COMBINING OBSERVATION MODELS IN DUAL EXPOSURE PROBLEMS USING THE KULLBACK-LEIBLER DIVERGENCE

M. Tallón[1], J. Mateos[1], S.D. Babacan[2], R. Molina[1], A.K. Katsaggelos[2]

[1]Departamento de Ciencias de la Computación e I.A.
Universidad de Granada, Granada, Spain
{mtallon, jmd, rms}@decsai.ugr.es

[2]EECS Department
Northwestern University, Evanston, IL USA
sdb@northwestern.edu, aggk@eecs.northwestern.edu

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. By taking a pair of blurred/noisy images it is possible to reconstruct a sharp image without noise. This paper is devoted to the combination of observation models in the blurred/noisy image pair reconstruction problem. By examining the difference between the blurred image and the blurred version of the noisy image a third observation model is obtained. Based on the minimization of a linear convex combination of Kullback-Leibler divergences between posterior distributions, a procedure to combine the three observation models is proposed in the paper. The estimated images are compared with images provided by other reconstruction methods.

1. INTRODUCTION

Taking high-quality photographs under low-light conditions is a challenging problem. A long exposure time can be set to capture as many photons from the scene as possible; but unfortunately the image becomes blurry without the use of a tripod. A short exposure shot produces a sharp very dark image, but if the ISO sensitivity is increased, the image becomes noisier. A computational photography approach to this problem consists of applying bracketing techniques and acquiring two almost simultaneous images of the above kind which are then combined using post-processing techniques.

To tackle the blurred/noisy restoration problem it was observed in [2], following the approach in [7] for multichannel image restoration, that it was possible to obtain an additional observation by examining the difference between the blurred image and the blurred version of the noisy image. In [2] inference was carried out assuming that the three observation models were independent.

The combination of observations models, like the combination of prior models, is a very challenging and interesting problem. However, while the combination of image prior has received some interest in the recent literature, see for instance [3] and references therein, no work has been reported on observation model combination.

In this paper we develop a method to combine the three above described observations based on the use of the Kullback-Leibler divergence between distributions. Taking into account that each combination of a given prior model and two of the above observation models produces a different posterior distribution of the underlying image, the use of variational posterior distribution approximation on each posterior produces as many posterior approximations as pairs of observation models can be formed. A unique approximation is obtained here by finding the distribution on the original image given the observations that minimizes a linear convex combination of the Kullback- Leibler divergences associated to each posterior distribution. We find this distribution in closed form.

The rest of this paper is organized as follows. In Sec. 2 we formulate the blurred/noisy image pair problem and mathematically introduce the third observation model. In sec. 3 the unknown variables are modeled using the hierarchical Bayesian framework and the posterior distribution associated to each pair of observation models is presented. Variational inference is used in Sec. 4 to find a unique posterior distribution approximation that takes into account the information provided by the three observation models. Finally, experimental results are presented in Sec. 5 and conclusions are drawn in Sec. 6.

This work was supported in part by the "Comisión Nacional de Ciencia y Tecnología" under contract TIC2007-65533

2. PROBLEM FORMULATION

Assuming that the blur is mainly caused by the shake of the camera during the long exposure time (a process which is modeled as a linear and space invariant operator), and that the two images are calibrated photometrically and geometrically in advance, the observation process can be written as

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

$$\mathbf{y}_2 = \mathbf{x} + \mathbf{n}_2, \tag{2}$$

where $\mathbf{y}_1$ and $\mathbf{y}_2$ are, respectively, the observed blurred and noisy images in the pair, $\mathbf{x}$ the unknown original image and $\mathbf{n}_1$ and $\mathbf{n}_2$ the shot noise assumed to be zero mean white Gaussian noise with variances β_1^{-1} and β_2^{-1}, respectively. Note that $\beta_1 >> \beta_2$ since the image $\mathbf{y}_1$ is mainly degraded by blur while $\mathbf{y}_2$ is degraded by a high amount of noise. We use matrix-vector notation throughout the paper, so that $\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{x}$, $\mathbf{n}_1$ 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 point spread function (PSF) $\mathbf{h}$ with support M, $M \leq N$.

Since both $\mathbf{y}_1$ and $\mathbf{y}_2$ are from the same scene, they are highly correlated. In [2] an additional degradation model that exploits the dependency between the observations $\mathbf{y}_1$ and $\mathbf{y}_2$, with a single unknown $\mathbf{H}$, is proposed. Combining (1) and (2) we obtain that given $\mathbf{H}$,

$$\mathbf{y}_1 - \mathbf{H}\mathbf{y}_2 = \mathbf{n}_{12}, \tag{3}$$

where $\mathbf{n}_{12} \sim \mathscr{N}(\mathbf{0}, \beta_1\mathbf{I} + \beta_2\mathbf{H}\mathbf{H}^t)$. We approximate here $|\beta_1\mathbf{I} + \beta_2\mathbf{H}\mathbf{H}^t|^{-1/2}$ by $\beta_{12}^{-N/2}$ where $\beta_{12} > 0$.

