PARAMETER ESTIMATION IN TOTAL VARIATION BLIND DECONVOLUTION

S. Derin Babacan[1], *Rafael Molina*[2], *Aggelos K. Katsaggelos*[1]

[1]Department of Electrical Engineering
and Computer Science
Northwestern University, Evanston, IL 60208, USA
sdb@northwestern.edu, aggk@eecs.northwestern.edu

[2]Departamento de Ciencias
de la Computación e I.A.
Universidad de Granada, 18071 Granada, Spain
rms@decsai.ugr.es

ABSTRACT

In this paper we present a methodology for parameter estimation in total variation (TV) blind deconvolution. By formulating the problem in a Bayesian framework, the unknown image, blur and the model parameters are simultaneously estimated. The resulting algorithms provide approximations to the posterior distributions of the unknowns by utilizing variational distribution approximations. We show that some of the current approaches towards TV-based blind deconvolution are special cases of our formulation. Experimental results are provided to demonstrate the performance of the algorithms.

1. INTRODUCTION

Blind deconvolution refers to a class of problems when the original unknown image is estimated from degraded observations with no information about the degradation and noise. It is encountered in many areas, such as astronomical imaging, photography, medical imaging, optics, and super-resolution applications, among others. Blind deconvolution is a very challenging problem because it is ill-posed, and the solution is not unique.

A standard formulation of the image degradation process is given by a linear space-invariant system, that is,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (1)$$

where $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{n}$ represent the original image, the observed image, and the noise, respectively, all in vector-form obtained by lexicographical ordering. The block-circulant matrix $\mathbf{H}$ represents the unknown blurring matrix which is formed by the point spread function (PSF) $\mathbf{h}$ of the degradation system. We assume that all vectors are of dimension $N \times 1$ and $\mathbf{H}$ of dimension $N \times N$. Note that the PSF support is to be equal to the image support in this formulation.

Following the recently proposed approaches in [1] and [2], which became very popular in the literature, the blind deconvolution problem can be formulated as a regularized least squares optimization as

$$(\mathbf{x}, \mathbf{h}) = \underset{\mathbf{x},\mathbf{h}}{\operatorname{argmin}} \frac{\beta}{2} \parallel \mathbf{y} - \mathbf{H}\mathbf{x} \parallel^2 + \alpha_{\text{im}} \text{TV}(\mathbf{x}) + \alpha_{\text{bl}} \text{TV}(\mathbf{h}), \quad (2)$$

where the total variation (TV) function is defined as

$$\text{TV}(\mathbf{x}) = \sum_i \sqrt{(\Delta_i^1(\mathbf{x}))^2 + (\Delta_i^2(\mathbf{x}))^2}, \quad (3)$$

with the operators $\Delta_i^1(\mathbf{x})$ and $\Delta_i^2(\mathbf{x})$ corresponding to the horizontal and vertical first order differences at pixel i, respectively. The TV function has the advantage of preserving the edge structure while imposing smoothness on the solutions.

The choice of the regularization parameters β, α_{im}, and α_{bl} is very critical in determining the convergence of the algorithms and the quality of the restored images. In general, it is hard to estimate the optimal parameters and a long supervised tuning process is needed. Some guidelines to choose them are presented in [1] and several methods are proposed for other blind deconvolution algorithms [2, 3, 4]. However, to our knowledge no work has been reported on the estimation of these parameters in TV-based blind deconvolution.

In this paper we propose a methodology for parameter estimation in TV-based blind deconvolution. By formulating the problem in Eq. (2) in the Bayesian framework, we model the unknowns $\mathbf{x}$ and $\mathbf{h}$, and the parameters β, α_{im}, and α_{bl} as stochastic quantities, and form their prior distributions. We utilize a variational approximation approach to estimate the posterior distribution and provide approximations to the posterior distributions from which estimates to the unknowns can be drawn. We will show that the TV-based algorithms in [1] and [2] are special cases of our framework. Our formulation results in two algorithms that are fully automated and require no user input.

The rest of this paper is organized as follows. In Section 2 the priors on the unknown in the Bayesian model are described. We present the variational methods for the inference in Section 3 and develop the proposed algorithms. Experimental results are shown in Section 4 and the paper is concluded in Section 5.

