of local kernels to be estimated. Here, the PSNR values of the restored images became the highest when the local kernel size is 5×5 , i.e., when it coincides with the size of the ground truth. From the graphs, we can observe that the variation in the estimate became stationary, or small enough to terminate (below 0.1), before 16 iterations regardless of kernels sizes.

Figure 9 shows the noise tolerance of the proposed method. Here, the proposed method was applied to versions of the blurry *Lena* and *Cameraman* images degraded with Gaussian noise of varying standard deviation, and its performance was measured in terms of improvement in signal-to-noise ratio (ISNR), i.e., the difference in PSNR between input and output. While the proposed method successfully improved image quality with modest noise, it began to fail at higher noise levels, i.e., ISNR values went below zero, e.g., around standard deviation 3 and 5 in the cases of *Lena* and *Cameraman*, respectively. We note that BD in general is sensitive to degradation of high-frequency image components due to noise, especially in the filter domain [2]. To



Fig. 7 Images of Cameraman (top) and their closeups (bottom) with PSNR values: (a) Latent. (b) Observed. (c) Restored by SI BD. (d) Restored by segmentation-based SV BD. (e) Restored by field-of-kernels-based SV BD (proposed).

Table 1 Computational time of the proposed method [seconds]

Image	SI BD	segmentation-based SV BD	field-of-kernels-based SV BD (proposed)
<i>Lena</i>	43	35	338
<i>Cameraman</i>	33	30	251

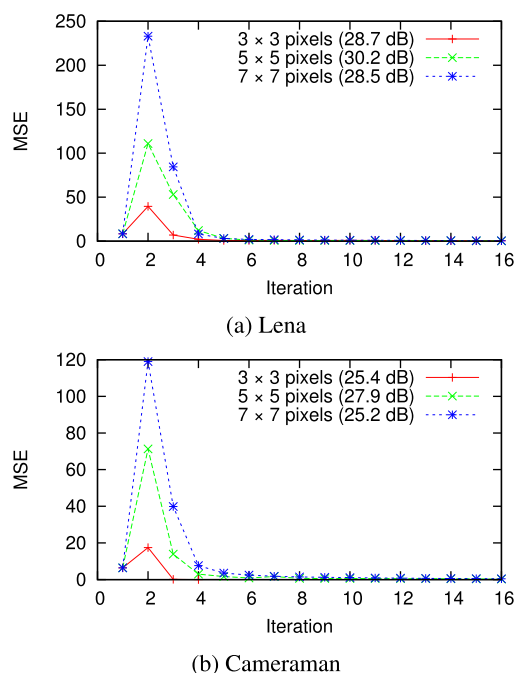


Fig. 8 Convergence of the proposed method for different local kernel sizes. The PSNR values of the corresponding result images are shown in parentheses.

improve the noise tolerance of the proposed method, we will need to combine our model with image priors that have stronger power to suppress noise, e.g., as proposed in [32].

Finally, using more practical images with large blur, we compared the proposed method with recent BD methods: the SI method by Fergus et al. [52], the SV method for rotational camera shake by Whyte et al. [20], the SV method

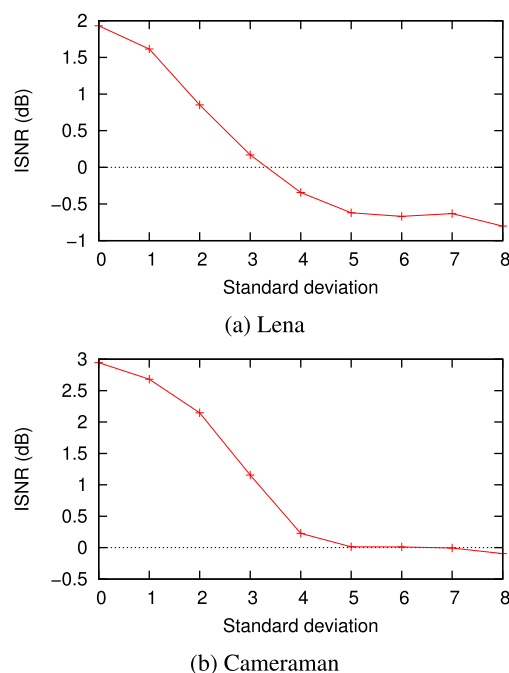


Fig. 9 Noise tolerance of the proposed method for varying noise levels measured in standard deviation.

parameterized with a motion path by Gupta et al. [53], and the SV method by Hirsch et al. [13], which is an extension of the method by Gupta et al. [53] with patch-wise parameterization, and thus can be regarded as a parametric approach to segmentation-based BD. To compare the proposed method with these methods, we applied the proposed method to images blurred by camera shake used in Whyte et al. [20] and Gupta et al. [53]. Here, the sizes of the input images were 354×265 and 768×512 pixels, respectively, and the size of each local kernel was set to 11×11 pixels. Note that these real images do not have corresponding ground truth images and kernels; thus, only visual comparison is possible here. The results are shown in Fig. 10 and Fig. 11, respectively. While the SI method of Fergus et al. [52] failed in handling

