Bayesian Blind Deconvolution From Differently Exposed Image Pairs

Sevket Derin Babacan, *Member, IEEE*, Jingnan Wang, Rafael Molina, *Member, IEEE*, and Aggelos K. Katsaggelos, *Fellow, IEEE*

***Abstract*—Photographs acquired under low-lighting conditions require long exposure times and therefore exhibit significant blurring due to the shaking of the camera. Using shorter exposure times results in sharper images but with a very high level of noise. In this paper, we address the problem of utilizing two such images in order to obtain an estimate of the original scene and present a novel blind deconvolution algorithm for solving it. We formulate the problem in a hierarchical Bayesian framework by utilizing prior knowledge on the unknown image and blur, and also on the dependency between the two observed images. By incorporating a fully Bayesian analysis, the developed algorithm estimates all necessary model parameters along with the unknown image and blur, such that no user-intervention is needed. Moreover, we employ a variational Bayesian inference procedure, which allows for the statistical compensation of errors occurring at different stages of the restoration, and also provides uncertainties of the estimates. Experimental results with synthetic and real images demonstrate that the proposed method provides very high quality restoration results and compares favorably to existing methods even though no user supervision is needed.**

***Index Terms*—Bayesian methods, blind deconvolution, image stabilization, parameter estimation, variational distribution approximations.**

I. Introduction

TAKING high-quality photographs under low-lighting conditions is a major challenge. A longer exposure time than usual is required to obtain an image with low-noise, but any motion of the camera during exposure causes blur in the recorded image. On the other hand, a short exposure time will result in an image with a very high level of noise. Possible hardware-based solutions include increasing the light sensitivity (ISO) of the camera sensor, which increases the noise level; increasing the aperture, which results in a smaller depth of field in the acquired image; and using a tripod to stabilize the camera which is not practical in many cases. Additionally, many digital cameras incorporate optical image stabilizers, either inside the camera body or inside the lens, that significantly help in reducing degradations caused by hand-held photography. However, when the exposure times are too long, as might be required in some conditions, these hardware based solutions can not provide satisfactory results. In these cases, digital image stabilization methods, applied at a postprocessing stage, provide a powerful means to obtain high-quality images using the low-quality observations.

In case a photograph is taken using a long exposure time under dim lighting, the resulting blur in the image can be removed by utilizing a single-image blind deconvolution algorithm. A number of methods are proposed for blind deconvolution of a single observation (see, for example, [2] for a recent review), and the specific case of restoring images degraded by camera shake is addressed in [3]–[7]. However, due to the challenging nature of the problem, obtaining a high-quality restoration result is very hard in most cases and requires significant user-supervision. Additionally, due to the difficulty in estimating the camera shake degradation, the restored images generally exhibit deconvolution artifacts such as ringing.

Another possible approach is to use a short exposure time to prevent blur at the expense of high noise, and then apply denoising algorithms to the sharp short-exposed image to remove the noise. Many advanced denoising methods are available in the literature (see, for instance, [8]–[10]). However, the noise level in such short-exposed images is generally so high that features of the underlying image are concealed, and the denoising algorithms cannot easily separate image and noise. An additional, and possibly more important problem is that due to the short exposure time, the images generally have low contrast and the colors might be partially lost.

Recently, deconvolution methods have been proposed that utilize an image pair instead of a single observation, where two images are taken with different exposure times [11]. Some digital cameras have exposure bracketing features which allow the user to acquire consecutive photographs with different exposure settings, which was mainly developed for high dynamic range applications [12], [13], but can also be used for image stabilization. Utilizing two images reduces the ill-posedness of the deconvolution problem, and generally results in much higher quality restorations than methods utilizing single observations [14]. A wide range of algorithms exists for the general problem of multiframe blind deconvolution [2], [15], [16]. The specific case of blind deconvolution from a pair of short- and long-exposure images has been considered in [11], [17], [18]. In [11], the blur point spread function (PSF) is first identified using the long-exposure image and the denoised version of the short-exposure

Manuscript received May 20, 2009; revised October 26, 2009; accepted April 30, 2010. Date of publication June 07, 2010; date of current version October 15, 2010. This work was supported in part by the "Comisión Nacional de Ciencia y Tecnología" under Contract TIC2007-65533 and the Spanish research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018). Preliminary results of this work were presented at the IEEE International Conference on Image Processing, Cairo, Egypt, July, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ercan E Kuruoglu.

S. D. Babacan was with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208 USA. He is now with the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Champaign, IL 61820 USA (e-mail: dbabacan@illinois.edu).

J. Wang and A. K. Katsaggelos are with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208 USA (e-mail: jwa180@eecs.northwestern.edu; aggk@eecs.northwestern.edu).

R. Molina is with the Departamento de Ciencias de la Computación e I.A. Universidad de Granada, Spain (e-mail: rms@decsai.ugr.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2010.2052263

image, where Tikhonov regularization and hysteresis thresholding is utilized to regularize the solution. This PSF estimate is then used in a classical image restoration method [19], [20] in order to obtain an estimate of the original image from the long-exposure image. A joint identification method is proposed in [17], where the unknown image and the PSF are estimated simultaneously. The image is modeled using a total-variation (TV) based prior, and the blur PSFs are estimated by imposing the constraint that the blur in the short exposure image is very small. No explicit blur model is utilized in this work, and denoising is applied to the blur estimates by thresholding in an ad hoc manner to enhance the estimates. Finally, sparsity priors with continuity constraints on the blurs are utilized in [18], and the image is modeled using a mixture-of-Gaussians prior on the image derivatives. However, the model is derived in a somewhat ad hoc manner, and the resulting algorithm has many parameters to tune, which makes it hard to apply to a wide range of images.

This paper addresses the problem of blind deconvolution from a short- and long-exposed image pair. We provide a systematic modeling of the unknowns within a novel hierarchical Bayesian formulation and develop a blind deconvolution algorithm which jointly estimates the unknown image and blur. We utilize a TV-prior on the image to model natural image statistics and to achieve robustness in the algorithm. The blur in the long-exposed image is modeled using a mixture prior which imposes both sparsity and positivity on the estimated blur PSF. We also model the coupling between the long- and short-exposed images using an additional observation model. Moreover, we incorporate a fully-Bayesian approach, where all required model parameters are estimated along with the unknowns. As a result, the proposed algorithm does not require user-intervention and the restoration process is adaptively steered between the long- and short-exposed images. Finally, we incorporate a variational Bayesian analysis, which provides estimates of the distributions of the unknowns. These distributions are implicitly used to incorporate the uncertainties of the estimates into the algorithm and to compensate for the estimation errors. We demonstrate with both synthetic and real image experiments that the proposed method provides very high quality restoration results and compares favorably to existing methods.

The rest of this paper is organized as follows. In Section II we formulate the image acquisition processes mathematically. The unknown variables in our model are cast into a hierarchical Bayesian framework as presented in Section III. The variational inference to estimate the unknowns and the proposed algorithm are presented in Section IV. Experimental results are presented in Section IV and conclusions are drawn in Section VII.

II. Problem Formulation

The degradations in the image pair can be modeled using a linear and space invariant degradation model, by assuming that the blur is mainly caused by the shake of the camera during the long exposure time. Under this assumption, the observation processes can mathematically be expressed as follows:

$$\mathbf{y}_1 = \mathbf{H}\mathbf{x} + \mathbf{n}_1 \tag{1}$$

$$\mathbf{y}_2 = \lambda_1 \mathbf{C}\mathbf{x} + \lambda_2 \mathbf{1} + \mathbf{n}_2 \tag{2}$$

where $\mathbf{y}_1$ and $\mathbf{y}_2$ are the observed images, $\mathbf{x}$ the unknown original image, $\mathbf{n}_1$ and $\mathbf{n}_2$ the noise components, and the rest of the quantities are explained in the following. We use matrix-vector notation throughout the paper, so that the images $\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{x}$, $\mathbf{n}_1$, and $\mathbf{n}_2$ are $N \times 1$ vectors, where N is the number of pixels in each image. The $N \times N$ matrix $\mathbf{H}$ models the blur point spread function (PSF) $\mathbf{h}$, which has support $M \leq N$. The selection of the support M of the PSF is important and will be explained in the experimental results section. Note that the explicit construction of the matrix $\mathbf{H}$ is not needed but it is used for notation purposes only.

Generally, the average luminance levels of the observations $\mathbf{y}_1$ and $\mathbf{y}_2$ are significantly different due to the different exposure times, and furthermore the images have to be geometrically registered. These geometric and photometric differences between the observed images are represented in (2) by the matrix $\mathbf{C}$ and the parameters λ_1 and λ_2, respectively. The geometric registration (or warping) matrix $\mathbf{C}$ represents the motion between the observations $\mathbf{y}_1$ and $\mathbf{y}_2$, and the parameters λ_1 and λ_2 represent the illumination differences and therefore, the photometric registration between the observations, and $\mathbf{1}$ is an $N \times 1$ vector of ones.

In this work we assume that the photometric and geometric calibration between the images $\mathbf{y}_1$ and $\mathbf{y}_2$ is calculated in a preprocessing stage, as is the case with the existing methods [11], [17], [18], [21]. The geometric registration can be performed using methods specifically designed for blurred/non-blurred image pairs (see, for example, [22]). In this work, we performed the geometric registration using publicly available registration software [23]. Alternatively, the registration algorithm in [24] can be used, as suggested by [21]. As demonstrated by our experimental results, crude initial registrations still result in high quality results due to the blur estimation process, which compensates for possible misalignments by appropriately shifting the estimated blur kernels. Photometric registration between the images is performed in a similar fashion by estimating the parameters λ_1 and λ_2 using the least squares solution utilizing the approximation $\mathbf{y}_2 \approx \lambda_1 \mathbf{C}\mathbf{y}_1 + \lambda_2 \mathbf{1}$ and the RANSAC algorithm [25], [26].

After the observed images are corrected using the computed geometric and photometric registration, the observation models in (1) and (2) can be simplified using $\mathbf{C} = \mathbf{I}$, $\lambda_1 = 1$, and $\lambda_2 = 0$, that is

$$\mathbf{y}_1 = \mathbf{H}\mathbf{x} + \mathbf{n}_1 \tag{3}$$

$$\mathbf{y}_2 = \mathbf{x} + \mathbf{n}_2. \tag{4}$$

These observation models will be utilized in the rest of the paper. Using (3) and (4), the blind deconvolution problem is then to find estimates of $\mathbf{x}$ and $\mathbf{h}$ utilizing $\mathbf{y}_1$ and $\mathbf{y}_2$ and prior knowledge about $\mathbf{x}$, $\mathbf{h}$, $\mathbf{n}_1$, and $\mathbf{n}_2$.

III. Hierarchical Bayesian Model

The proposed hierarchical model is composed of two stages. In the first stage, prior distributions are utilized on the unknown image $\mathbf{x}$ and blur $\mathbf{h}$, and conditional distributions are utilized for the observations $\mathbf{y}_1$ and $\mathbf{y}_2$. These distributions in the first stage depend on certain parameters, called *hyperparameters*, which are modeled by hyperprior distributions in the second stage. The explicit forms of these distributions are presented in the following subsections.

A. First Stage: Observation Models

The observation model we use in this paper has the Gaussian form shown in (5) at the bottom of the page, where Σ is the covariance matrix of the distribution of $\mathbf{y}_1$ and $\mathbf{y}_2$ given $\mathbf{x}$, $\mathbf{h}$, β_1, β_2, and β_{12}. Before calculating $|\Sigma|$ let us describe the energy terms in this probability distribution.