2. HIERARCHICAL BAYESIAN MODELING

In Bayesian models, all unknown parameters are treated as stochastic quantities, and probability distributions are assigned to them. The unknown parameters $\mathbf{x}$ and $\mathbf{h}$ are assigned *prior* distributions $\text{p}(\mathbf{x}|\alpha_{\text{im}})$ and $\text{p}(\mathbf{h}|\alpha_{\text{bl}})$, which model our knowledge about the nature of the original image and the blur, respectively. The observation $\mathbf{y}$ is also a random variable with the corresponding *conditional* distribution $\text{p}(\mathbf{y}|\mathbf{x}, \mathbf{h}, \beta)$. Clearly, these distributions depend on the model parameters α_{im}, α_{bl}, and β, which are called *hyperparameters*.

To alleviate the ill-posed nature of the blind deconvolution problem, prior knowledge about the unknown image and the blur is incorporated through the use of the prior distributions. In our case, when the hyperparameters are not assumed to be known, they have to be estimated simultaneously with the unknown parameters. To achieve this, we utilize a hierarchical model with two steps.

In the first stage of the Bayesian formulation, we model the observation process, the image, and the blur. Assuming the degradation noise is additive and independent zero-mean Gaussian, the probability distribution of the observation in Eq. (1) can be expressed as

$$\text{p}(\mathbf{y}|\mathbf{x}, h, \beta) \propto \beta^{N/2} \exp\left[-\frac{\beta}{2} \parallel \mathbf{y} - \mathbf{H}\mathbf{x} \parallel^2\right], \quad (4)$$

where β^{-1} is the noise variance. We adopt the TV function for the image prior, that is,

$$\text{p}(\mathbf{x}|\alpha_{\text{im}}) \propto \frac{1}{Z_1(\alpha_{\text{im}})} \exp\left[-\alpha_{\text{im}} \text{TV}(\mathbf{x})\right], \quad (5)$$

This work has been partially supported by the Spanish research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018) and the Ministerio de Educacion y Ciencia under contract TIN2007-65533.

with the partition function $Z_1(\alpha_{\text{im}})$.

Simultaneous Autoregression (SAR), Conditional Autoregression (CAR), and Gaussian models are some of the commonly used models for the unknown blur [4, 5, 6]. These models are very efficient in estimating smooth PSFs, for instance, a Gaussian PSF modeling long-term atmospheric turbulence. However, for PSFs with sharp transitions, such as motion blur and out-of-focus blur, smooth priors cannot capture the sharp transitions and tend to smooth the estimated PSF. For such PSFs with discontinuities, the TV function is more suitable since it does not overpenalize large gradients. We define the prior on the blur as

$$\mathrm{p}(\mathbf{h}|\alpha_{\text{bl}}) \propto \frac{1}{Z_2(\alpha_{\text{bl}})} \exp\left[-\alpha_{\text{bl}} \mathrm{TV}(\mathbf{h})\right], \tag{6}$$

with $Z_2(\alpha_{\text{bl}})$ the partition function.

Calculating closed form solutions for $Z_1(\alpha_{\text{im}})$ and $Z_2(\alpha_{\text{bl}})$ presents a major difficulty. Instead, quadratic approximations are used for these partition functions (see [5]) so that the image and blur priors can be expressed as

$$\begin{aligned} \mathrm{p}(\mathbf{x}|\alpha_{\text{im}}) &\propto \alpha_{\text{im}}^{N/2} \exp\left[-\alpha_{\text{im}} \mathrm{TV}(\mathbf{x})\right], &(7)\\ \mathrm{p}(\mathbf{h}|\alpha_{\text{bl}}) &\propto \alpha_{\text{bl}}^{M/2} \exp\left[-\alpha_{\text{bl}} \mathrm{TV}(\mathbf{h})\right]. &(8)\end{aligned}$$

In the second stage of the hierarchical model, we model each hyperparameter with a corresponding hyperprior. We utilize flat improper hyperpriors on all hyperparameters, that is,

$$\mathrm{p}(\alpha_{\text{im}}) \propto \text{const}, \qquad \mathrm{p}(\alpha_{\text{bl}}) \propto \text{const}, \qquad \mathrm{p}(\beta) \propto \text{const}. \tag{9}$$