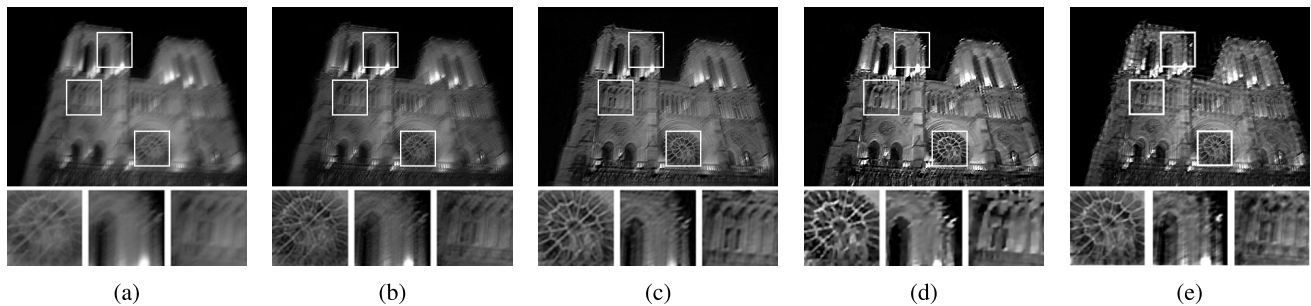


Fig. 10 Comparison with the previous methods using a practical image with large blur from Whyte et al. [20]. The input image to the proposed method was released by the authors of [20], while the result images by the previous methods were taken from [13]. These images were converted to grayscale and gamma-corrected for visual comparison, since the colors of the images from [13] were not matched to that of the input image. (a) Observed. (b) Restored by the SI method of Fergus et al. [52]. (c) Restored by the parametric SV method of Whyte et al. [20]. (d) Restored by the parametric SV method of Hirsch et al. [13]. (e) Restored by the proposed nonparametric SV method.



Fig. 11 Comparison with the previous methods using a practical image with large blur from Gupta et al. [53]. The input image to the proposed method was released by the authors of [53], while the result images by the previous methods were taken from [13]. (a) Observed. (b) Restored by the parametric SV method of Gupta et al. [53]. (c) Restored by the parametric SV method of Hirsch et al. [13]. (d) Restored by the proposed nonparametric SV method.

nonuniform blur, the proposed method successfully reduced blur to some extent. The deblurring performance of the proposed method was visually comparable with the parametric SV methods of Whyte et al. and Gupta et al. [20], [53], although the proposed method produced some artifacts in highly blurry regions, e.g., in the middle closeup image of Fig. 10(e). This is possibly due to inaccuracy in kernel estimation, which could be seen as the price for the flexibility offered by the nonparametric approach; still, such artifacts might be suppressed by using more effective image priors [32], which we consider as one of promising future extensions of the proposed method. Meanwhile, the previous parametric methods, especially the one of Hirsch et al. [13], yielded sharper results in several regions, e.g., in the middle closeup image of Fig. 11(c); however, they also exhibited oversmoothing in regions with textures, as seen in the middle and right closeup images of Fig. 10(d), which could happen when the assumptions under their parameterizations did not hold exactly. Note that the proposed method has a clear advantage over these previous methods; i.e., while these previous methods are specialized to camera motion blur, the proposed method has the potential to handle defocus blur, camera motion blur, object motion blur, and even mixture of them, without the need for blur type

identification for each image.

7. Conclusion

In this paper, we have proposed a SV BD method based on a field-of-kernels model. By modeling a spatially-varying blur kernel as a smooth field of local kernels, the proposed VB-based method can flexibly handle different kernels between pixels. Our experimental results confirm that the proposed method can successfully deblur images affected by nonuniform blur, which is intractable for previous approaches such as SI BD and segmentation-based SV BD.

One possible future direction is to introduce more effective priors to the field of kernels. Incorporating elaborate image priors and using multiple observed images will also enhance the effectiveness of the proposed method. Accelerated implementation of the proposed algorithm is another topic of interest. Finally, integration with parametric approaches and other applications such as SR are interesting future extensions of the proposed SV BD methodology.