The objective of the blurred/noisy restoration problem is to obtain estimates of $\mathbf{x}$ and $\mathbf{h}$ utilizing $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{n}_{12}$ and prior knowledge about $\mathbf{x}$, $\mathbf{h}$, $\mathbf{n}_1$, $\mathbf{n}_2$.

3. HIERARCHICAL BAYESIAN MODEL

In this work, we adopt the hierarchical Bayesian framework which consists of two stages. In the first stage, we define prior distributions on the unknown image, $\mathbf{x}$, and blur, $\mathbf{h}$, and we propose two different probability distributions from the degradation models described in the previous sections. These distributions defined in the first stage depend on certain parameters, called *hyperparameters*, which are modeled by hyperprior distributions in the second stage. Let us now describe those probability distributions.

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 use a mixture prior of D exponential distributions on each coefficient of the PSF, that is,

$$\mathrm{p}(\mathbf{h}|\{\tau_d\},\{\sigma_d\}) = \prod_{j=1}^{M}\left(\sum_{d=1}^{D}\tau_d \,\mathrm{Expon}\left(h_j \mid \sigma_d\right)\right) = \prod_{j=1}^{M}\left(\sum_{d=1}^{D}\tau_d \,\sigma_d e^{-\sigma_d h_j}\right) \tag{4}$$

with τ_d the mixture coefficient for class d and σ_d the parameter of each exponential distribution. In addition to imposing sparsity, this prior also imposes positivity on the blur coefficients h_j. This mixture-of-exponentials prior has also been utilized before for modeling PSFs resulting from camera shake [4, 6] and in independent component analysis [6].

The total variation function is used as the prior model for the image because it preserves the edges in the image, not over-penalizing them, while imposing smoothness [1]. So, we can write the prior distribution of the image $\mathbf{x}$ as

$$\mathrm{p}(\mathbf{x}|\alpha) = c\,\alpha^{N/2}\exp\left[-\frac{1}{2}\alpha\sum_{i=1}^{N}\sqrt{(\Delta_i^h(\mathbf{x}))^2+(\Delta_i^v(\mathbf{x}))^2}\right], \tag{5}$$

where c is a constant and the operators $\Delta_i^h(\mathbf{x})$ and $\Delta_i^v(\mathbf{x})$ correspond to horizontal and vertical first order differences, at pixel i, respectively.

Based in the degradation models in (1), (2) and (3), we can define the following two observation models, one obtained from (1) and (2) as

$$\mathrm{p}_1(\mathbf{y}_1,\mathbf{y}_2|\mathbf{x},\mathbf{h},\beta_1,\beta_2) \propto \beta_1^{N/2}\beta_2^{N/2}\exp\left[-\frac{\beta_1}{2}\parallel \mathbf{y}_1-\mathbf{H}\mathbf{x}\parallel^2-\frac{\beta_2}{2}\parallel \mathbf{y}_2-\mathbf{x}\parallel^2\right], \tag{6}$$

and another one obtained from (2) and (3) as

$$\mathrm{p}_2(\mathbf{y}_1,\mathbf{y}_2|\mathbf{x},\mathbf{h},\beta_2,\beta_{12}) \propto \beta_2^{N/2}\beta_{12}^{N/2}\exp\left[-\frac{\beta_{12}}{2}\parallel \mathbf{y}_1-\mathbf{H}\mathbf{y}_2\parallel^2-\frac{\beta_2}{2}\parallel \mathbf{y}_2-\mathbf{x}\parallel^2\right]. \tag{7}$$

Although other observation models can also be defined, we will concentrate here on the two above models, the extension of the theory to be developed to alternative observation models is straightforward.

Note that in principle, we could have considered a single observation model combining the three quadratic forms, $\beta_1\parallel \mathbf{y}_1-\mathbf{H}\mathbf{x}\parallel^2$, $\beta_2\parallel \mathbf{y}_2-\mathbf{x}\parallel^2$, and $\beta_{12}\parallel \mathbf{y}_1-\mathbf{H}\mathbf{y}_2\parallel^2$, this would require the calculation of the partition function. However, proposing two different models will allow us to theoretically study how to perform the combination and, also, determine the best combination of both observation models, as we will make clear in the following sections.