The first energy term, $\beta_1 \parallel \mathbf{y}_1 - \mathbf{H}\mathbf{x} \parallel^2$, is the energy of the probability distribution associated with the observation in (3). The second energy term, $\beta_2 \parallel \mathbf{y}_2 - \mathbf{x} \parallel^2$, corresponds to the energy of the probability distribution associated with the observation in (4). To incorporate the third energy term, $\beta_{12} \parallel \mathbf{y}_1 - \mathbf{H}\mathbf{y}_2 \parallel^2$, we make use of the coprimeness condition employed in some multichannel blind deconvolution methods (see, for example, [27]). Combining (3) and (4) we obtain

$$\mathbf{y}_1 - \mathbf{H}\mathbf{y}_2 | \mathbf{h}, \beta_1, \beta_2 \sim \mathcal{N}(\mathbf{0}, \beta_1^{-1}\mathbf{I} + \beta_2^{-1}\mathbf{H}\mathbf{H}^T). \quad (6)$$

Note that $\mathbf{y}_1 - \mathbf{H}\mathbf{y}_2$ does not represent a new set of observations since its conditional distribution has been obtained from the already observed $\mathbf{y}_1$ and $\mathbf{y}_2$. However, $\mathbf{y}_1 - \mathbf{H}\mathbf{y}_2$ provides additional information to the one provided by $\mathbf{y}_1$ and $\mathbf{y}_2$ when the energy $(\mathbf{y}_1 - \mathbf{H}\mathbf{y}_2)^T(\beta_1^{-1}\mathbf{I} + \beta_2^{-1}\mathbf{H}\mathbf{H}^T)^{-1}(\mathbf{y}_1 - \mathbf{H}\mathbf{y}_2)$ is replaced by $\beta_{12}(\mathbf{y}_1 - \mathbf{H}\mathbf{y}_2)^T(\mathbf{y}_1 - \mathbf{H}\mathbf{y}_2)$, with $\beta_{12} > 0$.

The conditional model in (5) does not exactly result from the models shown in (3) and (4) but it is still a conditional probability distribution and so it can be used in combination with the prior models for image and blur to carry out Bayesian inference on the dual exposure problem. Note that the third energy term in the above observation model provides information about the blurring function only and in some sense helps to decouple the estimation of $\mathbf{x}$ from the estimation of $\mathbf{h}$ (unlike the first energy term in (5)).

Let us now proceed to calculate the partition function. We have that

$$\Sigma = \begin{pmatrix} (\beta_1 + \beta_{12})\mathbf{I} & -\beta_{12}\mathbf{H} \\ -\beta_{12}\mathbf{H}^T & \beta_2\mathbf{I} + \beta_{12}\mathbf{H}^T\mathbf{H} \end{pmatrix}^{-1} \quad (7)$$

and consequently

$$|\Sigma|^{-1} = |(\beta_1+\beta_{12})\mathbf{I}||\beta_2\mathbf{I}+\beta_{12}\mathbf{H}^T\mathbf{H} - \frac{\beta_{12}^2}{\beta_1+\beta_{12}}\mathbf{H}^T\mathbf{H}|. \quad (8)$$

Then since $\log|\beta_2(\beta_1 + \beta_{12})\mathbf{I} + \beta_{12}\beta_1\mathbf{H}^T\mathbf{H}|$ changes slowly as $\mathbf{H}^T\mathbf{H}$ changes we write

$$|\Sigma|^{-1/2} \approx |(\beta_1\beta_2 + \beta_1\beta_{12} + \beta_2\beta_{12})\mathbf{I}|^{1/2}. \quad (9)$$

Finally, for an observation model produced by three independent models and generating a total of $2N$ observations, each of them with the same number of observations, $2N/3$, we can write from (9)

$$|\Sigma|^{-1/2} \approx (\beta_1\beta_2\beta_{12})^{N/3}. \quad (10)$$

We have experimentally observed that using this modeling produces a good restored image and a good estimation of the blur. Consequently we write (11), shown at the bottom of the page. It is important to note that the additional observation model incorporates the strong dependency between the observations $\mathbf{y}_1$ and $\mathbf{y}_2$ into the inference procedure (see [28] for a similar approach). Finally, this model can also potentially account for small registration errors between the observed images $\mathbf{y}_1$ and $\mathbf{y}_2$.

B. First Stage: Prior Model on the Blur

Since the blur is mainly caused by the shaking of the camera during the long exposure time, it exhibits the characteristics of the nonuniform motion blur. Hence, it is expected to be very sparse, i.e., most of the PSF coefficients being zero or very small. In order to exploit this information, we utilize a mixture prior of D exponential distributions on each PSF coefficient, that is

$$\mathrm{p}(\mathbf{h}|\{\tau_{jd}\},\{\sigma_{jd}\}) = \prod_{j=1}^{M}\left(\sum_{d=1}^{D}\tau_{jd}\,\mathrm{Expon}\;(h_j \mid \sigma_{jd})\right) \quad (12)$$

with τ_{jd} the mixture coefficients for each pixel j and

$$\mathrm{Expon}\,(h_j \mid \sigma_{jd}) = \begin{cases} \sigma_{jd}\exp(-\sigma_{jd}h_j), & \text{if } h_j \geq 0 \\ 0, & \text{if } h_j < 0. \end{cases} \quad (13)$$

with σ_{jd} the parameters of each exponential distribution.

Note that this blur prior enforces sparsity to a great extent, and the degree of sparsity is increased by increasing the number of mixture coefficients D [29]. In addition to imposing sparsity, note that (13) also imposes positivity on the blur coefficients h_j. This property makes the prior especially useful, since unlike most previous works the positivity constraint is imposed during the formulation and subsequent optimization process, and not artificially after the optimization, which can move the estimates

$$\mathrm{p}(\mathbf{y}_1,\mathbf{y}_2|\mathbf{x},\mathbf{h},\beta_1,\beta_2,\beta_{12}) = (2\pi)^{-N}|\Sigma|^{-1/2}\exp\left[-\frac{\beta_1}{2}\parallel\mathbf{y}_1-\mathbf{H}\mathbf{x}\parallel^2 - \frac{\beta_2}{2}\parallel\mathbf{y}_2-\mathbf{x}\parallel^2 - \frac{\beta_{12}}{2}\parallel\mathbf{y}_1-\mathbf{H}\mathbf{y}_2\parallel^2\right] \quad (5)$$

$$\begin{aligned}\mathrm{p}(\mathbf{y}_1,\mathbf{y}_2|\mathbf{x},\mathbf{h},\beta_1,\beta_2,\beta_{12}) \propto \beta_1^{N/3}\beta_2^{N/3}\beta_{12}^{N/3}\exp\left[-\frac{\beta_1}{2}\parallel\mathbf{y}_1-\mathbf{H}\mathbf{x}\parallel^2\right]\exp\left[-\frac{\beta_2}{2}\parallel\mathbf{y}_2-\mathbf{x}\parallel^2\right] \\ \times\exp\left[-\frac{\beta_{12}}{2}\parallel\mathbf{y}_1-\mathbf{H}\mathbf{y}_2\parallel^2\right]. \quad (11)\end{aligned}$$

away from their optimal values. Note that this mixture-of-exponentials prior has also been utilized before for modeling PSFs resulting from camera shake [3], [29] and in independent component analysis [29].

C. First Stage: Prior Model on the Image

The unknown image $\mathbf{x}$ is expected to be mostly smooth except at the locations of discontinuities (e.g., edges). Therefore, as the prior model on the image $\mathbf{x}$, we utilize the total variation function because it preserves the edges in the image by not over-penalizing discontinuities while imposing smoothness [30]. Specifically, we utilize the following approximation of the TV prior [31]

$$\mathrm{p}(\mathbf{x}|\alpha_{\mathrm{im}}) = c\,\alpha_{\mathrm{im}}^{N/2}\exp\left[-\frac{1}{2}\alpha_{\mathrm{im}}\mathrm{TV}(\mathbf{x})\right] \tag{14}$$

where c is a constant and

$$\mathrm{TV}(\mathbf{x}) = \sum_{j=1}^{N}\sqrt{(\Delta_j^h(\mathbf{x}))^2 + (\Delta_j^v(\mathbf{x}))^2}. \tag{15}$$

The operators $\Delta_j^h(\mathbf{x})$ and $\Delta_j^v(\mathbf{x})$ correspond to, respectively, horizontal and vertical first order differences, at pixel j, that is, $\Delta_j^h(\mathbf{x}) = x_j - x_{l(j)}$ and $\Delta_j^v(\mathbf{x}) = x_j - x_{a(j)}$, where $l(j)$ and $a(j)$ denote the nearest neighbors of j, to the left and above, respectively.

D. Second Stage: Hyperpriors on the Hyperparameters

In the second stage of the hierarchical model, we model the hyperparameters α_{im}, β_1, β_2, β_{12}, $\{\sigma_{jd}\}$ and $\{\tau_{jd}\}$ by hyperprior distributions. In Bayesian models, hyperprior distributions are generally chosen to be conjugate distributions, i.e., they have the same functional form as the product of the conditional distribution and the priors. This choice of hyperpriors simplifies the analytical derivation of the inference procedure. Based on this, we employ conjugate Gamma distributions for the hyperpriors $\mathrm{p}(\alpha_{\mathrm{im}})$, $\mathrm{p}(\beta_1)$, $\mathrm{p}(\beta_2)$, $\mathrm{p}(\beta_{12})$, $\mathrm{p}(\sigma_{jd})$, and Dirichlet distributions on the mixture coefficients $\mathrm{p}(\tau_{jd})$, that is shown in (16)–(21) at the bottom of the page, with shape parameters $a^o_{\alpha_{\mathrm{im}}}$, $a^o_{\beta_1}$, $a^o_{\beta_2}$, $a^o_{\beta_{12}}$ and inverse scale parameters $b^o_{\alpha_{\mathrm{im}}}$, $b^o_{\beta_1}$, $b^o_{\beta_2}$, $b^o_{\beta_{12}}$. The shape and inverse scale parameters of the Gamma distributions are set to a small common value compared to the range of intensity values of the image (e.g., 0.001 for $\mathbf{x}$ in $[0, 255]$ range), and $c^o_{\tau_{jd}}$ is set to 1 to obtain vague hyperpriors which make the estimation process rely more on the observations than on prior knowledge. Note, however, that these hyperprior distributions are very flexible in incorporating additional information provided by the user. If some prior knowledge on the value of some of the hyperparameters (for instance, about the noise variances in the observed images) is available, this information can easily be incorporated into the estimation procedure by choosing the shape and inverse scale parameters of the corresponding distributions accordingly (see, for example, [31], [32] for such incorporation of prior knowledge). Moreover, note that using nonzero values for the shape and inverse scale parameters aid in avoiding trivial solutions such as delta PSF estimates. Note also that the additional observation model in (11) has an important role in preventing the blur from being estimated as a delta function unless $\mathbf{y}_1 - \mathbf{y}_2$ can be considered as Gaussian independent noise.

Finally, combining the first and second stage of the hierarchical model we obtain (22), shown at the bottom of the page.