References

- [1] A. Levin, Y. Weiss, F. Durand, and W.T. Freeman, "Understand-

- ing and evaluating blind deconvolution algorithms," *Proc. CVPR*, pp.1964–1971, 2009.
- [2] P. Ruiz, X. Zhou, J. Mateos, R. Molina, and A.K. Katsaggelos, "Variational bayesian blind image deconvolution: a review," *Digital Signal Processing*, May 2015.
 - [3] M. Tico and K. Pulli, "Low-light imaging solutions for mobile devices," 2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers, pp.851–855, 2009.
 - [4] R. Raskar, A. Agrawal, and J. Tumblin, "Coded exposure photography: motion deblurring using fluttered shutter," vol.25, no.3, pp.795–804, 2006.
 - [5] R. Szeliski, *Computer vision: algorithms and applications*, Springer, London, UK, 2011.
 - [6] L. Denis, E. Thiébaud, and F. Soulez, "Fast model of space-variant blurring and its application to deconvolution in astronomy," *Proc. ICIP*, pp.2817–2820, 2011.
 - [7] M. Blume, D. Zikic, W. Wein, and N. Navab, "A new and general method for blind shift-variant deconvolution of biomedical images," *Proc. MICCAI*, pp.743–750, Springer, 2007.
 - [8] T. Hyun Kim, B. Ahn, and K. Mu Lee, "Dynamic scene deblurring," *Proc. ICCV*, pp.3160–3167, 2013.
 - [9] J.G. Nagy, V.P. Pauca, R.J. Plemmons, and T.C. Torgersen, "Space-varying restoration of optical images," *JOSA A*, vol.14, no.12, pp.3162–3174, 1997.
 - [10] J.G. Nagy and D.P. O'Leary, "Restoring images degraded by spatially variant blur," *SISC*, vol.19, no.4, pp.1063–1082, 1998.
 - [11] M.S.C. Almeida and L.B. Almeida, "Blind deblurring of foreground-background images," *Proc. ICIP*, pp.1301–1304, 2009.
 - [12] A. Chakrabarti, T. Zickler, and W.T. Freeman, "Analyzing spatially-varying blur," *Proc. CVPR*, pp.2512–2519, IEEE, 2010.
 - [13] M. Hirsch, C.J. Schuler, S. Harmeling, and B. Schölkopf, "Fast removal of non-uniform camera shake," *Proc. ICCV*, pp.463–470, 2011.
 - [14] J. Pan, Z. Hu, Z. Su, H.Y. Lee, and M.H. Yang, "Soft-segmentation guided object motion deblurring," *Proc. CVPR*, pp.459–468, 2016.
 - [15] X. Zhang, R. Wang, X. Jiang, W. Wang, and W. Gao, "Spatially variant defocus blur map estimation and deblurring from a single image," *J. Vis. Commun. Image R.*, vol.35, pp.257–264, Feb. 2016.
 - [16] D. Miraut, J. Ballé, and J. Portilla, "Efficient shift-variant image restoration using deformable filtering (part ii): Psf field estimation," *EURASIP*, vol.2012, no.1, p.193, 2012.
 - [17] L. Denis, E. Thiébaud, F. Soulez, J.-M. Becker, and R. Mourya, "Fast approximations of shift-variant blur," *International Journal of Computer Vision*, vol.115, no.3, pp.253–278, 2015.
 - [18] J.K. Sroubek and Y.M. Lu, "Decomposition of space-variant blur in image deconvolution," *IEEE Signal Process. Lett.*, vol.23, no.3, pp.346–350, 2009.
 - [19] M. Hirsch, S. Sra, B. Schölkopf, and S. Harmeling, "Efficient filter flow for space-variant multiframe blind deconvolution," vol.1, p.2, 2010.
 - [20] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce, "Non-uniform deblurring for shaken images," *Proc. IEEE CVPR*, pp.233–240, 2010.
 - [21] H. Zhang and D. Wipf, "Non-uniform camera shake removal using a spatially-adaptive sparse penalty," *Proc. NIPS*, pp.1556–1564, 2013.
 - [22] T.H. Kim and K.M. Lee, "Segmentation-free dynamic scene deblurring," *Proc. CVPR*, pp.2766–2773, June 2014.
 - [23] Y. Xu, X. Hu, and S. Peng, "Blind motion deblurring using optical flow," *Optik*, vol.126, no.1, pp.87–94, Jan. 2015.
 - [24] E. Kee, S. Paris, S. Chen, and J. Wang, "Modeling and removing spatially-varying optical blur," *Proc. ICCP*, pp.1–8, 2011.
 - [25] S.H. Chan and T.Q. Nguyen, "Single image spatially variant out-of-focus blur removal," *Proc. ICIP*, pp.677–680, IEEE, 2011.
 - [26] H. Cheong, E. Chae, E. Lee, G. Jo, and J. Paik, "Fast image restoration for spatially varying defocus blur of imaging sensor," *Sensors*, vol.15, no.1, pp.880–898, 2015.