Finally, combining the first and second stages, we form the joint probability distribution as follows

$$\begin{aligned} &\mathrm{p}(\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta, \mathbf{x}, \mathbf{h}, \mathbf{y}) \\ &= \mathrm{p}(\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta)\mathrm{p}(\mathbf{x}|\alpha_{\text{im}})\mathrm{p}(\mathbf{h}|\alpha_{\text{bl}})\mathrm{p}(\mathbf{y}|\mathbf{x}, \mathbf{h}, \beta). \end{aligned} \tag{10}$$

3. VARIATIONAL INFERENCE

Let us denote the set of hyperparameters by $\Omega = (\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta)$, and the set of all unknowns by $\Theta = (\mathbf{x}, \mathbf{h}, \alpha_{\text{im}}, \alpha_{\text{bl}}, \beta)$. The Bayesian inference is based on the posterior distribution given by

$$\mathrm{p}(\Theta|\mathbf{y}) = \frac{\mathrm{p}(\Omega)\mathrm{p}(\mathbf{x}|\alpha_{\text{im}})\mathrm{p}(\mathbf{h}|\alpha_{\text{bl}})\mathrm{p}(\mathbf{y}|\mathbf{x}, h, \beta)}{\mathrm{p}(\mathbf{y})}. \tag{11}$$

The closed form of the posterior distribution cannot be calculated because $\mathrm{p}(\mathbf{y})$ cannot be found analytically. Therefore, we consider instead a simpler parametric approximation $\mathrm{q}(\Theta)$ to the posterior $\mathrm{p}(\Theta|\mathbf{y})$ which is found by minimizing the Kullback-Leibner (KL) divergence [7], given by

$$\begin{aligned} C_{KL}(\mathrm{q}(\Theta) \parallel \mathrm{p}(\Theta|\mathbf{y})) &= \int \mathrm{q}(\Theta) \log\left(\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta|\mathbf{y})}\right) d\Theta \\ &= \int \mathrm{q}(\Theta) \log\left(\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta, \mathbf{y})}\right) d\Theta + \text{const}. \end{aligned} \tag{12}$$

In order to obtain a tractable approximation, the family of distributions $\mathrm{q}(\Theta)$ are restricted by utilizing the mean field approximation so that $\mathrm{q}(\Theta) = \mathrm{q}(\alpha_{\text{im}})\mathrm{q}(\alpha_{\text{bl}})\mathrm{q}(\beta)\mathrm{q}(\mathbf{x})\mathrm{q}(\mathbf{h})$. The major difficulty with the TV function is that it makes the KL distance difficult to evaluate. To overcome this problem, we define two functionals $\mathrm{M}(\alpha_{\text{im}}, \mathbf{x}, \mathbf{w})$ and $\mathrm{T}(\alpha_{\text{bl}}, \mathbf{h}, \mathbf{u})$ for any $N-$dimensional vector $\mathbf{w} \in (R^+)^N$ and $N-$dimensional vector $\mathbf{u} \in (R^+)^N$, according to

$$\begin{aligned} &\mathrm{M}(\alpha_{\text{im}}, \mathbf{x}, \mathbf{w}) \\ &= c_1 \alpha_{\text{im}}^{N/2} \exp\left[-\frac{\alpha_{\text{im}}}{2} \sum_i \frac{(\Delta_i^1(\mathbf{x}))^2 + (\Delta_i^2(\mathbf{x}))^2 + w_i}{\sqrt{w_i}}\right], &(13)\\ &\mathrm{T}(\alpha_{\text{bl}}, \mathbf{h}, \mathbf{u}) \\ &= c_2 \alpha_{\text{bl}}^{N/2} \exp\left[-\frac{\alpha_{\text{bl}}}{2} \sum_i \frac{(\Delta_i^1(\mathbf{h}))^2 + (\Delta_i^2(\mathbf{h}))^2 + u_i}{\sqrt{u_i}}\right], &(14)\end{aligned}$$

with c_1 and c_2 constants. By considering the following inequality, which states that for any $a \geq 0$ and $b > 0$

$$\sqrt{ab} \leq \frac{a+b}{2} \Rightarrow \sqrt{a} \leq \frac{a+b}{2\sqrt{b}}. \tag{15}$$

and comparing Eqs. (13) and (14) with Eqs. (7) and (8), respectively, we obtain the following lower bounds for the priors

$$\begin{aligned} \mathrm{p}(\mathbf{x}|\alpha_{\text{im}}) &\geq \mathrm{M}(\alpha_{\text{im}}, \mathbf{x}, \mathbf{w}), &(16)\\ \mathrm{p}(\mathbf{h}|\alpha_{\text{bl}}) &\geq \mathrm{T}(\alpha_{\text{bl}}, \mathbf{h}, \mathbf{u}). &(17)\end{aligned}$$