IV. Variational Bayesian Inference

In Bayesian formulations, the inference is based on the posterior distribution, which in our case is intractable. Therefore, in this work we utilize variational distribution approximations. Let us denote by Θ the set of unknowns, i.e., $\Theta = \{\mathbf{x}, \mathbf{h}, \alpha_{\mathrm{im}}, \beta_1, \beta_2, \beta_{12}, \{\sigma_{jd}\}, \{\tau_{jd}\}\}$. The goal is to approximate the posterior distribution $\mathrm{p}(\Theta|\mathbf{y}_1, \mathbf{y}_2)$ by another distribution $\mathrm{q}(\Theta)$ which allows a tractable analysis. The approximating distribution $\mathrm{q}(\Theta)$ is found by minimizing

$$\mathrm{p}(\alpha_{\mathrm{im}}) = \mathrm{Gamma}\left(\alpha_{\mathrm{im}}|a^o_{\alpha_{\mathrm{im}}}, b^o_{\alpha_{\mathrm{im}}}\right) = \frac{(b^o_{\alpha_{\mathrm{im}}})^{a^o_{\alpha_{\mathrm{im}}}}}{\Gamma(a^o_{\alpha_{\mathrm{im}}})}\alpha_{\mathrm{im}}^{a^o_{\alpha_{\mathrm{im}}}-1}\exp\left[-\alpha_{\mathrm{im}}b^o_{\alpha_{\mathrm{im}}}\right] \tag{16}$$

$$\mathrm{p}(\beta_1) = \mathrm{Gamma}\left(\beta_1|a^o_{\beta_1}, b^o_{\beta_1}\right) \tag{17}$$

$$\mathrm{p}(\beta_2) = \mathrm{Gamma}\left(\beta_2|a^o_{\beta_2}, b^o_{\beta_2}\right) \tag{18}$$

$$\mathrm{p}(\beta_{12}) = \mathrm{Gamma}\left(\beta_{12}|a^o_{\beta_{12}}, b^o_{\beta_{12}}\right) \tag{19}$$

$$\mathrm{p}(\sigma_{jd}) = \mathrm{Gamma}\left(\sigma_{jd}|a^o_{\sigma_{jd}}, b^o_{\sigma_{jd}}\right), \qquad j = 1, \ldots, M,\ d = 1, \ldots, D \tag{20}$$

$$\mathrm{p}(\{\tau_{jd}\}_{d=1}^{D}) = \mathrm{Dirichlet}\left(\{\tau_{jd}\}_{d=1}^{D}|c^o_{\tau_{jd}}\right), \qquad j = 1, \ldots, M \tag{21}$$

$$\begin{aligned}\mathrm{p}(\mathbf{x}, \mathbf{h}, \mathbf{y}_1, \mathbf{y}_2, \alpha_{\mathrm{im}}, \beta_1, \beta_2, \beta_{12}, \{\tau_{jd}\}, \{\sigma_{jd}\}) &= \mathrm{p}(\mathbf{x}|\alpha_{\mathrm{im}})\,\mathrm{p}(\mathbf{h}|\{\tau_{jd}\}, \{\sigma_{jd}\})\mathrm{p}(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}, \mathbf{h}, \beta_1, \beta_2, \beta_{12}) \\ &\quad \times \mathrm{p}(\alpha_{\mathrm{im}})\mathrm{p}(\beta_1)\,\mathrm{p}(\beta_2)\mathrm{p}(\beta_{12})\prod_{j=1}^{M}\left[\mathrm{p}(\{\tau_{jd}\}_{d=1}^{D})\prod_{d=1}^{D}\mathrm{p}(\sigma_{jd})\right].\end{aligned} \tag{22}$$

the Kullback-Leibler (KL) divergence between $\mathrm{q}(\Theta)$ and $\mathrm{p}(\Theta|\mathbf{y}_1,\mathbf{y}_2)$, which is given by

$$C_{KL}(\mathrm{q}(\Theta)\parallel \mathrm{p}(\Theta|\mathbf{y}_1,\mathbf{y}_2)) = \int \mathrm{q}(\Theta)\log\left(\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta|\mathbf{y}_1,\mathbf{y}_2)}\right)d\Theta \quad (23)$$

$$= \int \mathrm{q}(\Theta)\log\left(\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta,\mathbf{y}_1,\mathbf{y}_2)}\right)d\Theta + \text{const.} \quad (24)$$

Generally, the only assumption in variational Bayesian analysis is that the approximating distribution $\mathrm{q}(\Theta)$ is factorizable. In this work, we use the following factorization:

$$\mathrm{q}(\Theta) = \mathrm{q}(\mathbf{x})\,\mathrm{q}(\mathbf{h})\,\mathrm{q}(\alpha_{\mathrm{im}})\,\mathrm{q}(\beta_1) \times \mathrm{q}(\beta_2)\,\mathrm{q}(\beta_{12})\prod_{j=1}^{M}\left[\mathrm{q}\left(\{\tau_{jd}\}_{d=1}^{D}\right)\prod_{d=1}^{D}\mathrm{q}(\sigma_{jd})\right] \quad (25)$$

with

$$\mathrm{q}(\mathbf{h}) = \prod_{j=1}^{M}\mathrm{q}(h_j). \quad (26)$$

Unfortunately the general results from variational Bayesian analysis cannot be directly utilized in this work, since the TV and mixture priors in our model render the calculation of the KL divergence in (24) not possible. The problems caused by the TV prior can be avoided by utilizing a majorization-minimization approach, whose details are given in [31]. We will provide a brief overview here as follows. Let us consider the following inequality, derived from the geometric-arithmetic mean inequality, which states that for any real numbers $a \geq 0$ and $b > 0$

$$\sqrt{ab} \leq \frac{a+b}{2} \Rightarrow \sqrt{a} \leq \frac{a+b}{2\sqrt{b}}. \quad (27)$$

Let us also define for α_{im}, $\mathbf{x}$, and an $N-$ dimensional vector $\mathbf{w} \in (R^+)^N$, with components w_i, $i = 1,\ldots,N$, the following functional:

$$\mathbf{M}(\alpha_{\mathrm{im}},\mathbf{x},\mathbf{w}) = c\,\alpha_{\mathrm{im}}^{N/2} \times \exp\left[-\frac{\alpha_{\mathrm{im}}}{2}\sum_i \frac{(\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2 + w_i}{\sqrt{w_i}}\right] \quad (28)$$

where c is the same constant as in (14). Using $a = (\Delta_i^h(\mathbf{x}))^2 + (\Delta_i^v(\mathbf{x}))^2$ and $b = w_i$ in the inequality (27), it can be seen that the functional $\mathbf{M}(\alpha_{\mathrm{im}},\mathbf{x},\mathbf{w})$ is a lower bound of the image prior $\mathrm{p}(\mathbf{x}|\alpha_{\mathrm{im}})$, that is

$$\mathrm{p}(\mathbf{x}|\alpha_{\mathrm{im}}) \geq \mathbf{M}(\alpha_{\mathrm{im}},\mathbf{x},\mathbf{w}). \quad (29)$$

The quadratic form of the bounding functional $\mathbf{M}(\alpha_{\mathrm{im}},\mathbf{x},\mathbf{w})$ renders the analytical derivation of the Bayesian inference tractable. Using the lower bound in (28), a lower bound of the joint probability distribution in (22) can be found in (30) at the bottom of the page, which leads to the following upper bound for the KL divergence in (24):

$$C_{KL}(\mathrm{q}(\Theta)\parallel \mathrm{p}(\Theta|\mathbf{y}_1,\mathbf{y}_2)) \leq C_{KL}(\mathrm{q}(\Theta)\parallel \mathbf{F}(\Theta,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2)) + \text{const.} \quad (31)$$

The upper bound $C_{KL}(\mathrm{q}(\Theta) \parallel \mathbf{F}(\Theta,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2))$ can be made tighter by minimizing it with respect to $\mathbf{w}$, since

$$C_{KL}(\mathrm{q}(\Theta)\parallel \mathrm{p}(\Theta|\mathbf{y}_1,\mathbf{y}_2)) \leq \min_{\mathbf{w}} C_{KL}(\mathrm{q}(\Theta)\parallel \mathbf{F}(\Theta,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2)) + \text{const.} \quad (32)$$

Therefore, by minimizing the upper bound $C_{KL}(\mathrm{q}(\Theta) \parallel \mathbf{F}(\Theta,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2))$ with respect to both $\mathrm{q}(\Theta)$ and $\mathbf{w}$, the upper bound can iteratively be made closer to the KL distance $C_{KL}(\mathrm{q}(\Theta) \parallel \mathrm{p}(\Theta|\mathbf{y}_1,\mathbf{y}_2))$. Thus, this upper bound can be used as an approximation to the original KL distance, and variational Bayesian analysis can be performed using this upper bound instead (see [31] for details on the theoretical justification). For each unknown $\theta \in \Theta$, the distribution approximation $\mathrm{q}(\theta)$ can then be found by alternating the minimization of $C_{KL}(\mathrm{q}(\Theta) \parallel \mathbf{F}(\Theta,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2))$ with respect to each $\mathrm{q}(\theta)$ by holding $\mathrm{q}(\Theta_\theta)$ constant, where Θ_θ denotes the set Θ with θ removed from the set. This approach results in the following general solution [33]:

$$\mathrm{q}(\theta) = \text{const} \times \exp\left(\mathrm{E}_{\mathrm{q}(\Theta_\theta)}\left[\log \mathbf{F}(\Theta,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2)\right]\right) \quad (33)$$

where $\mathrm{E}_{\mathrm{q}(\Theta_\theta)}[\cdot]$ denotes the expectation with respect to the distribution $\mathrm{q}(\Theta_\theta)$. In order to solve (33), an additional approximation is needed when using mixture priors. Specifically, we utilize Jensen's inequality as follows [29]:

$$\log\left[\prod_{j=1}^{M}\left(\sum_{d=1}^{D}\tau_{jd}\,\mathrm{Expon}\,(h_j \mid \sigma_{jd})\right)\right] \geq \sum_{j=1}^{M}\sum_{d=1}^{D}\mu_{jd}\log\left(\frac{\tau_{jd}}{\mu_{jd}}\mathrm{Expon}\,(h_j \mid \sigma_{jd})\right) \quad (34)$$

with $\sum_{d=1}^{D}\mu_{jd} = 1$, $j = 1,\ldots,M$. An analysis of the closeness of this bound can be found in [29]. The auxiliary variables μ_{jd} need to be computed along with the unknowns Θ, as will be shown in the next section. Using (34), we obtain a lower bound of $\log \mathbf{F}(\Theta,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2)$ as follows. Let us denote by $\bar{\mathbf{F}}(\Theta,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2)$ the product of

$$\begin{aligned}\mathrm{p}(\Theta) \geq & \mathbf{M}(\alpha_{\mathrm{im}},\mathbf{x},\mathbf{w})\mathrm{p}(\mathbf{h}|\{\tau_{jd}\},\{\sigma_{jd}\})\mathrm{p}(\mathbf{y}_1,\mathbf{y}_2|\mathbf{x},\mathbf{h},\beta_1,\beta_2,\beta_{12}) \\ & \times \mathrm{p}(\alpha_{\mathrm{im}})\mathrm{p}(\beta_1)\,\mathrm{p}(\beta_2)\mathrm{p}(\beta_{12})\prod_{j=1}^{M}\left[\mathrm{p}\left(\{\tau_{jd}\}_{d=1}^{D}\right)\prod_{d=1}^{D}\mathrm{p}(\sigma_{jd})\right] \\ = & \mathbf{F}(\Theta,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2)\end{aligned} \quad (30)$$

the terms in $\mathbf{F}(\Theta,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2)$ except $\mathrm{p}(\mathbf{h}|\{\tau_{jd}\},\{\sigma_{jd}\})$, that is, $\mathbf{F}(\Theta,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2) = \mathrm{p}(\mathbf{h}|\{\tau_{jd}\},\{\sigma_{jd}\})\bar{\mathbf{F}}(\Theta,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2)$. Then, (see (35) and (36) at the bottom of the page) with $\boldsymbol{\mu} = \{\mu_{jd}\,|\,j = 1,\ldots,M, d = 1,\ldots D\}$. Utilizing this lower bound, we replace the general solution in (33) by

$$\mathrm{q}(\theta) = \text{const} \times \exp\left(\mathrm{E}_{\mathrm{q}(\Theta_\theta)}\left[\mathbf{B}(\Theta,\mathbf{w},\boldsymbol{\mu},\mathbf{y}_1,\mathbf{y}_2)\right]\right). \quad (37)$$

Applying this general solution (37) to each unknown results in an iterative procedure, which converges to the best approximation of the true posterior distribution $\mathrm{p}(\Theta|\mathbf{y}_1,\mathbf{y}_2)$ by distributions of the form in (25). This iterative procedure provides estimates $\mathrm{q}(\theta)$ to the distributions of the unknowns. In this work, we utilize the means of these distributions as the point estimates of the unknowns. Finally, note that in the case of $D = 1$, the solutions provided by (33) and (37) are equal.