 - [27] X. Zhu, S. Cohen, S. Schiller, and P. Milanfar, "Estimating spatially varying defocus blur from a single image," *IEEE Trans. Image Process.*, vol.22, no.12, pp.4879–4891, Dec. 2014.
 - [28] F. Couzinie-Devy, J. Sun, K. Alahari, and J. Ponce, "Learning to estimate and remove non-uniform image blur," *Proc. CVPR*, pp.1075–1082, 2013.
 - [29] S. Tiwari, V.P. Shukla, S.R. Biradar, and A.K. Singh, "Blur parameters identification for simultaneous defocus and motion blur," *CSI Trans. ICT*, vol.2, no.1, pp.11–22, 2014.
 - [30] L. Fang, H. Liu, F. Wu, X. Sun, and H. Li, "Separable kernel for image deblurring," *Proc. CVPR*, pp.2885–2892, 2014.
 - [31] S.D. Babacan, R. Molina, and A.K. Katsaggelos, "Variational bayesian blind deconvolution using a total variation prior," *IEEE Trans. Image Process.*, vol.18, no.1, pp.12–26, Jan. 2009.
 - [32] S.D. Babacan, R. Molina, M.N. Do, and A.K. Katsaggelos, "Bayesian blind deconvolution with general sparse image priors," *Proc. ECCV*, pp.341–355, 2012.
 - [33] D. Wipf and H. Zhang, "Revisiting bayesian blind deconvolution," *J. Mach. Learn. Res.*, vol.15, no.8, pp.3595–3634, Jan. 2014.
 - [34] J. Pan, D. Sun, H. Pfister, and M.H. Yang, "Blind image deblurring using dark channel prior," *Proc. CVPR*, pp.1628–1636, 2016.
 - [35] X. Zhou, J. Mateos, F. Zhou, R. Molina, and A.K. Katsaggelos, "Variational dirichlet blur kernel estimation," *IEEE Trans. Image Process.*, vol.24, no.12, pp.5127–5139, 2015.
 - [36] L. Bar, B. Berkels, M. Rumpf, and G. Sapiro, "A variational framework for simultaneous motion estimation and restoration of motion-blurred video," *Proc. ICCV*, pp.1–8, 2007.
 - [37] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Blind motion deblurring using multiple images," *Journal of computational physics*, vol.228, no.14, pp.5057–5071, 2009.
 - [38] Y.-H. Fan, X.-M. Wei, and S.-Y. Qin, "Fast and robust deblurring method with multi-frame images based on psf estimation and total variation optimization," *Optik*, vol.124, no.16, pp.2285–2291, 2013.
 - [39] H. Zhang, D. Wipf, and Y. Zhang, "Multi-observation blind deconvolution with an adaptive sparse prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.36, no.8, pp.1628–1643, 2014.
 - [40] S.D. Babacan, R. Molina, and A.K. Katsaggelos, "Variational bayesian super resolution," *IEEE Trans. Image Process.*, vol.20, no.4, pp.984–999, April 2011.
 - [41] E. Faramarzi, D. Rajan, and M.P. Christensen, "Unified blind method for multi-image super-resolution and single/multi-image blur deconvolution," *IEEE Trans. Image Process.*, vol.22, no.6, pp.2101–2114, 2013.
 - [42] C. Liu and D. Sun, "On bayesian adaptive video super resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.36, no.2, pp.346–360, June 2014.
 - [43] H. Zhang and L. Carin, "Multi-shot imaging: joint alignment, deblurring, and resolution-enhancement," *Proc. CVPR*, pp.2925–2932, 2014.
 - [44] Z. Ma, R. Liao, X. Tao, L. Xu, J. Jia, and E. Wu, "Handling motion blur in multi-frame super-resolution," *Proc. CVPR*, pp.5224–5232, 2015.
 - [45] C.M. Bishop, *Pattern recognition and machine learning*, Springer, New York, NY, USA, 2006.
 - [46] G. Parisi, *Statistical field theory*, Addison-Wesley, 1988.
 - [47] H. Attias, "Inferring parameters and structure of latent variable models by variational bayes," *Uncertainty in Artificial Intelligence: Proceedings of the Fifteenth Conference*, pp.21–30, Morgan Kaufmann, 1999.
 - [48] T. Jaakkola, "Tutorial on variational approximation methods," *Advanced mean field methods: theory and practice*, p.129, 2001.
 - [49] R. Rockafellar, *Convex analysis*, Princeton University Press, 1972.
 - [50] T. Minka, "Inferring a gaussian distribution," *Media Lab Note*, 1998.
 - [51] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.
 - [52] R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, and W.T. Freeman, "Removing camera shake from a single photograph," vol.25, no.3, pp.787–794, 2006.
 - [53] A. Gupta, N. Joshi, C. Lawrence Zitnick, M. Cohen, and B. Curless,