The proposed prior and observation models depends on a set of parameters whose value is crucial in determining the performance of the algorithm. For their modeling we are going to employ Gamma distributions for the parameters α, β_1, β_2, β_{12} and σ_d, $d=1,\ldots,D$, that is,

$$\mathrm{p}(\omega) = \mathrm{Gamma}\left(\omega|a_\omega^o,b_\omega^o\right) \tag{8}$$

where ω denotes a hyperparameter and a_ω^o and b_ω^o are the shape and inverse scale parameters of the Gamma distribution. For the mixture parameters τ_d we use the Dirichlet distribution with parameters $c_{\tau_d}^o$, $d=1,\ldots,D$

$$\mathrm{p}(\{\tau_d\}_{d=1}^D) = \mathrm{Dirichlet}\left(\{\tau_d\}_{d=1}^D|\{c_{\tau_d}^o\}_{d=1}^D\right). \tag{9}$$

Finally, combining (5), (8) and (9) with the two observation models in (6) and (7) we obtain the joint distributions $\mathrm{p}_1(\cdot)$ and $\mathrm{p}_2(\cdot)$ given by

$$\begin{aligned}\mathrm{p}_i(\mathbf{y}_1,\mathbf{y}_2,\Omega,\boldsymbol{\beta}_i) &= \mathrm{p}(\mathbf{x}|\alpha)\mathrm{p}(\alpha)\\ &\times\ \mathrm{p}(\mathbf{h}|\{\tau_d\},\{\sigma_d\})\prod_{d=1}^{D}\mathrm{p}(\tau_d)\mathrm{p}(\sigma_d)\\ &\times\ \mathrm{p}_i(\mathbf{y}_1,\mathbf{y}_2|\mathbf{x},\mathbf{h},\boldsymbol{\beta}_i)\mathrm{p}(\boldsymbol{\beta}_i),\end{aligned} \tag{10}$$

for $i=1,2$, where $\Omega=\{\mathbf{x},\mathbf{h},\alpha,\{\tau_d\},\{\sigma_d\}\}$, $\boldsymbol{\beta}_1=\{\beta_1,\beta_2\}$ and $\boldsymbol{\beta}_2=\{\beta_2,\beta_{12}\}$.

4. 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 use the variational approach to approximate it. Let us denote by Θ the set of unknowns, i.e., $\Theta=\{\Omega,\boldsymbol{\beta}\}$ with $\boldsymbol{\beta}=\{\beta_1,\beta_2,\beta_{12}\}$. The goal is to calculate the posterior distribution, in our case is intractable so we approximate the posterior $\mathrm{p}(\Theta|\mathbf{y}_1,\mathbf{y}_2)$ by another distribution $\mathrm{q}(\Theta)$ which allows a tractable analysis. Here we propose to approximate this distribution by the distribution minimizing the following linear convex combination of Kullback-Leibler (KL) divergence measures

$$\hat{\mathrm{q}}(\Theta) = \underset{\mathrm{q}(\Theta)}{\mathrm{argmin}}\sum_{i=1}^{2}\lambda_i C_{KL}(\mathrm{q}(\Omega)\mathrm{q}(\boldsymbol{\beta}_i)\parallel \mathrm{p}_i(\Omega,\boldsymbol{\beta}_i|\mathbf{y}_1,\mathbf{y}_2)), \tag{11}$$

with $\lambda_i\geq 0$, $\lambda_1+\lambda_2=1$,

$$\begin{aligned}\mathrm{q}(\Omega) &= \mathrm{q}(\mathbf{x})\mathrm{q}(\mathbf{h})\mathrm{q}(\alpha)\mathrm{q}(\{\tau_d\}_{d=1}^D)\prod_{d=1}^{D}\mathrm{q}(\sigma_d),\\ \mathrm{q}(\boldsymbol{\beta}_1) &= \mathrm{q}(\beta_1)\mathrm{q}(\beta_2)\\ \mathrm{q}(\boldsymbol{\beta}_2) &= \mathrm{q}(\beta_2)\mathrm{q}(\beta_{12})\\ q(\mathbf{h}) &= \prod_{j=1}^{M}\mathrm{q}(h_j),\\ \mathrm{q}(\Theta) &= \mathrm{q}(\Omega)\mathrm{q}(\beta_1)\mathrm{q}(\beta_2)\mathrm{q}(\beta_{12}),\end{aligned}$$

and the KL divergence given by

$$C_{KL}(\mathrm{q}(\Omega)\mathrm{q}(\boldsymbol{\beta}_i)\parallel \mathrm{p}_i(\Omega,\boldsymbol{\beta}_i|\mathbf{y}_1,\mathbf{y}_2)) = \int \mathrm{q}(\Omega)\mathrm{q}(\boldsymbol{\beta}_i)\log\left(\frac{\mathrm{q}(\Omega)\mathrm{q}(\boldsymbol{\beta}_i)}{\mathrm{p}_i(\mathbf{y}_1,\mathbf{y}_2,\Omega,\boldsymbol{\beta}_i)}\right)d\Omega d\boldsymbol{\beta}_i+\mathrm{const.} \tag{12}$$

The estimation of λ_1 and λ_2 will not be addressed in this paper but we will show experimentally that a non-degenerate combination of divergences, that is, $\lambda_1,\lambda_2>0$, provides better results than a degenerate one. We want to note also that the model in [2] correspond to choose $\lambda_1=\lambda_2=1$.