V. Calculation of Posterior Distribution Approximations

In this section, we provide the explicit forms of each $\mathrm{q}(\cdot)$ distribution. In the following, the means of the distributions will be denoted by $\langle\cdot\rangle = \mathbf{E}_{\mathrm{q}(\theta)}[\cdot]$, when the corresponding distribution is clear from the context.

The distribution $\mathrm{q}(\mathbf{x})$ is calculated from (37) as a multivariate Gaussian distribution, that is

$$\mathrm{q}(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\langle\mathbf{x}\rangle,\Sigma_{\mathbf{x}}) \quad (38)$$

where its mean and covariance are given by

$$\langle\mathbf{x}\rangle = \Sigma_{\mathbf{x}}\left(\langle\beta_1\rangle\langle\mathbf{H}\rangle^T\mathbf{y}_1 + \langle\beta_2\rangle\mathbf{y}_2\right) \quad (39)$$

$$\Sigma_{\mathbf{x}}^{-1} = \langle\alpha_{\mathrm{im}}\rangle(\Delta^h)^T\mathbf{W}(\Delta^h) + \langle\alpha_{\mathrm{im}}\rangle(\Delta^v)^T\mathbf{W}(\Delta^v) + \langle\beta_1\rangle\langle\mathbf{H}^T\mathbf{H}\rangle + \langle\beta_2\rangle\mathbf{I} \quad (40)$$

with

$$w_j = \langle(\Delta_j^h(\mathbf{x}))^2 + (\Delta_j^v(\mathbf{x}))^2\rangle,\ j = 1,\ldots,N \quad (41)$$

$$\mathbf{W} = \mathrm{diag}\left(\frac{1}{\sqrt{w_j}}\right),\ j = 1,\ldots,N. \quad (42)$$

The mean $\langle\mathbf{x}\rangle$ of the distribution $\mathrm{q}(\mathbf{x})$ is used as the image estimate, which is calculated by applying a conjugate gradient method in (39). It can be seen from (40) that the matrix $\mathbf{W}$ in (42) is the spatial adaptivity matrix which controls the amount of smoothing at each pixel location depending on the intensity variation at that pixel, as expressed by the vector $\mathbf{w}$ representing the total variation of the estimated image. It therefore controls the trade-off between the data fidelity and image smoothness. Additionally, the parameters $\langle\beta_1\rangle$, $\langle\beta_2\rangle$ and $\langle\beta_{12}\rangle$ control the fidelity of the image estimate to the observed images $\mathbf{y}_1$ and $\mathbf{y}_2$. Since they are also estimated simultaneously with the image (as shown below), the estimation process is automatically steered towards the more reliable observation. The reliability is expressed by the constraints imposed on the image and blur estimates using their corresponding prior distributions. For instance, if the noise level in the short-exposed image is low, the estimation process relies more on $\mathbf{y}_2$ by increasing $\langle\beta_2\rangle$, as this provides smoother PSF and image estimates.

Next we find the distribution approximations $\mathrm{q}(h_j)$ of the blur PSF coefficients from (37) as rectified Gaussian distributions, given by (see the appendix for derivation details)

$$\mathrm{q}(h_j) = \mathcal{N}^R\left(h_j|\hat{h}_j,\tilde{h}_j\right) \quad (43)$$

with parameters shown in (44) and (45) at the bottom of the page, where $(\cdot)_{ij}$ denotes the $(i,j)^{th}$ element of a matrix, and $\mathbf{X}$ and $\mathbf{Y}_2$ are convolution matrices constructed from $\mathbf{x}$ and $\mathbf{y}_2$,

$$\begin{aligned}\log\mathbf{F}(\Theta,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2) &= \log\left[\mathrm{p}(\mathbf{h}|\{\tau_{jd}\},\{\sigma_{jd}\})\bar{\mathbf{F}}(\Theta,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2)\right] = \log\mathrm{p}(\mathbf{h}|\{\tau_{jd}\},\{\sigma_{jd}\}) + \log\bar{\mathbf{F}}(\Theta,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2)\\ &\geq \sum_{j=1}^{M}\sum_{d=1}^{D}\mu_{jd}\log\left(\frac{\tau_{jd}}{\mu_{jd}}\mathrm{Expon}\left(h_j\mid\sigma_{jd}\right)\right) + \log\bar{\mathbf{F}}(\Theta,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2) \quad (35)\\ &= \mathbf{B}(\Theta,\mathbf{w},\boldsymbol{\mu},\mathbf{y}_1,\mathbf{y}_2) \quad (36)\end{aligned}$$

$$\hat{h}_j = \left(\tilde{h}_j\right)^{-1}\left[-\sum_{d=1}^{D}\langle\sigma_{jd}\rangle\mu_{jd} + \sum_{n=1}^{N}\left\langle\beta_1 X_{nj}\left((y_1)_n - \sum_{\substack{m=1\\m\neq j}}^{M}X_{nm}h_m\right)\right\rangle + \sum_{n=1}^{N}\langle\beta_{12}\rangle(Y_2)_{nj}\left((y_1)_n - \sum_{\substack{m=1\\m\neq j}}^{M}(Y_2)_{nm}\langle h_m\rangle\right)\right] \quad (44)$$

$$\tilde{h}_j = \sum_{n=1}^{N}\langle\beta_1\rangle\langle X_{nj}^2\rangle + \langle\beta_{12}\rangle\sum_{n=1}^{N}(Y_2)_{nj}^2 \quad (45)$$

respectively. The mean $\langle h_j \rangle$ and variance $\mathrm{var}(h_j)$ of the distributions $\mathrm{q}(h_j)$ are given by [29]

$$\langle h_j \rangle = \hat{h}_j + \sqrt{\frac{2}{\pi \tilde{h}_j}} \frac{1}{\mathrm{erfcx}\left(-\hat{h}_j \sqrt{\frac{\tilde{h}_j}{2}}\right)} \tag{46}$$

$$\mathrm{var}(h_j) = \tilde{h}_j^{-1} + \sqrt{\frac{1}{\pi \tilde{h}_j}} \frac{\hat{h}_j}{\mathrm{erfcx}\left(-\hat{h}_j \sqrt{\frac{\tilde{h}_j}{2}}\right)} \tag{47}$$

where $\mathrm{erfcx}(\cdot)$ is the scaled complementary error function.

In the next step, we calculate the distributions of the hyperparameters from (37) as

$$\mathrm{q}(\alpha_{\mathrm{im}}) = \mathrm{Gamma}\left(\alpha_{\mathrm{im}} | \bar{a}_{\alpha_{\mathrm{im}}}, \bar{b}_{\alpha_{\mathrm{im}}}\right) \tag{48}$$

$$\mathrm{q}(\beta_1) = \mathrm{Gamma}\left(\beta_1 | \bar{a}_{\beta_1}, \bar{b}_{\beta_1}\right) \tag{49}$$

$$\mathrm{q}(\beta_2) = \mathrm{Gamma}\left(\beta_2 | \bar{a}_{\beta_2}, \bar{b}_{\beta_2}\right) \tag{50}$$

$$\mathrm{q}(\beta_{12}) = \mathrm{Gamma}\left(\beta_{12} | \bar{a}_{\beta_{12}}, \bar{b}_{\beta_{12}}\right) \tag{51}$$

$$\mathrm{q}(\sigma_{jd}) = \mathrm{Gamma}\left(\sigma_{jd} | \bar{a}_{\sigma_{jd}}, \bar{b}_{\sigma_{jd}}\right) \tag{52}$$

$$\mathrm{q}(\{\tau_{jd}\}_{d=1}^{D}) = \mathrm{Dirichlet}\left(\{\tau_{jd}\}_{d=1}^{D} | \{\bar{c}_{\tau_{jd}}\}_{d=1}^{D}\right). \tag{53}$$

Note that the posterior distribution approximations have the same shapes as their corresponding prior distributions due to the use of conjugate priors. We utilize the means of these distributions as their estimates, which are given by

$$\langle \alpha_{\mathrm{im}} \rangle = \frac{\bar{a}_{\alpha_{\mathrm{im}}}}{\bar{b}_{\alpha_{\mathrm{im}}}} = \frac{a^o_{\alpha_{\mathrm{im}}} + \frac{N}{2}}{b^o_{\alpha_{\mathrm{im}}} + \sum_j \sqrt{w_j}} \tag{54}$$

$$\langle \beta_1 \rangle = \frac{\bar{a}_{\beta_1}}{\bar{b}_{\beta_1}} = \frac{a^o_{\beta_1} + \frac{N}{3}}{b^o_{\beta_1} + \frac{1}{2}\langle \| \mathbf{y}_1 - \mathbf{H}\mathbf{x} \|^2 \rangle} \tag{55}$$

$$\langle \beta_2 \rangle = \frac{\bar{a}_{\beta_2}}{\bar{b}_{\beta_2}} = \frac{a^o_{\beta_2} + \frac{N}{3}}{b^o_{\beta_2} + \frac{1}{2}\langle \| \mathbf{y}_2 - \mathbf{x} \|^2 \rangle} \tag{56}$$

$$\langle \beta_{12} \rangle = \frac{\bar{a}_{\beta_{12}}}{\bar{b}_{\beta_{12}}} = \frac{a^o_{\beta_{12}} + \frac{N}{3}}{b^o_{\beta_{12}} + \frac{1}{2}\langle \| \mathbf{y}_1 - \mathbf{H}\mathbf{y}_2 \|^2 \rangle} \tag{57}$$

$$\langle \sigma_{jd} \rangle = \frac{\bar{a}_{\sigma_{jd}}}{\bar{b}_{\sigma_{jd}}} = \frac{a^o_{\sigma_{jd}} + \mu_{jd}}{b^o_{\sigma_{jd}} + \mu_{jd}\, h_j},\ j = 1, \ldots, M,\ d = 1, \ldots, D \tag{58}$$

$$\bar{c}_{\tau_{jd}} = c^o_{\tau_{jd}} + \mu_{jd} \tag{59}$$

$$\langle \tau_{jd} \rangle = \frac{\bar{c}_{\tau_{jd}}}{\sum_{d=1}^{D} \bar{c}_{\tau_{jd}}}. \tag{60}$$

The variances of these distributions can be used to assess the certainty of the algorithm about the estimated parameters. Finally, the auxiliary variables μ_{jd} are computed by first taking the expectation of (36) with respect to h_j and σ_{jd}, and then maximizing it with respect to the auxiliary variables. This results in the following update:

$$\mu_{jd} \propto \langle \tau_{jd} \rangle \, \mathrm{Expon}\left(\langle h_j \rangle | \langle \sigma_{jd} \rangle\right),\ j = 1, \ldots, M \tag{61}$$

with the condition

$$\sum_{d=1}^{D} \mu_{jd} = 1,\ j = 1, \ldots, M. \tag{62}$$

TABLE I
PROPOSED ALGORITHM

Set initial image estimate $<\mathbf{x}>^{(0)} = \mathbf{y}_1$
Calculate initial estimates of $<h_j>$, β_1, β_2, β_{12}, α_{im}, $\{\sigma_{jd}\}$ and $\{\tau_{jd}\}$ using $<\mathbf{x}>^{(0)}$, $\mathbf{y}_1$, and $\mathbf{y}_2$.
For $k = 1, 2, \ldots$ until convergence:
1) Find image distribution $\mathrm{q}^k(\mathbf{x})$ using (39)-(40)
2) Find blur PSF coefficient distributions $\mathrm{q}^k(h_j)$ using (44), (45) and (46).
3) Find hyperparameter estimates using (54)-(60)
4) Find auxiliary variables $\{\mu_{jd}\}$ using (61)

The proposed algorithm is summarized in Table I. The explicit forms of the expectations $\langle \mathbf{H}^T\mathbf{H} \rangle$ in (40), $\langle (\Delta_j^h(\mathbf{x}))^2 + (\Delta_j^v(\mathbf{x}))^2 \rangle$ in (41), and $\langle \| \mathbf{y}_1 - \mathbf{H}\mathbf{x} \|^2 \rangle$ in (55), $\langle \| \mathbf{y}_2 - \mathbf{x} \|^2 \rangle$ in (56), and $\langle \| \mathbf{y}_1 - \mathbf{H}\mathbf{y}_2 \|^2 \rangle$ in (57) are given by (63)–(67), shown at the bottom of the page, where

$$\Upsilon_{\mathbf{h}} = \mathrm{diag}\left(\sum_{j=1}^{M} \mathrm{var}(h_j)\right) \tag{68}$$

with $\mathrm{var}(h_j)$ given in (47).