Therefore, we have the following lower bound for the joint probability distribution from Eq. (10)

$$\begin{aligned} \mathrm{p}(\Theta, \mathbf{y}) &\geq \mathrm{p}(\Omega)\mathrm{M}(\alpha_{\text{im}}, \mathbf{x}, \mathbf{w})\mathrm{T}(\alpha_{\text{bl}}, \mathbf{h}, \mathbf{u})\mathrm{p}(\mathbf{y}|\mathbf{x}, \mathbf{h}, \beta) \\ &= \mathrm{F}(\Theta, \mathbf{w}, \mathbf{u}, \mathbf{y}). \end{aligned} \tag{18}$$

For $\theta \in \{\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta, \mathbf{x}, \mathbf{h}\}$ let us denote by Θ_θ the subset of Θ with θ removed; for instance, if $\theta = \mathbf{x}$, $\Theta_{\mathbf{x}} = (\alpha_{\text{im}}, \alpha_{\text{bl}}, \beta, \mathbf{h})$. Then, using the lower bound in (18) we can define an upper bound for the Kullback-Leibner distance given by

$$\begin{aligned} &C_{KL}(\mathrm{q}(\Theta) \parallel \mathrm{p}(\Theta|\mathbf{y})) \leq C_{KL}(\mathrm{q}(\Theta) \parallel \mathrm{F}(\Theta, \mathbf{w}, \mathbf{u}, \mathbf{y})) \\ &= \int \mathrm{q}(\theta) \left(\int \mathrm{q}(\Theta_\theta) \log\left(\frac{\mathrm{q}(\theta)\mathrm{q}(\Theta_\theta)}{\mathrm{F}(\Theta, \mathbf{w}, \mathbf{u}, \mathbf{y})}\right) d\Theta_\theta\right) d\theta. \end{aligned} \tag{19}$$

Note that the minimization of the KL distance which is very difficult to evaluate because of the TV functions, can be replaced by minimization of this upper bound. We utilize an alternating minimization (AM) procedure [8] to find estimates of the posterior distributions. For each unknown θ, the posterior $\mathrm{q}(\theta)$ can be computed by holding $\mathrm{q}(\Theta_\theta)$ constant and solving

$$\mathrm{q}(\theta) = \arg\min_{\mathrm{q}(\theta)} C_{KL}(\mathrm{q}(\Theta_\theta)\mathrm{q}(\theta) \parallel \mathrm{F}(\Theta, \mathbf{w}, \mathbf{u}, \mathbf{y})). \tag{20}$$

The solution is found by differentiating the integral on the right hand side with respect to $\mathrm{q}(\theta)$ and setting it equal to zero, which results in (see Eq. (2.28) in [9]),

$$\hat{\mathrm{q}}(\theta) = \text{const} \times \exp\left(\mathrm{E}_{\mathrm{q}(\Theta_\theta)}\left[\, \log \mathrm{F}(\Theta, \mathbf{w}, \mathbf{u}, \mathbf{y}) \,\right]\right), \tag{21}$$

where

$$\mathrm{E}_{\mathrm{q}(\Theta_\theta)}\left[\, \log \mathrm{F}(\Theta, \mathbf{w}, \mathbf{u}, \mathbf{y}) \,\right] = \int \log \mathrm{F}(\Theta, \mathbf{w}, \mathbf{u}, \mathbf{y}) \mathrm{q}(\Theta_\theta) d\Theta_\theta.$$

Applying this general solution to each unknown in an alternating fashion results in the iterative procedure shown above.