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 (12) not possible. The problems caused by the TV prior can be avoided by utilizing a majorization-minimization approach, whose details are given in [1], which finds a bound for the distribution in (5) which makes the analytical derivation of the Bayesian inference tractable. Let us consider the functional $\mathbf{M}(\alpha,\mathbf{x},\mathbf{w})$, where $\mathbf{w}\in(R^+)^N$ is an $N-$dimensional vector with components w_i, $i=1,\ldots,N$,

$$\mathbf{M}(\alpha,\mathbf{x},\mathbf{w}) = c\,\alpha^{N/2}\exp\left[-\frac{\alpha}{2}\sum_{i=1}^{N}\frac{(\Delta_i^h(\mathbf{x}))^2+(\Delta_i^v(\mathbf{x}))^2+w_i}{\sqrt{w_i}}\right], \tag{13}$$

where c is the same constant as in (5). It can be shown (see [1] for the details) that the functional $\mathbf{M}(\alpha, \mathbf{x}, \mathbf{w})$ is a lower bound of the image prior $\mathrm{p}(\mathbf{x}|\alpha)$, that is,

$$\mathrm{p}(\mathbf{x}|\alpha) \geq \mathbf{M}(\alpha, \mathbf{x}, \mathbf{w}). \tag{14}$$

Using this lower bound a lower bound of the joint probability distributions in (10) can be found, that is,

$$\begin{aligned}
\mathrm{p}_i(\mathbf{y}_1, \mathbf{y}_2, \Omega, \boldsymbol{\beta}_i) &\geq \mathbf{M}(\alpha, \mathbf{x}, \mathbf{w})\mathrm{p}(\alpha) \\
&\times \mathrm{p}(\mathbf{h}|\{\tau_d\}, \{\sigma_d\}) \prod_{d=1}^{D} \mathrm{p}(\tau_d)\mathrm{p}(\sigma_d) \\
&\times \mathrm{p}_i(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}, \mathbf{h}, \boldsymbol{\beta}_i)\mathrm{p}(\boldsymbol{\beta}_i) \\
&= \mathbf{F}_i(\Omega, \boldsymbol{\beta}_i, \mathbf{w}, \mathbf{y}_1, \mathbf{y}_2),
\end{aligned} \tag{15}$$

for $i = 1, 2$, which leads to the following upper bound for the KL divergence in (12)

$$\begin{aligned}
&C_{KL}(\mathrm{q}(\Omega)\mathrm{q}(\boldsymbol{\beta}_i) \parallel \mathrm{p}_i(\Omega, \boldsymbol{\beta}_i|\mathbf{y}_1, \mathbf{y}_2)) \\
&\qquad \leq C_{KL}(q(\Omega)\mathrm{q}(\boldsymbol{\beta}_i) \parallel \mathbf{F}_i(\Omega, \boldsymbol{\beta}_i, \mathbf{w}, \mathbf{y}_1, \mathbf{y}_2)) + \text{const.}
\end{aligned} \tag{16}$$

An additional approximation of equation (4) is needed when using mixture priors. Specifically, we utilize Jensen's inequality as follows [6]

$$\begin{aligned}
\log(\mathrm{p}(\mathbf{h}|\{\tau_d\}, \{\sigma_d\})) &= \log \left[\prod_{j=1}^{M} \left(\sum_{d=1}^{D} \tau_d \,\mathrm{Expon}\left(h_j \mid \sigma_d\right) \right) \right] \\
&\geq \sum_{j=1}^{M} \sum_{d=1}^{D} \mu_{jd} \log \left(\frac{\tau_d}{\mu_{jd}} \mathrm{Expon}\left(h_j \mid \sigma_d\right) \right),
\end{aligned} \tag{17}$$

with $\mu_{jd} \geq 0$, $\sum_{d=1}^{D} \mu_{jd} = 1$, $\quad j = 1, \ldots, M$. An analysis of the closeness of this bound can be found in [6]. The auxiliary variables μ_{jd} need to be computed along with the unknowns Θ, as will be shown later.