Note that the explicit calculation of the covariance matrix $\Sigma_{\mathbf{x}}$ is only needed in (64)–(66), which is impractical due to its huge size of $N \times N$. As mentioned above, the image estimate in (39) is calculated using a conjugate gradient method, which does not require $\Sigma_{\mathbf{x}}$ to be constructed. In order to avoid the high computational complexity of computing $\Sigma_{\mathbf{x}}$ in (64)–(66), in these equations we approximate $\Sigma_{\mathbf{x}}$ as a diagonal matrix using the inverses of the diagonal elements of $\Sigma_{\mathbf{x}}^{-1}$ in (40). We have conducted extensive experiments with small images which permit the explicit construction of $\Sigma_{\mathbf{x}}$ to verify the validity of this approximation. We found out empirically that this approximation results in very similar estimates and has a minor effect in the estimation process. Note that similar approximations have also been utilized in other Bayesian recovery methods [34]–[36].

Finally we note the following. In the proposed framework, the distributions of the latent variables are estimated instead of their point estimates, which has several advantages. First, the uncertainty of the estimates can be calculated by examining the vari-

$$\langle \mathbf{H}^T\mathbf{H} \rangle = \langle \mathbf{H} \rangle^T \langle \mathbf{H} \rangle + \Upsilon_{\mathbf{h}} \tag{63}$$

$$\langle (\Delta_j^h(\mathbf{x}))^2 + (\Delta_j^v(\mathbf{x}))^2 \rangle = (\Delta_j^h(\langle \mathbf{x} \rangle))^2 + (\Delta_j^v(\langle \mathbf{x} \rangle))^2 + \mathrm{trace}\left(\Sigma_{\mathbf{x}}\left((\Delta_j^h)(\Delta_j^h)^T + (\Delta_j^v)(\Delta_j^v)^T\right)\right) \tag{64}$$

$$\begin{aligned} \langle \| \mathbf{y}_1 - \mathbf{H}\mathbf{x} \|^2 \rangle = {} & \| \mathbf{y}_1 - \langle \mathbf{H} \rangle \langle \mathbf{x} \rangle \|^2 + \mathrm{trace}\left(\langle \mathbf{H} \rangle^T \langle \mathbf{H} \rangle \Sigma_{\mathbf{x}}\right) \\ & + \mathrm{trace}\left(\Upsilon_{\mathbf{h}} \langle \mathbf{x} \rangle \langle \mathbf{x} \rangle^T\right) + \mathrm{trace}\left(\Sigma_{\mathbf{x}} \Upsilon_{\mathbf{h}}\right) \end{aligned} \tag{65}$$

$$\langle \| \mathbf{y}_2 - \mathbf{x} \|^2 \rangle = \| \mathbf{y}_2 - \langle \mathbf{x} \rangle \|^2 + \mathrm{trace}\left(\Sigma_{\mathbf{x}}\right) \tag{66}$$

$$\langle \| \mathbf{y}_1 - \mathbf{H}\mathbf{y}_2 \|^2 \rangle = \| \mathbf{y}_1 - \langle \mathbf{H} \rangle \mathbf{y}_2 \|^2 + \mathrm{trace}\left(\Upsilon_{\mathbf{h}} \mathbf{Y}_2 \mathbf{Y}_2^T\right) \tag{67}$$

ances of the estimated distributions. Second, these uncertainties are incorporated into the estimation procedure using the expectations given in (63)–(67), so that when estimating an unknown, the algorithm accounts for the possible errors in the estimates of other variables. This incorporation of uncertainties significantly improves the algorithm performance and it is also very useful in avoiding the high number of local minima resulting in low-quality estimates (see [37] for a related discussion). Note that if lower computational complexity is desired, degenerate distributions can be assumed for all unknowns, that is, all distributions are approximated by delta functions placed at their modes. This results in setting all covariances in the proposed method equal to zero, and it is equivalent to providing *maximum a posteriori* (MAP) estimates of the unknowns. Finally, note that instead of using the means of the distributions as the estimates of the unknowns, one can apply a sampling algorithm to draw different values from the distributions. Although this approach will lead to a much slower procedure, it can be used to avoid local minima in cases where initial estimates of the unknowns are far from the desired solutions and the degradations are extremely severe, and also to obtain alternative estimates when the uncertainty values provided by the variational analysis are large and the obtained image and blur are not satisfactory.

It should also be noted that the proposed framework can easily be extended to handle a higher number of input images with possibly more than one blurred image. The proposed framework provides the main mathematical basis for more general cases with multiple input images and multiple blur PSFs, and the resulting algorithms are very similar to the one presented in this work.

VI. Experimental Results

In this section, we demonstrate the performance of the proposed method with experiments with synthetic and real images. We first present experiments with synthetically generated degraded image pairs, to demonstrate the accuracy of the estimation of the unknown image and blur. We then show the application of the proposed method to real degraded image pairs and compare it with existing methods.

The following algorithm and experimental setup is utilized in all experiments. The observed image $\mathbf{y}_1$ is used as the initial estimate of $\mathbf{x}$. The initial estimate of $\mathbf{h}$ is obtained from (44) with the covariance matrix $\Sigma_{\mathbf{x}}$ set equal to zero. This initial estimate corresponds to the maximum likelihood estimate of $\mathbf{h}$, and it can be obtained efficiently in the frequency domain by taking the ratio of the Fourier transforms of the observations $\mathbf{y}_1$ and $\mathbf{y}_2$. Although the initial PSF calculated in this fashion is a very crude estimate of the true PSF, it provides a very fast initialization of the algorithm. The algorithm is able to provide very accurate results even with this crude initialization. The blur support M is chosen as the smallest support that covers the most significant entries of the initial PSF estimate. This operation can be performed manually or by simple thresholding followed by a convex hull algorithm. The blur support M chosen in this fashion is generally much smaller than the image support N, which improves the computational complexity of the method. Note that similar methods have been utilized by existing deconvolution methods, where the blur support is generally selected manually.

(a) (b)

Fig. 1. (a) Original Ephesus image, (b) observed noisy image simulating a short-exposure acquisition.

In all experiments, the number of mixture distributions is set to $D = 3$, but other values ($D = 2$ or $D = 4$) gave similar results. Utilizing single exponential distributions per pixel ($D = 1$) generally resulted in less sparse PSF estimates with higher estimation noise. All other parameters are calculated using (54)–(61). Note that except possibly the PSF support M, all required parameters of the algorithm are initialized automatically. As convergence criterion we use $\| \langle \mathbf{x} \rangle^{(k)} - \langle \mathbf{x} \rangle^{(k-1)} \|_2^2 \, / \, \| \langle \mathbf{x} \rangle^{(k-1)} \|_2^2 < 10^{-5}$, where $\langle \mathbf{x} \rangle^{(k)}$ and $\langle \mathbf{x} \rangle^{(k-1)}$ are the image estimates at iterations k and $k-1$, respectively. The convergence is generally achieved within 20 iterations, where each iteration takes approximately 20 seconds using our nonoptimized Matlab code running on a Pentium Core2 CPU at 2.66 GHz, depending on the severity of the degradations in the input images.

For the synthetic image degradations, the image shown in Fig. 1(a) is used to create the observed images $\mathbf{y}_1$ and $\mathbf{y}_2$. The range of the image is $[0, 255]$. The observed image $\mathbf{y}_2$, shown in Fig. 1(b), is obtained by adding white Gaussian noise of variance 220 to the original image ($\mathrm{SNR} = 7$ dB). This image suffers from a very high level of noise: the mean-squared-error (MSE) between this image and the original image is 196.20. We create five different observations $\mathbf{y}_1$ by blurring the original image by five different blur PSFs shown at the bottom row of Fig. 2, which are typical examples of PSFs resulting from the motion of the camera during long exposures. White Gaussian noise with a variance of 0.16 is added to the blurred images to obtain the final observed images $\mathbf{y}_1$ with signal-to-noise-ratios (SNR) of 40 dB, which are shown at the top row of Fig. 2. Note that although the noise level is low, as is typically the case in long-exposure images, the images are severely degraded by PSFs with large-supports compared to the image size. The MSEs between the original image and the observed images are given in Table II. The support of the PSFs used in this experiment are 21×21, and the original image is of size 430×270.

Each observed image $\mathbf{y}_1$ in Fig. 2 along with the observed image $\mathbf{y}_2$ in Fig. 1(b) is provided to the proposed algorithm as an image pair. The restored images obtained by the proposed method corresponding to different PSFs are shown at the top

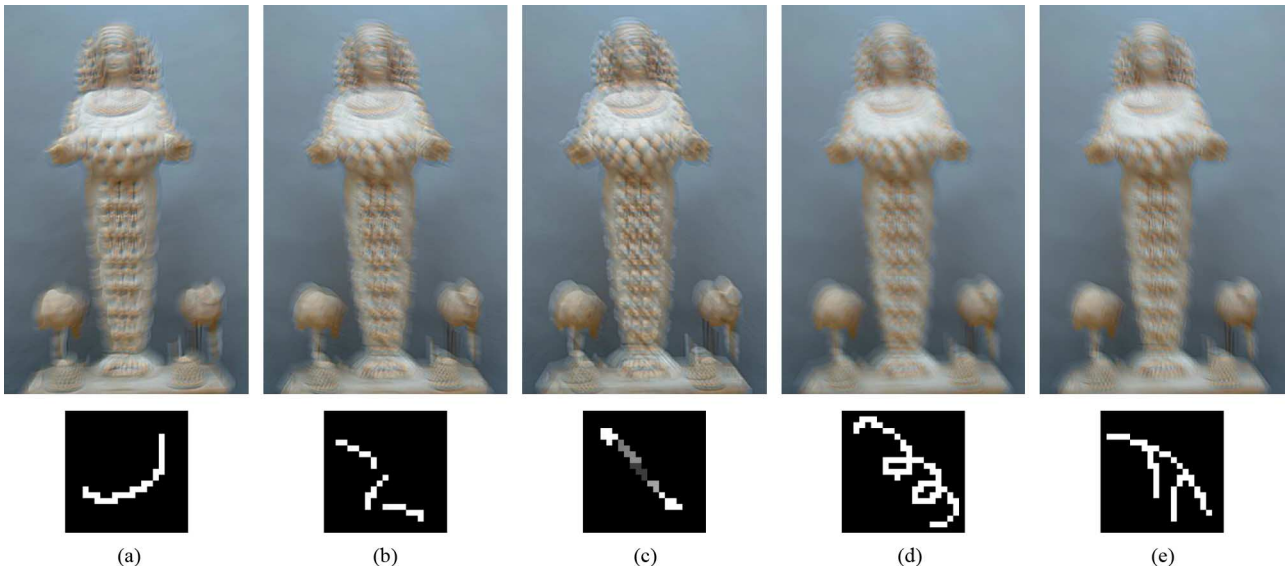


Fig. 2. Blurred images simulating long-exposure photographs. The point spread function (PSF) used to generate each image is shown below the corresponding image. All PSFs have support 21×21 pixels and the images are of size 430×270 pixels. The values of the PSFs are linearly mapped to the $[0, 255]$ range for visualization purposes.