“Single image deblurring using motion density functions,” *Computer Vision–ECCV 2010*, pp.171–184, 2010.

[54] K.B. Petersen and M.S. Pedersen, “The matrix cookbook,” 2012. Version 20121115.

Appendix A: Derivation of the VB inference Algorithm

In this section, we derive the algorithm of the VB inference for SV BD, which is outlined in Sect. 4. Under our Bayesian model, the update equation for each latent variable is obtained by maximizing the modified variational lower bound f' with respect to the corresponding approximate posterior.

To begin with, we derive the expression of f' in (11). By definition, f' is obtained by substituting (7) and (8) into (9) and replacing $p(x|a)$ with $p'(x|a, v)$, i.e.,

$$\begin{aligned} f'[q(x), q(w), q(b), q(a), q(z), v] \\ = \mathbb{E} \left[\ln \frac{p(y|x, w, b)p'(x|a, v)p(w|z)p(b)p(a)p(z)}{q(x)q(w)q(b)q(a)q(z)} \right], \end{aligned} \quad (\text{A} \cdot 1)$$

where \mathbb{E} denotes expectation with respect to all approximate posteriors. Using properties of logarithms, we can decompose the right-hand side of (A·1) as follows:

$$\begin{aligned} f'[q(x), q(w), q(b), q(a), q(z), v] \\ = \mathbb{E}[\ln p(y|x, w, b)] + \mathbb{E}[\ln p'(x|a, v)] + \mathbb{E}[\ln p(w|z)] \\ + \mathbb{E}[\ln p(a)] + \mathbb{E}[\ln p(b)] + \mathbb{E}[\ln p(z)] \\ + H[x] + H[w] + H[b] + H[a] + H[z], \end{aligned} \quad (\text{A} \cdot 2)$$

where H denotes entropy, i.e., $H[\alpha] = \mathbb{E}[-\ln q(\alpha)]$ for any α . Here, we can take expectations independently for each variable owing to mean-field approximation in (8). Since $q(w)$, $q(b)$, $q(a)$, $q(z)$ are assumed to be degenerate, the values of their variables always coincide with their modes, and their entropies are zero. Moreover, the hyperpriors $p(b)$, $p(a)$, $p(z)$ in our model are uniform and thus constant. Therefore, by substituting the modes \hat{w} , \hat{b} , \hat{a} , \hat{z} for w , b , a , z , respectively, setting $H[w]$, $H[b]$, $H[a]$, $H[z]$ to zero, and gathering the constant terms $\mathbb{E}[\ln p(b)]$, $\mathbb{E}[\ln p(a)]$, $\mathbb{E}[\ln p(z)]$, we can further simplify (A·2) as follows:

$$\begin{aligned} f'[q(x), q(w), q(b), q(a), q(z), v] \\ = \mathbb{E}[\ln p(y|x, \hat{w}, \hat{b})] + \mathbb{E}[\ln p'(x|\hat{a}, v)] \\ + \ln p(\hat{w}|\hat{z}) + H[x] + \text{const.}, \end{aligned} \quad (\text{A} \cdot 3)$$

where only expectations with respect to $q(x)$ remain. To evaluate them, we will use the following identity that holds for any $\alpha, \beta \in \mathbb{R}^v$ that follows from basis properties of random vectors [54]:

$$\begin{aligned} \mathbb{E} \left[\sum_{i=1}^v \beta_i \alpha_i^2 \right] &= \mathbb{E} [\alpha^T \mathbf{B} \alpha] \\ &= \mathbb{E} [\alpha]^T \mathbf{B} \mathbb{E} [\alpha] + \text{tr}(\mathbf{B} \text{Var}[\alpha]) \\ &= \|\mathbf{B}^{\frac{1}{2}} \mathbb{E} [\alpha]\|_2^2 + \text{tr}(\mathbf{B} \text{Var}[\alpha]), \end{aligned} \quad (\text{A} \cdot 4)$$

where $\text{Var}[\alpha]$ is the variance of α , and \mathbf{B} is the diagonal

matrix of β . First, taking the logarithmic expectations of both sides of (2) with $w = \hat{w}$ and $b = \hat{b}$, we obtain

$$\begin{aligned} \mathbb{E}[\ln p(y|x, \hat{w}, \hat{b})] \\ = \frac{n}{2} \ln \hat{b} - \frac{\hat{b}}{2} \mathbb{E} \left[\sum_{i=1}^n (y - \mathbf{H}_{\hat{w}} \mathbf{x})_i^2 \right] + \text{const.} \end{aligned} \quad (\text{A} \cdot 5)$$