Using (17), we obtain a lower bound of $\log \mathbf{F}_i(\Omega, \boldsymbol{\beta}_i, \mathbf{w}, \mathbf{y}_1, \mathbf{y}_2)$ as follows,

$$\begin{aligned}
&\log \mathbf{F}_i(\Omega, \boldsymbol{\beta}_i, \mathbf{w}, \mathbf{y}_1, \mathbf{y}_2) = \log \mathbf{M}(\alpha, \mathbf{x}, \mathbf{w}) + \log \mathrm{p}(\alpha) \\
&\quad + \sum_{d=1}^{D} \log \mathrm{p}(\tau_d)\mathrm{p}(\sigma_d) + \log \mathrm{p}(\mathbf{h}|\{\tau_d\}, \{\sigma_d\}) \\
&\quad + \log \mathrm{p}_i(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}, \mathbf{h}, \boldsymbol{\beta}_i)\mathrm{p}(\boldsymbol{\beta}_i) \\
&\quad \geq \log \mathbf{M}(\alpha, \mathbf{x}, \mathbf{w}) + \log \mathrm{p}(\alpha) \\
&\quad + \sum_{j=1}^{M} \sum_{d=1}^{D} \mu_{jd} \log \left(\frac{\tau_d}{\mu_{jd}} \mathrm{Expon}\left(h_j \mid \sigma_d\right) \right) \\
&\quad + \sum_{d=1}^{D} \log[\mathrm{p}(\tau_d)\mathrm{p}(\sigma_d)] + \log \mathrm{p}_i(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}, \mathbf{h}, \boldsymbol{\beta}_i)\mathrm{p}(\boldsymbol{\beta}_i) \\
&\quad = \mathbf{B}_i(\Omega, \boldsymbol{\beta}_i, \mathbf{w}, \boldsymbol{\mu}, \mathbf{y}_1, \mathbf{y}_2),
\end{aligned} \tag{18}$$

with $\boldsymbol{\mu} = \{\mu_{jd} \,|\, j = 1, \ldots, M, d = 1, \ldots D\}$.

Utilizing this lower bound, we obtain the solutions

$$\mathrm{q}(\beta_1) = \text{const} \times \exp(\langle \mathbf{B}_1(\Omega, \boldsymbol{\beta}_1, \mathbf{w}, \boldsymbol{\mu}, \mathbf{y}_1, \mathbf{y}_2) \rangle_\Omega) \tag{19}$$

$$\mathrm{q}(\beta_{12}) = \text{const} \times \exp(\langle \mathbf{B}_2(\Omega, \boldsymbol{\beta}_2, \mathbf{w}, \boldsymbol{\mu}, \mathbf{y}_1, \mathbf{y}_2) \rangle_\Omega) \tag{20}$$

where $\langle \cdot \rangle_\Omega = \mathbf{E}_{\mathrm{q}(\Omega)}[\cdot]$, and $\mathbf{E}_{\mathrm{q}(\Omega)}$ denotes the expectation with respect to the distribution $\mathrm{q}(\Omega)$. Furthermore, to calculate the rest of the distributions, $\mathrm{q}(\gamma)$, $\gamma \in \{\Omega, \beta_2\}$, we have to take into account both divergences, obtaining

$$\mathrm{q}(\gamma) = \text{const} \times \exp\left(\left\langle \sum_{i=1}^{2} \lambda_i \mathbf{B}_i(\Omega, \boldsymbol{\beta}_i, \mathbf{w}, \boldsymbol{\mu}, \mathbf{y}_1, \mathbf{y}_2) \right\rangle_{\Theta_\gamma} \right) \tag{21}$$

where Θ_γ denotes the set of unknown with γ removed.

Calculating the above distributions for 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 (11). In this work, we utilize the means of these distributions as the point estimates of the unknowns. Let us now to make explicit the form of each of these distributions.

The distribution $\mathrm{q}(\mathbf{x})$ is calculated from (21) as a multivariate Gaussian distribution, that is, $\mathrm{q}(\mathbf{x}) = \mathscr{N}(\mathbf{x}| \langle \mathbf{x} \rangle, \Sigma_\mathbf{x})$ where its mean and covariance are given by

$$\langle \mathbf{x} \rangle = \Sigma_\mathbf{x} \left(\lambda_1 \langle \beta_1 \rangle \langle \mathbf{H} \rangle^T \mathbf{y}_1 + \langle \beta_2 \rangle \mathbf{y}_2 \right) \tag{22}$$

$$\begin{aligned}
\Sigma_\mathbf{x}^{-1} &= \langle \alpha \rangle (\Delta^h)^T \mathbf{W}(\Delta^h) + \langle \alpha \rangle (\Delta^v)^T \mathbf{W}(\Delta^v) \\
&\quad + \lambda_1 \langle \beta_1 \rangle \left\langle \mathbf{H}^T \mathbf{H} \right\rangle + \langle \beta_2 \rangle \mathbf{I}
\end{aligned} \tag{23}$$

with

$$w_j = (\Delta_j^h(\langle \mathbf{x} \rangle))^2 + (\Delta_j^v(\langle \mathbf{x} \rangle))^2 \,, j = 1, \ldots, N, \tag{24}$$

$$\mathbf{W} = \mathrm{diag}\left(\frac{1}{\sqrt{w_j}} \right), \; j = 1, \ldots, N. \tag{25}$$

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 (22). It can be seen that the matrix $\mathbf{W}$ in (25) is an 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.