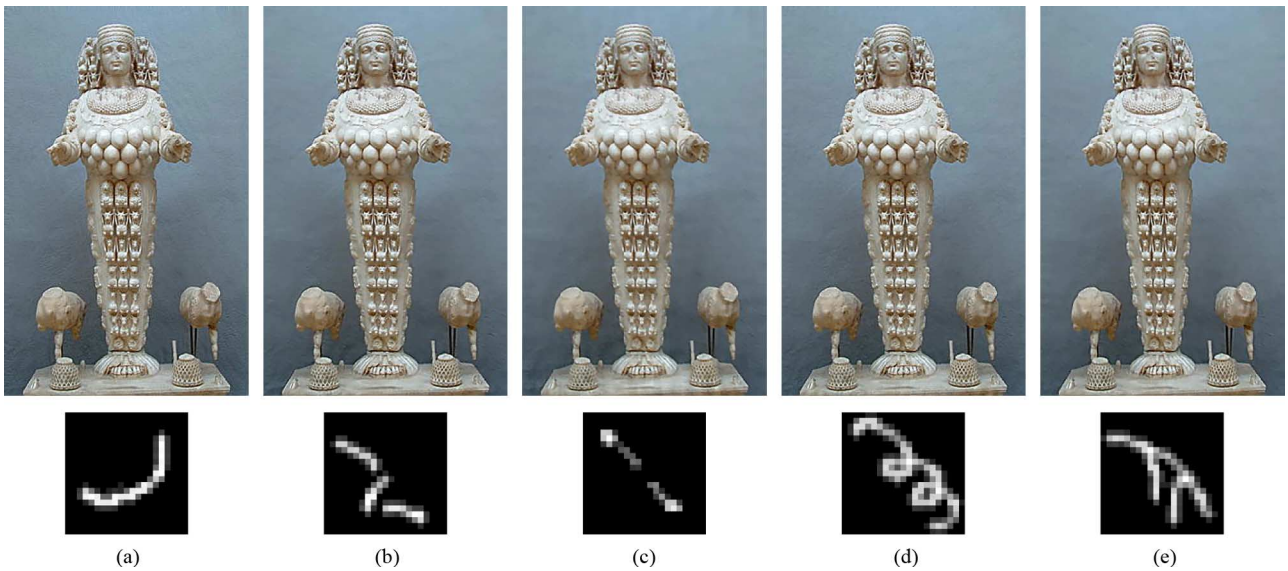
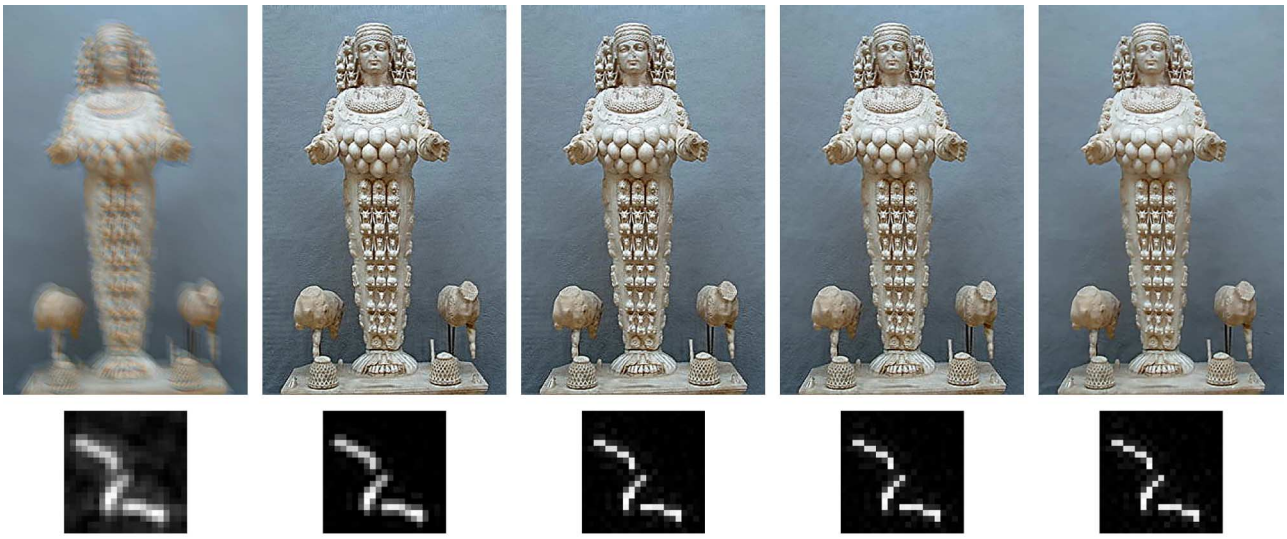


Fig. 3. Restoration results using the proposed algorithm. The restored images are shown in the top row, and the corresponding recovered PSFs are shown below the images. The values of the PSFs are linearly mapped to the $[0, 255]$ range for visualization purposes.

row of Fig. 3, and the recovered PSF for each case is shown at the bottom row of Fig. 3. By comparing with the original image in Fig. 1(a), it is clear that the proposed algorithm provides restored images with very high visual quality in all cases. The MSEs between the original image and the restored images are given in Table II. It is clear that the restored images are very close to the original image. The corresponding MSEs between the original and recovered PSFs are $9\times10^{-7}, 8\times10^{-6}, 5\times10^{-6}$, 9×10^{-7} and 3×10^{-6}. Both quantitative MSE results and visual inspection of the recovered PSFs suggest that the algorithm is very successful in estimating the original PSFs. Moreover, note that the recovered images do not exhibit any deconvolution artifacts such as ringing or noise amplification, due to the accurate estimation of the PSFs and the spatially-varying smoothing due to the total variation prior. The image and blur estimates during the iterative procedure in the proposed method for the blur shown in Fig. 2(b) are shown in Fig. 4. The estimated variances of the hyperparameters β_1, β_2 and β_{12} are provided in

TABLE II
MEAN-SQUARED ERRORS (MSES) FOR THE SYNTHETIC EXPERIMENTS

	Blur kernel #				
	1	2	3	4	5
Blurred Image	485.20	451.50	508.52	412.44	417.51
Noisy Image	196.65	196.65	196.65	196.65	196.65
deconvblind	135.21	182.87	198.27	247.43	224.28
Denoising [8]	44.33	44.33	44.33	44.33	44.33
Shan *et. al.* [4]	35.12	45.47	45.36	72.27	64.00
Proposed Method with $\beta_{12} = 0$	108.76	131.13	129.06	124.37	118.72
Proposed Method	11.29	24.5	42.60	18.12	21.39



(a) (b) (c) (d) (e)

Fig. 4. Restoration results during the iterative procedure for the case shown in Fig. 2(b). The restored images are shown in the top row, and the corresponding recovered PSFs are shown below the images. (a) Initial estimates, estimates at (b) iteration 3, (c) iteration 7, (d) iteration 15, and (e) iteration 20. The values of the PSFs are linearly mapped to the $[0, 255]$ range for visualization purposes.

(a) (b) (c) (d) (e)

Fig. 5. Restoration results using the *deconvblind* routine in MATLAB.

TABLE III
ESTIMATED VARIANCES OF THE HYPERPARAMETERS β_1, β_2, AND β_{12}

	Blur kernel #				
	1	2	3	4	5
$\text{var}(\beta_1)$	6.0×10^{-4}	1.6×10^{-4}	2.5×10^{-3}	7.5×10^{-4}	5.8×10^{-4}
$\text{var}(\beta_2)$	3.6×10^{-10}	3.9×10^{-10}	4.4×10^{-10}	4.6×10^{-10}	4.4×10^{-10}
$\text{var}(\beta_{12})$	3.3×10^{-7}	3.6×10^{-7}	2.4×10^{-7}	1.6×10^{-7}	9.9×10^{-7}

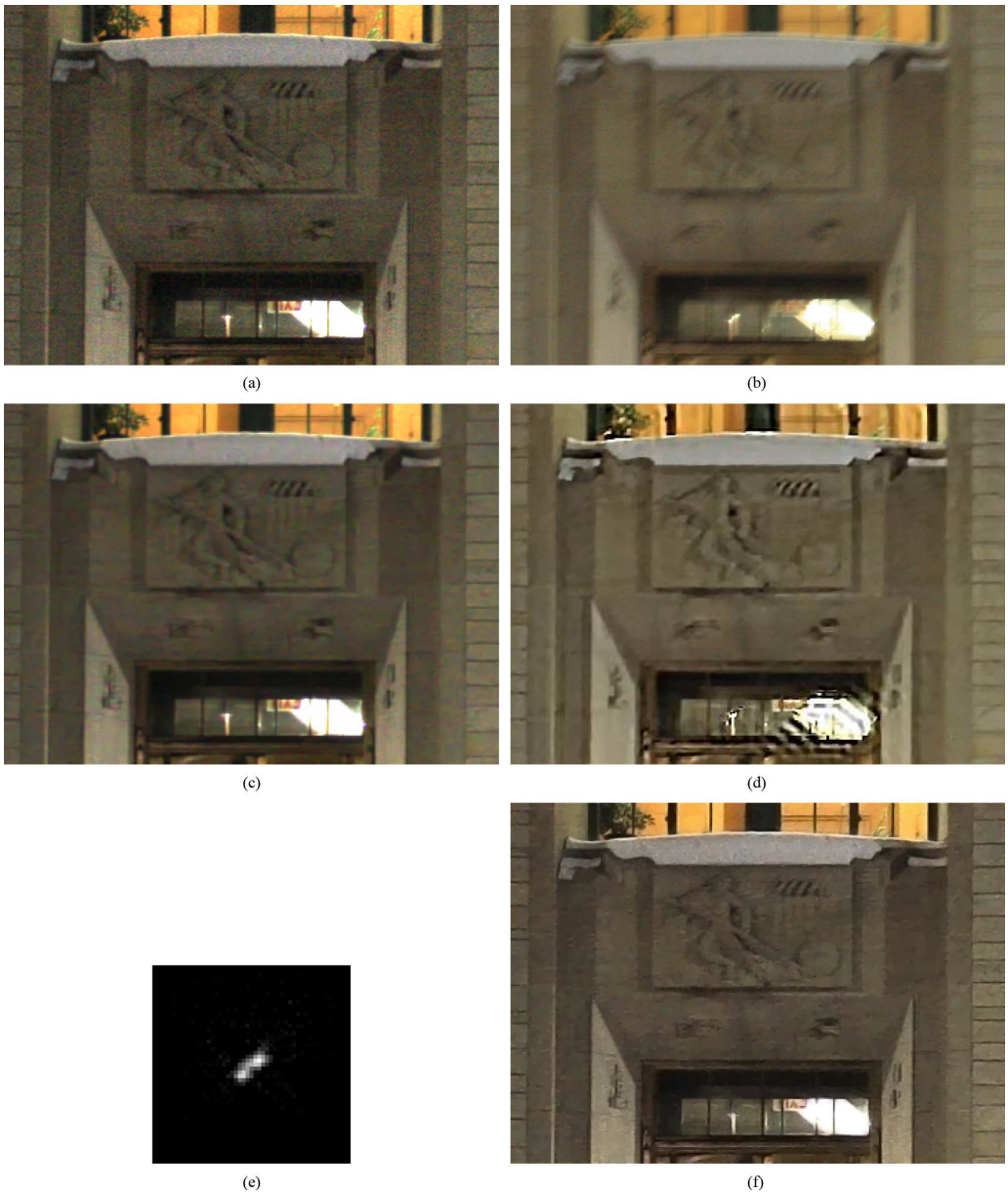
Table III, which, as mentioned above, can be used to evaluate the certainty of the algorithm about the estimated hyperparameters.