Then, substituting (A·4) with $\alpha = y - \mathbf{H}_{\hat{w}} \mathbf{x}$, $\beta = \mathbf{1}$, we obtain

$$\begin{aligned} \mathbb{E}[\ln p(y|x, \hat{w}, \hat{b})] \\ = \frac{n}{2} \ln \hat{b} - \frac{\hat{b}}{2} (\|y - \mathbf{H}_{\hat{w}} \mu_x\|_2^2 + \text{tr}(\mathbf{H}_{\hat{w}} \Sigma_x \mathbf{H}_{\hat{w}}^T)) + \text{const.}, \end{aligned} \quad (\text{A} \cdot 6)$$

where we have evaluated the expectation and variance as follows:

$$\mathbb{E}[y - \mathbf{H}_{\hat{w}} \mathbf{x}] = y - \mathbf{H}_{\hat{w}} \mu_x, \quad (\text{A} \cdot 7)$$

$$\text{Var}[y - \mathbf{H}_{\hat{w}} \mathbf{x}] = \mathbf{H}_{\hat{w}} \Sigma_x \mathbf{H}_{\hat{w}}^T. \quad (\text{A} \cdot 8)$$

Second, taking the logarithmic expectations of both sides of (10) with $a = \hat{a}$, we obtain

$$\begin{aligned} \mathbb{E}[\ln p'(x|\hat{a}, v)] \\ = \frac{n}{2} \ln \hat{a} - \frac{\hat{a}}{2} \left(\sum_{k=1}^2 \mathbb{E} \left[\sum_{i=1}^n v_i (\mathbf{G}_k \mathbf{x})_i \right] + \sum_{i=1}^n v_i^{-1} \right) + \text{const.} \end{aligned} \quad (\text{A} \cdot 9)$$

Then, substituting (A·4) with $\alpha = \mathbf{G}_k \mathbf{x}$ and $\beta = v$, we obtain

$$\begin{aligned} \mathbb{E}[\ln p'(x|\hat{a}, v)] \\ = \frac{n}{2} \ln \hat{a} \\ - \frac{\hat{a}}{2} \left(\sum_{k=1}^2 (\|V^{\frac{1}{2}} \mathbf{G}_k \mu_x\|_2^2 + \text{tr}(\mathbf{V} \mathbf{G}_k \Sigma_x \mathbf{G}_k^T)) + \text{tr}(\mathbf{V}^{-1}) \right) \\ + \text{const.}, \end{aligned} \quad (\text{A} \cdot 10)$$

where we have used $\sum_{i=1}^n v_i^{-1} = \text{tr}(\mathbf{V}^{-1})$ and evaluated the expectation and variance as follows:

$$\mathbb{E}[\mathbf{G}_k \mathbf{x}] = \mathbf{G}_k \mu_x, \quad (\text{A} \cdot 11)$$

$$\text{Var}[\mathbf{G}_k \mathbf{x}] = \mathbf{G}_k \Sigma_x \mathbf{G}_k^T. \quad (\text{A} \cdot 12)$$

Third, taking the logarithms of both sides of (4) with $w = \hat{w}$ and $z = \hat{z}$, we obtain

$$\ln p(\hat{w}|\hat{z}) = \frac{mn}{2} \ln \hat{z} - \frac{\hat{z}}{2} \sum_{k=1}^2 \|\mathbf{F}_k \hat{w}\|_2^2 + \text{const.} \quad (\text{A} \cdot 13)$$

Finally, we have

$$H[x] = \frac{1}{2} \ln |\Sigma_x| + \text{const.}, \quad (\text{A} \cdot 14)$$

since $q(x)$ is assumed to be Gaussian [54]. Substituting (A·6), (A·10), (A·13), and (A·14) into (A·3), we obtain f' as in (11).

Now, we derive the update equation of each parameter

that maximizes f' given other parameters fixed.

(1) Auxiliary Parameter

Taking the derivative of the right-hand side of (11) with respect to \mathbf{V} and setting it to zero, we obtain

$$-\frac{\hat{a}}{2} \left(\sum_{k=1}^2 (\mathbf{G}_k \boldsymbol{\mu}_x \boldsymbol{\mu}_x^T \mathbf{G}_k^T + \mathbf{G}_k \boldsymbol{\Sigma}_x \mathbf{G}_k^T) - \mathbf{V}^{-2} \right) = 0, \quad (\text{A} \cdot 15)$$

where we have used $\|\mathbf{V}^{\frac{1}{2}} \mathbf{G}_k \boldsymbol{\mu}_x\|_2^2 = \text{tr}(\mathbf{V} \mathbf{G}_k \boldsymbol{\mu}_x \boldsymbol{\mu}_x^T \mathbf{G}_k^T)$ and properties on derivative of traces [54]. Simplifying this with respect to \mathbf{V} , we obtain

$$\mathbf{V}^{-2} = \sum_{k=1}^2 (\mathbf{G}_k \boldsymbol{\mu}_x \boldsymbol{\mu}_x^T \mathbf{G}_k^T + \mathbf{G}_k \boldsymbol{\Sigma}_x \mathbf{G}_k^T). \quad (\text{A} \cdot 16)$$