Next we find the distribution approximations $\mathrm{q}(h_j)$ of the blur PSF coefficients. From (21), $\mathrm{q}(h_j)$ are rectified Gaussian distributions, given by $\mathrm{q}(h_j) = \mathscr{N}^R \left(h_j | \hat{h}_j, \tilde{h}_j\right)$ with parameters

$$\begin{aligned}
\hat{h}_j = \left(\tilde{h}_j\right)^{-1} \Bigg[&- \sum_{d=1}^{D} \langle \sigma_d \rangle \mu_{jd} \\
&+ \lambda_1 \langle \beta_1 \rangle \sum_{n=1}^{N} \left\langle X_{nj} \left((y_1)_n - \sum_{\substack{m=1 \\ m \neq j}}^{M} X_{nm} h_m \right) \right\rangle \\
&+ (1 - \lambda_1) \langle \beta_{12} \rangle \sum_{n=1}^{N} (Y_2)_{nj} \left((y_1)_n - \sum_{\substack{m=1 \\ m \neq j}}^{M} (Y_2)_{nm} \langle h_m \rangle \right) \Bigg],
\end{aligned} \tag{26}$$

$$\tilde{h}_j = \lambda_1 \langle \beta_1 \rangle \sum_{n=1}^{N} \left\langle X_{nj}^2 \right\rangle + (1 - \lambda_1) \langle \beta_{12} \rangle \sum_{n=1}^{N} (Y_2)_{nj}^2, \tag{27}$$

where $\mathbf{X}$ and $\mathbf{Y}_2$ are convolution matrices constructed from $\mathbf{x}$ and $\mathbf{y}_2$, respectively, $(\cdot)_{ij}$ denotes the $(i, j)^{th}$ element of a matrix, and we used the fact that $\lambda_2 = 1 - \lambda_1$. The mean $\langle h_j \rangle$ of the distributions $\mathrm{q}(h_j)$, that is, our point estimate for h_j, is given by [6]

$$\langle h_j \rangle = \hat{h}_j + \sqrt{\frac{2}{\pi \tilde{h}_j}} \frac{1}{\mathrm{erfcx}(-\hat{h}_j \sqrt{\frac{\tilde{h}_j}{2}})}, \tag{28}$$

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

We then calculate the distributions of the hyperparameters $\omega \in \{\alpha, \beta_1, \beta_2, \beta_{12}, \sigma_d\}$ from (19), (20), and (21) as

$$\mathrm{q}(\omega) = \mathrm{Gamma}\left(\omega | \bar{a}_\omega, \bar{b}_\omega\right), \tag{29}$$

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

Figure 1: (a) original image, (b) observed noisy image simulating a short-exposure acquisition, (c) blurred image simulating long-exposure photographs. The blur used to generate this image is shown below the image. (d) using the method in [2], and (e) Estimated image and its corresponding blur using the proposed method with $\lambda_1 = 0.5$.

which produces

$$\mathrm{E}(\alpha_{\mathrm{im}}) = \frac{\bar{a}^{(\alpha_{\mathrm{im}})}}{\bar{b}^{(\alpha_{\mathrm{im}})}} = \frac{a^{(\alpha_{\mathrm{im}})} + \frac{N}{2}}{b^{(\alpha_{\mathrm{im}})} + \sum_j \sqrt{w_j}} \tag{30}$$

$$\mathrm{E}(\beta_1) = \frac{\bar{a}^{(\beta_1)}}{\bar{b}^{(\beta_1)}} = \frac{a^{(\beta_1)} + \frac{N}{2}}{b^{(\beta_1)} + \frac{1}{2}\langle \| \mathbf{y}_1 - \mathbf{H}\mathbf{x} \|^2 \rangle} \tag{31}$$

$$\mathrm{E}(\beta_2) = \frac{\bar{a}^{(\beta_2)}}{\bar{b}^{(\beta_2)}} = \frac{a^{(\beta_2)} + \frac{N}{2}}{b^{(\beta_2)} + \frac{1}{2}\langle \| \mathbf{y}_2 - \mathbf{x} \|^2 \rangle} \tag{32}$$

$$\mathrm{E}(\beta_{12}) = \frac{\bar{a}^{(\beta_{12})}}{\bar{b}^{(\beta_{12})}} = \frac{a^{(\beta_{12})} + \frac{N}{2}}{b^{(\beta_{12})} + \frac{1}{2}\langle \| \mathbf{y}_1 - \mathbf{Y}_2\mathbf{h} \|^2 \rangle} \tag{33}$$

$$\mathrm{E}(\sigma_d) = \frac{\bar{a}^{(\sigma_d)}}{\bar{b}^{(\sigma_d)}} = \frac{a^{(\sigma_d)} + \sum_j \mu_{jd}}{b^{(\sigma_d)} + \sum_j \mu_{jd}\, \langle h_j \rangle} \tag{34}$$