We also compare the results of the proposed algorithm with the classical single-image blind deconvolution algorithm implemented by the *deconvblind* routine in MATLAB, which utilizes a modified form of the Richardson-Lucy algorithm [19], [20]. To achieve the best possible restoration results we provided the algorithm with the PSF estimates obtained by the proposed method shown at the bottom row of Fig. 3, to be used as initial PSF estimates. Even with this unrealistic scenario, the quality of the resulting restored images, shown in Fig. 5, is much lower than that of the proposed method. The corresponding MSEs between the original and restored images are given in Table II. Note also that in most cases the blur is not completely removed, and significant ringing artifacts are present in the restored images. On the other hand, the proposed algorithm provides restored images of very high quality.

Finally, we compare the proposed algorithm with a state-of-the-art single image deblurring algorithm presented in [4]. Since this method is utilizing a single image, to provide a fair comparison, we provided the algorithm the original PSFs and applied it to the long-exposure image shown in Fig. 2. Note that this scenario is again not realistic, but presents the best possible results that can be obtained by [4]. The restored images are shown in Fig. 6 and the corresponding MSE values are given in Table II. Notice that the proposed method results in lower MSE values for all blur PSFs, and it can be observed by comparing Figs. 3 and 6 that it provides sharper images than [4]. The restored images by the proposed algorithm are also closer to the original

(a) (b) (c) (d) (e)

Fig. 6. Non-blind restoration results using the method in [4].



(a) (b) (c) (d) (e) (f)

Fig. 7. Outdoor image pair. (a) Short exposure image (brightness level is corrected), (b) long exposure image, (c) denoised short exposure image, (d) restored image using the algorithm [4], (e) recovered PSF using the proposed method (support: 51×51), and (f) restored image using the proposed algorithm.

image than the ones provided by [4], although the original PSF is provided to this method. Overall, these results show that the proposed method is more effective in preserving the image characteristics in the restoration process.

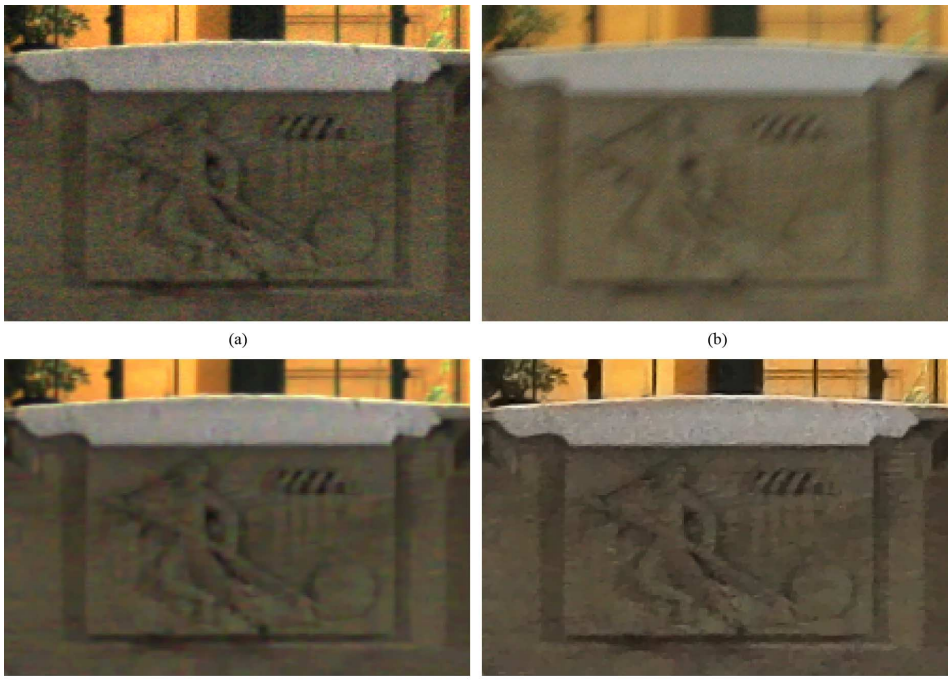


Fig. 8. Center regions of the images shown in Fig. 7. (a) Short exposure image, (b) long exposure image, (c) denoised short exposure image, (d) result of the proposed algorithm.

Next we apply the proposed method to a real image pair acquired by a compact digital camera. Figs. 7(a) and 7(b) show an image pair acquired outdoors at $\mathrm{ISO} = 800$ with exposure times 1/100 and 1/3 s, respectively. Due to low light conditions the noise level in the short exposure image is quite high. Moreover, it can be observed from Fig. 7(a) that the noise in the short exposure image is colored and therefore does not follow the assumed observation model in (4). In addition, certain parts of the images are highly saturated (e.g., part of the window above the door), which introduces an additional difficulty in blur estimation due to its nonlinearity. The application of the denoising algorithm in [8] to the short exposure image is shown in Fig. 7(c). We manually tuned the parameters of the denoising algorithm for each color channel, and show the result with the highest visual quality. The image and blur estimates provided by the proposed algorithm are shown in Fig. 7(e) and (f). The support of the PSF estimate is 51×51. The center regions of the images are shown in their original size in Fig. 8 for a closer inspection. It can be observed that despite the challenging nature of the input images, the algorithm provides significant improvement both in removing the blur and revealing sharp details in the image, as well as, in correcting the color loss apparent in the short-exposure image. Although the denoising method is also successful in removing the acquisition noise, it can not correct the color loss in the short exposure image, and its result is softer than the one provided by the proposed method. Finally, as an additional reference, we applied the algorithm in [4] to the long-exposure image and initialized the algorithm with the PSF estimate shown in Fig. 7(e). The corresponding restored image is shown in Fig. 7(d). Notice that this image suffers from a high level of ringing artifacts, especially around the saturated areas. Additionally, the blur is not completely removed and the edges are much softer than the image provided by the proposed method.

Finally, we present a comparison of the proposed method with the image stabilization method proposed in [17] and the denoising algorithm in [8] on a real image set. The images shown in Figs. 9(a) and 9(b) (published in [17]) are taken with exposure times 1/100 sec and 1 sec, respectively. The result of applying the denoising algorithm in [8] to the short exposure image is shown in Fig. 9(c). It is clear that although the noise level is significantly reduced, the contrast is very low, and there is a significant red color cast. The restored image in Fig. 9(d) is obtained by the algorithm in [17], which requires knowledge of the variances of the noise in the observations. Next, we applied the method [4] to the long exposure image where the PSF estimated by the proposed method (shown in Fig. 9(g)) is provided to this method. The restored image is shown in Fig. 9(e). Finally, the image and PSF estimates provided by the proposed algorithm are shown in Figs. 9(f) and 9(g), respectively, where the estimated support of the PSF is 41×41. Note that although the proposed method is fully-automated, the restored image is clearly sharper than that of [17] with almost no ringing artifacts. This is especially evident in the area around the letters. Moreover, the restored image by the proposed method is sharper than the denoised image and has a higher contrast with correctly restored colors. Finally, the image estimated by [4] exhibits a high level of ringing artifacts, and therefore the proposed method clearly provides an image with higher visual quality.

In summary, experimental results with both synthetic and real image sets demonstrate that the proposed algorithm is very effective in providing high quality restored images, although no image- and observation-specific parameter tuning has been performed.

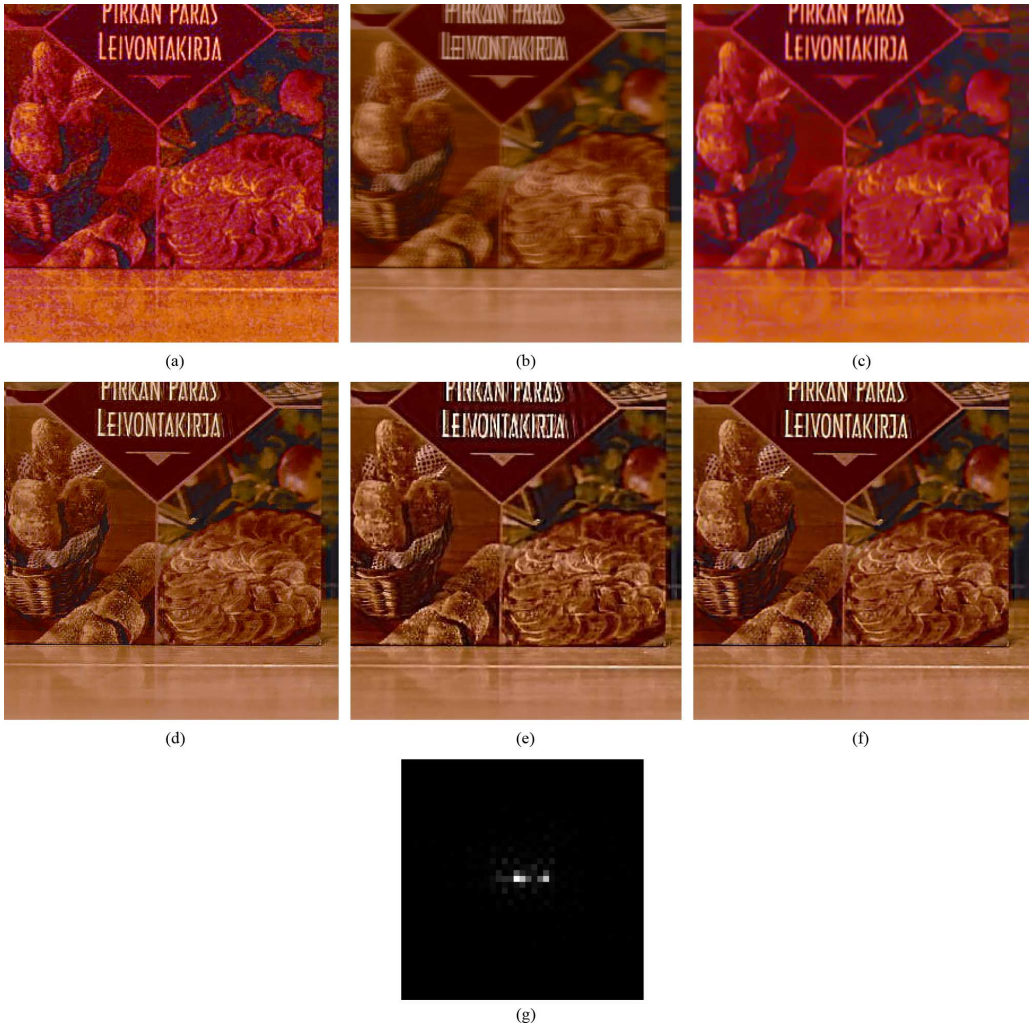


Fig. 9. Real image example (courtesy of [17]). (a) Short exposure image, (b) long exposure image, (c) denoised short exposure image, (d) restored image using [17], (e) restored image using the algorithm [4], (f) restored image using the proposed algorithm, and (g) recovered PSF using the proposed algorithm (support : 41×41).

VII. Conclusion

In this paper we presented a novel Bayesian formulation for blind deconvolution from image pairs acquired using long- and short-exposure times. The unknown image, blur and all model parameters, including the noise variances, are estimated solely from the observations without prior knowledge or user intervention. On the other hand, the proposed framework is very flexible so that when some prior knowledge about the unknowns is available, it can easily be incorporated into the algorithm. The developed algorithm simultaneously estimates the distributions of the unknowns which allows for the computation of the estimation uncertainties and also incorporates these uncertainties into the restoration procedure. The algorithm does not rely on nonrobust and input-dependent ad hoc methods (such as blur thresholding or blur denoising). Moreover, although the proposed method does not require user-intervention but instead provides a fully automated estimation of the algorithmic parameters, experimental results demonstrate that it results in very high quality restored images even with high degradations in both synthetic and real image cases, and compares favorably with existing methods.

Future work includes the incorporation of the geometric and photometric registration in the Bayesian framework. This will allow simultaneous estimation of the registration parameters along with the unknown image and the blur, which has the potential of providing more accurate registration estimates then preprocessing methods.