Noting that \mathbf{V} is the diagonal matrix of \mathbf{v} , we obtain

$$v_i = \left(\sum_{k=1}^2 (\mathbf{G}_k \boldsymbol{\mu}_x \boldsymbol{\mu}_x^T \mathbf{G}_k^T + \mathbf{G}_k \boldsymbol{\Sigma}_x \mathbf{G}_k^T) \right)_{ii}^{-\frac{1}{2}}. \quad (\text{A} \cdot 17)$$

Using that $(\mathbf{G}_k \boldsymbol{\mu}_x \boldsymbol{\mu}_x^T \mathbf{G}_k^T)_{ii} = (\mathbf{G}_k \boldsymbol{\mu}_x)_i^2$, we obtain each v_i as in (12).

(2) Weights

Taking the derivative of the right-hand side of (11) with respect to \hat{b} and setting it to zero, we obtain

$$\frac{n}{2} \hat{b}^{-1} - \frac{1}{2} (\|\mathbf{y} - \mathbf{H}_{\hat{w}} \boldsymbol{\mu}_x\|_2^2 + \text{tr}(\mathbf{H}_{\hat{w}} \boldsymbol{\Sigma}_x \mathbf{H}_{\hat{w}}^T)) = 0. \quad (\text{A} \cdot 18)$$

Simplifying this with respect to \hat{b} , we obtain

$$\hat{b} = n (\|\mathbf{y} - \mathbf{H}_{\hat{w}} \boldsymbol{\mu}_x\|_2^2 + \text{tr}(\mathbf{H}_{\hat{w}} \boldsymbol{\Sigma}_x \mathbf{H}_{\hat{w}}^T))^{-1}. \quad (\text{A} \cdot 19)$$

Using that $\|\mathbf{y} - \mathbf{H}_{\hat{w}} \boldsymbol{\mu}_x\|_2^2 = \sum_{i=1}^n (\mathbf{y} - \mathbf{H}_{\hat{w}} \boldsymbol{\mu}_x)_i^2$ and that $\text{tr}(\mathbf{H}_{\hat{w}} \boldsymbol{\Sigma}_x \mathbf{H}_{\hat{w}}^T) = \sum_{i=1}^n (\mathbf{H}_{\hat{w}} \boldsymbol{\Sigma}_x \mathbf{H}_{\hat{w}}^T)_{ii}$, we obtain \hat{b} as in (15). Likewise, we obtain \hat{a} and \hat{z} as in (14) and (13), respectively.

(3) Latent Image

Taking the derivative of the right-hand side of (11) with respect to $\boldsymbol{\mu}_x$ and setting it to zero, we obtain

$$-\hat{b} \mathbf{H}_{\hat{w}}^T (\mathbf{H}_{\hat{w}} \boldsymbol{\mu}_x - \mathbf{y}) - \hat{a} \sum_{k=1}^2 \mathbf{G}_k^T \mathbf{V} \mathbf{G}_k \boldsymbol{\mu}_x = 0, \quad (\text{A} \cdot 20)$$

where we have used the properties on derivative of squared norms [54]. Simplifying this with respect to $\boldsymbol{\mu}_x$, we obtain

$$\left(\hat{b} \mathbf{H}_{\hat{w}}^T \mathbf{H}_{\hat{w}} + \hat{a} \sum_{k=1}^2 \mathbf{G}_k^T \mathbf{V} \mathbf{G}_k \right) \boldsymbol{\mu}_x = \hat{b} \mathbf{H}_{\hat{w}}^T \mathbf{y}. \quad (\text{A} \cdot 21)$$

Premultiplying both sides with $(\hat{b} \mathbf{H}_{\hat{w}}^T \mathbf{H}_{\hat{w}} + \hat{a} \sum_{k=1}^2 \mathbf{G}_k^T \mathbf{V} \mathbf{G}_k)^{-1}$, we obtain $\boldsymbol{\mu}_x$ as in (16). Furthermore, taking the derivative of the right-hand side of (11) with respect to $\boldsymbol{\Sigma}_x$ and setting it to zero, we obtain

$$-\frac{\hat{b}}{2} \mathbf{H}_{\hat{w}}^T \mathbf{H}_{\hat{w}} - \frac{\hat{a}}{2} \sum_{k=1}^2 \mathbf{G}_k^T \mathbf{V} \mathbf{G}_k + \frac{1}{2} \boldsymbol{\Sigma}_x^{-1} = 0, \quad (\text{A} \cdot 22)$$

where we have used properties on derivatives of traces and logarithms of determinants [54]. Simplifying this with respect to $\boldsymbol{\Sigma}_x$, we obtain

$$\boldsymbol{\Sigma}_x^{-1} = \hat{b} \mathbf{H}_{\hat{w}}^T \mathbf{H}_{\hat{w}} + \hat{a} \sum_{k=1}^2 \mathbf{G}_k^T \mathbf{V} \mathbf{G}_k. \quad (\text{A} \cdot 23)$$

Inverting both sides, we obtain $\boldsymbol{\Sigma}_x$ as in (17).