Furthermore

$$\mathrm{q}(\{\tau_d\}_{d=1}^D) = \mathrm{Dirichlet}\left(\{\tau_d\}_{d=1}^D | \{\bar{c}_{\tau_d}\}_{d=1}^D\right), \tag{35}$$

where $\bar{c}_{\tau_d} = c^o_{\tau_d} + \sum_j \mu_{jd}$ and so

$$\langle \tau_d \rangle = \frac{\bar{c}_{\tau_d}}{\sum_{d=1}^D \bar{c}_{\tau_d}}, \tag{36}$$

Finally, the auxiliary variables μ_{jd} are computed again from (21) as

$$\mu_{jd} \propto \langle \tau_d \rangle \, \mathrm{Expon}\left(\langle h_j \rangle \,|\, \langle \sigma_d \rangle\right), \ \ j = 1, \ldots, M \tag{37}$$

with the condition

$$\sum_{d=1}^D \mu_{jd} = 1, \quad j = 1, \ldots, M \tag{38}$$

To summarize, the proposed algorithm is written as follows:

Algorithm 1 *Estimation of the image* $\mathbf{x}$*, the blur* $\mathbf{h}$ *and the needed parameters*

1. *Set initial image estimate* $\langle \mathbf{x} \rangle^{(0)} = \mathbf{y}_1$
2. *Calculate initial estimates of* $\langle h_j \rangle$, β_1, β_2, β_{12}, α, $\{\sigma_{jd}\}$ *and* $\{\tau_d\}$ *using* $\langle \mathbf{x} \rangle^{(0)}$, $\mathbf{y}_1$, $\mathbf{y}_2$ *and* λ.
3. *For* $k = 1, 2, \ldots$ *until convergence:*
   (a) *Find image distribution* $\mathrm{q}^k(\mathbf{x})$ *using (22)-(23)*
   (b) *Find blur PSF coefficient distributions* $\mathrm{q}^k(h_j)$ *using (26)-(28)*
   (c) *Find hyperparameter estimates from the distributions (29) and (35)*
   (d) *Find auxiliary variables* $\{\mu_{jd}\}$ *using (37)*

5. EXPERIMENTAL RESULTS

We have tested the proposed algorithm with synthetic and real images. In a first experiment, synthetic images are used to test the accuracy of the estimation and numerically and visually demonstrate that the combination of observation models proposed provides better results than using only one model or assuming that the models are independent [2]. Then the proposed algorithm is applied to real degraded image pairs and its results compared to existing methods.

In all the experiment we set the initial values as follows: As described in Alg. 1, the initial estimation of $\mathbf{x}$, $\langle \mathbf{x} \rangle^0$, is set to observed image $\mathbf{y}_1$. We chose $D = 2$ which means that the blur will be comprised of elements of two classes, one for the elements close to zero and the other for the elements with higher value. The shape and inverse scale parameters of the Gamma distributions are set to a small common value (0.001), $c^o_{\tau_d}$ is set to 1 to obtain vague hyperpriors which make the estimation process rely more on the observations than on prior knowledge. The initial blur estimation is obtained as $\mathrm{argmin}_{\mathbf{h}} \|\mathbf{Y}_2\mathbf{h} - \mathbf{y}_1\|^2$ which can be efficiently computed in the Fourier domain. The blur support M is chosen as the smallest support that covers the most significant entries of the initial PSF estimate. The initial value of the parameters β_1, β_2, β_{12} and α are calculated from (29). Since we have two different classes, initial value for τ_d is set as the proportion of pixels of the initial blur estimator that are smaller or greater than half of the maximum value of $\mathbf{h}$, for $d = 1$ and $d = 2$, respectively, and σ_d to the inverse of the mean of the pixels in each class. Then $\{\mu_{jd}\}$ is estimated using (37). In the experiments we varied λ_1 from 0 to 1 with a step of 0.1 and run the algorithm until the convergence criterion $\left\|\langle \mathbf{x} \rangle^i - \langle \mathbf{x} \rangle^{i-1}\right\|^2 / \left\|\langle \mathbf{x} \rangle^{i-1}\right\|^2 < 10^{-5}$ is met. We want to note that we only restore the luminance of the color images. The resulting color images are composed by the restored luminance and the chrominance of the observed blurred image.

For the synthetic experiment we generated the images in the pair from the original image depicted in Fig. 1a. The image $\mathbf{y}_2$ was obtained by adding a zero-mean Gaussian noise of variance 700.8 to the luminance of the original image to obtain the noisy image in Fig. 1b with a SNR of 7dB. The blurred image, depicted in the top row of Fig. 1c, was obtained by convolving the original image with the blurring function in the bottom row of Fig. 1c and adding zero-

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

Figure 3: (a) Noisy and (b) blurred images in the pair. (c) Restored image and blur using the method in [2], and (d) restored image and blur using the proposed method with $\lambda_1 = 0.8$.