Appendix
Calculation of the Approximation to the Blur PSF Posterior Distribution

In this section we show the calculation of the distribution in (43) with parameters (44) and (45). Using (36) and (37), we have

$$\mathrm{q}(h_j) \propto \exp\left[-\frac{\langle\beta_1\rangle}{2}\langle\|\,\mathbf{y}_1-\mathbf{X}\mathbf{h}\|^2\rangle-\frac{\langle\beta_{12}\rangle}{2}\langle\|\,\mathbf{y}_1-\mathbf{Y}_2\mathbf{h}\|^2\rangle\right] \times\left[\sum_{d=1}^{D}\mu_{jd}\log\left(\frac{\langle\tau_{jd}\rangle}{\mu_{jd}}\langle\sigma_{jd}\rangle\exp\left(-\langle\sigma_{jd}\rangle h_j\right)\right)\right] \quad \text{(A-1)}$$

The first and second terms in the exponent can be written as shown in (A-2) and (A-3) shown at the top of the next page. Substituting these identities in (A-1) and ignoring the terms not

$$\langle \| \mathbf{y}_1 - \mathbf{X}\mathbf{h} \|^2 \rangle = \left\langle \sum_{n=1}^{N} \left((y_1)_n - \sum_{\substack{m=1 \\ m \neq j}}^{M} X_{nm} h_m - X_{nj} h_j \right)^2 \right\rangle \tag{A-2}$$

$$\langle \| \mathbf{y}_1 - \mathbf{Y}_2\mathbf{h} \|^2 \rangle = \left\langle \sum_{n=1}^{N} \left((y_1)_n - \sum_{\substack{m=1 \\ m \neq j}}^{M} (Y_2)_{nm} h_m - (Y_2)_{nj} h_j \right)^2 \right\rangle. \tag{A-3}$$

$$\begin{aligned} \mathrm{q}(h_j) \propto \exp \Bigg[& -\frac{\langle \beta_1 \rangle}{2} \sum_{n=1}^{N} \langle X_{nj}^2 \rangle h_j^2 + \langle \beta_1 \rangle \sum_{n=1}^{N} \left\langle X_{nj} \left((y_1)_n - \sum_{\substack{m=1 \\ m \neq j}}^{M} X_{nm} h_m \right) \right\rangle h_j - \frac{\langle \beta_{12} \rangle}{2} \\ & \sum_{n=1}^{N} (Y_2)_{nj}^2 h_j^2 + \langle \beta_{12} \rangle \sum_{n=1}^{N} (Y_2)_{nj} \left((y_1)_n - \sum_{\substack{m=1 \\ m \neq j}}^{M} (Y_2)_{nm} \langle h_m \rangle \right) h_j - \sum_{d=1}^{D} \langle \sigma_{jd} \rangle \mu_{jd} h_j \Bigg]. \end{aligned} \tag{A-4}$$

containing h_j, we obtain (A-4) shown at the top of the page. Note that (A-4) is in standard form of the rectified Gaussian distribution (see [29, App. A.3]. The explicit forms of the distribution parameters can be easily calculated from (A-4) to obtain (44) and (45). The mean and variance of this distribution given in (46) and (47) follow from the properties of the rectified Gaussian distribution.

ACKNOWLEDGMENT

The authors would like to thank M. Tico for providing the real image set and the results obtained by the algorithm in [17] shown in Fig. 9.

REFERENCES

[1] S. Babacan, J. Wang, R. Molina, and A. Katsaggelos, "Bayesian blind deconvolution from differently exposed image pairs," in *Proc. IEEE Int. Conf. Image Process.*, Cairo, Egypt, Jul. 2009, pp. 133–136.

[2] T. E. Bishop, S. D. Babacan, B. Amizic, A. K. Katsaggelos, T. Chan, and R. Molina, "Blind image deconvolution: Problem formulation and existing approaches," in *Blind Image Deconvolution: Theory and Applications*, P. Campisi and K. Egiazarian, Eds. Boca Raton, FL: CRC Press, 2007, ch. 1.

[3] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. Freeman, "Removing camera shake from a single photograph," in *Proc. SIGGRAPH Conf.* , Boston, MA, 2006, vol. 25, pp. 787–794.

[4] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," in *Proc. SIGGRAPH Conf.*, 2008.

[5] M. Ben-Ezra and S. Nayar, "Motion deblurring using hybrid imaging," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2003, vol. I, pp. 657–664.

[6] J. Jia, "Single image motion deblurring using transparency," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[7] N. Joshi, R. Szeliski, and D. Kriegman, "Psf estimation using sharp edge prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[8] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[9] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, vol. 2, pp. 60–65.

[10] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2866–2878, Oct. 2006.

[11] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, "Image deblurring with blurred/noisy image pairs," in *Proc. SIGGRAPH Conf.*, New York, 2007, p. 1.

[12] S. Mann, "Compositing multiple pictures of the same scene," in *Proc. 46th Annu. IS&T Conf.*, Cambridge, MA, May 1993, pp. 50–52.

[13] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. 24th Annu. Conf. Comput. Graph. Interact. Techniques*, New York, 1997, pp. 369–378.

[14] A. K. Katsaggelos, "A multiple input image restoration approach," *J. Vis. Commun. Image Represent.*, vol. 1, pp. 93–103, Sep. 1990.

[15] A. Rav-Acha and S. Peleg, "Two motion-blurred images are better than one," *Pattern Recognit. Lett.*, vol. 26, no. 3, pp. 311–317, 2005.

[16] F. 'roubek and J. Flusser, "Multichannel blind iterative image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 9, pp. 1094–1106, Sep. 2003.

[17] M. Tico and M. Vehvilainen, "Image stabilization based on fusing the visual information in differently exposed images," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 19, 2007, vol. 1, pp. 117–120.

[18] J. Chen, L. Yuan, C.-K. Tang, and L. Quan, "Robust dual motion deblurring," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[19] H. W. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Amer.*, vol. 62, no. 1, pp. 55–59, Jan. 1972.

[20] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astron. J.*, vol. 79, p. 745, Jun. 1974.

[21] P.-Y. Lu, T.-H. Huang, M.-S. Wu, Y.-T. Cheng, and Y.-Y. Chuang, "High dynamic range image reconstruction from hand-held cameras," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 509–516.

[22] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, "Blurred/non-blurred image alignment using sparseness prior," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.

[23] StereoPhoto Maker [Online]. Available: http://stereo.jpn.org/eng/stphmkr/

[24] G. Ward, "Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures," *J. Graph., GPU, Game Tools*, vol. 8, no. 2, pp. 17–30, 2003.

[25] D. Capel and A. Zisserman, "Computer vision applied to super resolution," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 75–86, May 2003.

[26] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman, "Bayesian methods for image super-resolution," *Comput. J.*, 2007.

[27] F. Sroubek and J. Flusser, "Multichannel blind deconvolution of spatially misaligned images," *IEEE Trans. Image Process.*, vol. 14, no. 7, pp. 874–883, Jul. 2005.

[28] F. Sroubek, G. Cristobal, and J. Flusser, "A unified approach to super-resolution and multichannel blind deconvolution," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2322–2332, Sep. 2007.

[29] J. Miskin, "Ensemble learning for independent component analysis," Ph.D. dissertation, Astrophys. Group, Univ. Cambridge, , 2000.

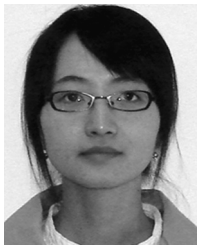
[30] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, pp. 259–268, 1992.

[31] S. D. Babacan, R. Molina, and A. Katsaggelos, "Parameter estimation in TV image restoration using variational distribution approximation," *IEEE Trans. Image Process.*, vol. 17, no. 3, pp. 326–339, Mar. 2008.
[32] R. Molina, J. Mateos, and A. K. Katsaggelos, "Blind deconvolution using a variational approach to parameter, image, and blur estimation," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3715–3727, Dec. 2006.
[33] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2006.
[34] M. E. Tipping and C. M. Bishop, "Bayesian image super-resolution," in *Advances in Neural Information Processing Systems 15 (NIPS)*. Cambridge, MA: MIT Press, 2003.
[35] S. Babacan, R. Molina, and A. Katsaggelos, "Variational Bayesian blind deconvolution using a total variation prior," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 12–26, Jan. 2009.
[36] A. Kanemura, S.-I. Maeda, and S. Ishii, "Superresolution with compound markov random fields via the variational EM algorithm," *Neural Netw.*, vol. 22, no. 7, pp. 1025–1034, 2009.
[37] A. Levin, Y. Weiss, F. Durand, and W. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1964–1971.

Sevket Derin Babacan (S'02–M'04) was born in Istanbul, Turkey, in 1981. He received the B.Sc. degree from Bogazici University, Istanbul, in 2004 and the M.Sc. and Ph.D. degrees from the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, in 2006 and 2009, respectively.

He is currently a Beckman Post-Doctoral Fellow with the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign. His primary research interests are inverse problems in image processing, computer vision, and computational photography.

Dr. Babacan was the recipient of an IEEE International Conference on Image Processing Paper Award in 2007.



Jingnan Wang received the B.S. degree from Zhejiang University, China and the M.Sc. degree from the Technische Universiteit Eindhoven, The Netherlands, and is currently pursuing the Ph.D. degree from the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL.

From 2007 to 2008, she worked as a research intern at Philips Research, Eindhoven, the Netherlands. In summer 2009, she interned with the Computational Video Group at Intuitive Surgical in California.

Ms. Wang was the recipient of the Philips Fellowship from the Technische Universiteit Eindhoven from 2006 to 2008 and the Philips Incentive Award from Philips Intellectual Property and Standards in 2009. She was also a recipient of a Murphy fellowship from Northwestern University in 2008 and 2009. Her research interests include variational Bayesian techniques for image restoration and image enhancement, motion analysis, shape analysis, and feature extraction.

Rafael Molina (M'88) was born in 1957. He received the degree in mathematics (statistics) in 1979 and the Ph.D. degree in optimal design in linear models in 1983.

He became Professor of computer science and artificial intelligence at the University of Granada, Granada, Spain, in 2000. His areas of research interest are image restoration (applications to astronomy and medicine), parameter estimation in image restoration, super resolution of images and video, and blind deconvolution.

Aggelos K. Katsaggelos (S'80–M'85–SM'92–F'98) received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Greece, in 1979 and the M.S. and Ph.D. degrees in electrical engineering from Georgia Institute of Technology, Atlanta, in 1981 and 1985, respectively.

In 1985, he joined the Electrical Engineering and Computer Science Department, Northwestern University, Evanston, IL, where he is currently a Professor. He was the holder of the Ameritech Chair of Information Technology from 1997 to 2003. He is also the Director of the Motorola Center for Seamless Communications, a member of the Academic Staff, the NorthShore University Health System, and an affiliated faculty member at the Department of Linguistics and the Argonne National Laboratory. He has published extensively, and he is the holder of 16 international patents. He is the coauthor of *Rate-Distortion Based Video Compression* (Kluwer, 1997), *Super-Resolution for Images and Video* (Claypool, 2007), and *Joint Source-Channel Video Transmission* (Claypool, 2007).

Dr. Katsaggelos was Editor-in-Chief of the *IEEE Signal Processing Magazine* from 1997 to 2002, a BOG Member of the IEEE Signal Processing Society from 1999 to 2001, and a member of the Publication Board of the *Proceedings of the IEEE* from 2003 to 2007, among his many professional activities. He is a Fellow of the SPIE (2009) and the recipient of the IEEE Third Millennium Medal (2000), the IEEE Signal Processing Society Meritorious Service Award (2001), an IEEE Signal Processing Society Best Paper Award (2001), an IEEE ICME Paper Award (2006), an IEEE ICIP Paper Award (2007), and an ISPA Paper Award (2009). He was a Distinguished Lecturer of the IEEE Signal Processing Society from 2007 to 2008.