(4) Field of Kernels

Substituting the definitions $\mathbf{H}_{\hat{w}} \boldsymbol{\mu}_x = \mathbf{H}_{\boldsymbol{\mu}_x} \hat{w}$ and $\text{tr}(\mathbf{H}_{\hat{w}} \boldsymbol{\Sigma}_x \mathbf{H}_{\hat{w}}^T) = \hat{w}^T \mathbf{H}_{\boldsymbol{\Sigma}_x}^2 \hat{w}$ into (11), we obtain

$$\begin{aligned} f'[q(x), q(w), q(b), q(a), q(a), v] \\ = -\frac{\hat{b}}{2} (\|\mathbf{y} - \mathbf{H}_{\boldsymbol{\mu}_x} \hat{w}\|_2^2 + \hat{w}^T \mathbf{H}_{\boldsymbol{\Sigma}_x}^2 \hat{w}) \\ - \frac{\hat{z}}{2} \sum_{k=1}^2 \|\mathbf{F}_k \hat{w}\|_2^2 + \text{const.}, \end{aligned} \quad (\text{A} \cdot 24)$$

where the terms constant with respect to \hat{w} are gathered. Then, taking the derivative of the right-hand side with respect to \hat{w} and setting it to zero, we obtain

$$-\hat{b} (\mathbf{H}_{\boldsymbol{\mu}_x}^T (\mathbf{H}_{\boldsymbol{\mu}_x} \hat{w} - \mathbf{y}) + \mathbf{H}_{\boldsymbol{\Sigma}_x}^2 \hat{w}) - \hat{z} \sum_{k=1}^2 \mathbf{F}_k^T \mathbf{F}_k \hat{w} = 0, \quad (\text{A} \cdot 25)$$

where we have used properties on derivative of squared norms [54]. Simplifying this with respect to \hat{w} , we obtain

$$\left(\hat{b} (\mathbf{H}_{\boldsymbol{\mu}_x}^T \mathbf{H}_{\boldsymbol{\mu}_x} + \mathbf{H}_{\boldsymbol{\Sigma}_x}^2) + \hat{z} \sum_{k=1}^2 \mathbf{F}_k^T \mathbf{F}_k \right) \hat{w} = \hat{b} \mathbf{H}_{\boldsymbol{\mu}_x}^T \mathbf{y} \quad (\text{A} \cdot 26)$$

Premultiplying both sides with $(\hat{b} (\mathbf{H}_{\boldsymbol{\mu}_x}^T \mathbf{H}_{\boldsymbol{\mu}_x} + \mathbf{H}_{\boldsymbol{\Sigma}_x}^2) + \hat{z} \sum_{k=1}^2 \mathbf{F}_k^T \mathbf{F}_k)^{-1}$, we obtain \hat{w} as in (16).



Motoharu Sonogashira received B.S. and M.S. degrees in informatics from Kyoto University, Kyoto, Japan in 2013 and 2015, respectively. He is currently pursuing his Ph.D. degree in informatics at Kyoto University. His research interests include image restoration, blind deconvolution, superresolution, optical flow, and variational Bayes.



Masaaki Iiyama received the B.S. degree in engineering informatics, and the M.S. and Ph.D. degrees in informatics from Kyoto University, Kyoto, Japan, in 1998, 2000, and 2006, respectively. Currently, he is an Associate Professor with the Academic Center for Computing and Media Studies, Kyoto University. His research interests include 3-D shape reconstruction, reflectance and scattering modeling, and computational photography.



Michihiko Minoh is a professor at Academic Center for Computing and Media Studies, Kyoto University, Japan. He received the BEng, MEng, and DEng degrees in information science from Kyoto University, in 1978, 1980, and 1983, respectively. He served as director of ACCMS from April 2006 to March 2010 and concurrently served as vice director in the Kyoto University Presidents Office from October 2008 to September 2010. Since October 2010, he has been vice-president, chief information officer at

Kyoto University, and director-general at Institute for Information Management and Communication, Kyoto University. His research interests include a variety area of Image Processing, Artificial Intelligence and Multimedia Applications, particularly, model centered framework for the computer system to help visual communication among humans and information media structure for human communication. He is a member of Information Processing Society of Japan, Institute of Electronics, Information and Communication Engineers of Japan, the IEEE Computer Society and Communication Society, and ACM.