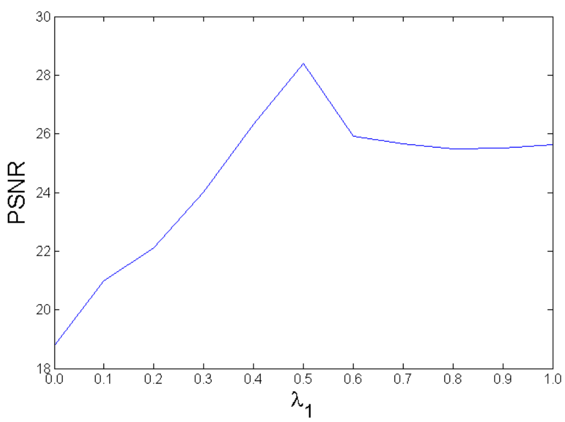

Figure 2: Obtained PSNR evolution as a function of λ_1 for the synthetic images pair.

mean Gaussian noise of variance 0.35 to obtain a SNR of 40dB. The pair formed by the blurred and the noisy images are the inputs of the algorithm.

To numerically evaluate the performance of the algorithm, the peak signal-to-noise ratio (PSNR) was used. Figure 2 shows the variation of the PSNR obtained with the proposed algorithm when λ_1 changes from 0.0 to 1.0, reaching its maximum value at $\lambda_1 = 0.5$ with a PSNR value of 28.4dB. The estimated image and blur obtained by the proposed algorithm with $\lambda_1 = 0.5$ are depicted in Figs. 1e. We want to note that the best result was obtained using a non-degenerate combination of the divergences, making clear that the combination of both models provides better results than using a single model (also note that $\lambda_1 = \lambda_2 = 0.5$ is not equivalent to the model in [2] which corresponds to setting $\lambda_1 = \lambda_2 = 1$). For comparison purposes we run the same experiments with the algorithm in [2] obtaining the image and blur shown in Fig. 1d with a PSNR of 27.9dB. The proposed algorithm provides better visual result than the algorithm in [2], with a better estimation of the blur, not as noisy, and sharper details in the image. Similar results were obtained when we synthetically blurred the original image with different PSFs. Usually a value for λ_1 between 0.4 and 0.7 resulted in the best PSNR.

We also tested our algorithm on a real image pair taken with a SLR camera with a fixed aperture of f/8. The noisy image was taken using an exposition time of 1/200 seconds and ISO 400 while the blurred one was taken using an exposition time of 1/3 seconds and ISO 200. The images were photometrically calibrated using histogram equalization and then geometrical calibration was carried out using SURF [5]+RANSAC to obtain the noisy and blurred images shown in Fig. 3a and 3b, respectively. Those images were the input of the proposed algorithm. The image obtained by the proposed algorithm using $\lambda_1 = 0.8$ is shown in Fig. 3d and the image obtained by the algorithm in [2] is shown in 3c. It is clear that, although both methods successfully removes a great part of the blur, the proposed method provides sharper details.

6. CONCLUSIONS

In this paper we have proposed a procedure based on variational Bayesian inference to combine observation models in the blurred/noisy image pair restoration problem. The procedure is based on finding the posterior distribution on the restored image given the observations that minimizes a linear convex combination of the Kullback-Leibler divergences associated to the prior and each pair of observation models. We have found this distribution in closed form. The estimated images compare favorably with images provided by other reconstruction methods. Future work will address the estimation of the weights assigned to each Kullback-Leibler divergence in the convex combination.

REFERENCES

[1] S. D. Babacan, R. Molina, and A. Katsaggelos. Parameter estimation in TV image restoration using variational distribution approximation. *IEEE Trans. Image Processing*, (3):326–339, March 2008.

[2] S. D. Babacan, J. Wang, R. Molina, and A. K. Katsaggelos. Bayesian blind deconvolution from differently exposed image pairs. *ICIP*, 2009.

[3] G. Chantas, N. Galatsanos, R. Molina, and A. Katsaggelos. Variational bayesian image restoration with a spatially adaptive product of total variation image priors. *IEEE Transactions on Image Processing*, 19(2):351–362, February 2010.

[4] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. Freeman. Removing camera shake from a single photograph. *ACM Transactions on Graphics, SIGGRAPH 2006 Conference Proceedings, Boston, MA*, 25:787–794, 2006.

[5] L. V. G. H Bay, T. Tuytelaars. Surf: Speeded up robust features. *ECCV*, 2006.

[6] J. Miskin. *Ensemble Learning for Independent Component Analysis*. PhD thesis, Astrophysics Group, University of Cambridge, 2000.

[7] F. Šroubek and J. Flusser. Multichannel blind deconvolution of spatially misaligned images. 7:45–53, July 2005.