We proceed by stating the solutions at each step of the algorithm. It is not difficult to see from step 3 that $\mathrm{q}^k(\mathbf{x})$ is an N-dimensional Gaussian distribution, rewritten as,

$$\mathrm{q}^k(\mathbf{x}) = \mathcal{N}\left(\mathbf{x} \mid \mathrm{E}^k(\mathbf{x}), \mathrm{cov}^k(\mathbf{x})\right).$$

The covariance and mean of this normal distribution can be calculated as

$$\begin{aligned} \mathrm{cov}^k(\mathbf{x}) = \big[&\beta^k \mathbf{H}^t \mathbf{H} + N\beta^k \mathrm{cov}^k(\mathbf{h}) \\ &+ \alpha_{\text{im}}^k (\Delta^1)^t W(\mathbf{w}^k)(\Delta^1) + \alpha_{\text{im}}^k (\Delta^2)^t W(\mathbf{w}^k)(\Delta^2)\big]^{-1}, \end{aligned} \tag{22}$$

Algorithm 1

1: Given observation $\mathbf{y}$, and initial estimates $\mathrm{q}^1(\mathbf{h})$, Ω^1
2: **while** Convergence criterion is not met **do**
3: Find

$$\mathrm{q}^k(\mathbf{x}) \propto \exp\left(\mathrm{E}_{\mathrm{q}(\Theta_{\mathbf{x}})}\left[\log \mathrm{F}(\mathbf{x},\mathbf{h}^k,\Omega^k,\mathbf{w}^k,\mathbf{u}^k,\mathbf{y})\right]\right)$$

4: Find

$$\mathrm{q}^{k+1}(\mathbf{h}) \propto \exp\left(\mathrm{E}_{\mathrm{q}(\Theta_{\mathbf{h}})}\left[\log \mathrm{F}(\mathbf{x}^k,\mathbf{h},\Omega^k,\mathbf{w}^k,\mathbf{u}^k,\mathbf{y})\right]\right)$$

5: Find

$$\mathbf{w}^{k+1} \propto \exp\left(\mathrm{E}_{\mathrm{q}(\Theta_{\mathbf{w}})}\left[\log \mathrm{F}(\mathbf{x}^k,\mathbf{h}^{k+1},\Omega^k,\mathbf{w},\mathbf{u}^k,\mathbf{y})\right]\right)$$

6: Find

$$\mathbf{u}^{k+1} \propto \exp\left(\mathrm{E}_{\mathrm{q}(\Theta_{\mathbf{u}})}\left[\log \mathrm{F}(\mathbf{x}^k,\mathbf{h}^{k+1},\Omega^k,\mathbf{w}^{k+1},\mathbf{u},\mathbf{y})\right]\right)$$

7: Find

$$\mathrm{q}^{k+1}(\Omega) \propto \exp\left(\mathrm{E}_{\mathrm{q}(\Theta_{\Omega})}\left[\log \mathrm{F}(\mathbf{x}^k,\mathbf{h}^{k+1},\Omega,\mathbf{w}^{k+1},\mathbf{u}^{k+1},\mathbf{y})\right]\right)$$

$$\mathrm{E}^k(\mathbf{x}) = \mathrm{cov}^k(\mathbf{x})\,\beta^k\mathbf{H}^t\mathbf{y}, \quad (23)$$

with $(\cdot)^t$ denoting the transpose operator and

$$W(\mathbf{w}^k) = diag\left(\frac{1}{\sqrt{w_i^k}}\right),\ i=1,\ldots,N. \quad (24)$$

Similarly, $\mathrm{q}^k(\mathbf{h})$ is found in the second step of the algorithm as an N-dimensional Gaussian distribution, given by

$$\mathrm{q}^{k+1}(\mathbf{h}) = \mathcal{N}\left(\mathbf{h}\mid \mathrm{E}^{k+1}(\mathbf{h}),\mathrm{cov}^{k+1}(\mathbf{h})\right), \quad (25)$$

with parameters

$$\mathrm{cov}^{k+1}(\mathbf{h}) = \big[\beta^k\mathbf{X}^t\mathbf{X} + N\beta^k\mathrm{cov}^k(\mathbf{x}) + \alpha_{\mathrm{bl}}^k(\Delta^1)^t U(\mathbf{u}^k)(\Delta^1) + \alpha_{\mathrm{bl}}^k(\Delta^2)^t U(\mathbf{u}^k)(\Delta^2)\big]^{-1}, \quad (26)$$

and

$$\mathrm{E}^{k+1}(\mathbf{h}) = \mathrm{cov}^{k+1}(\mathbf{h})\beta^k\mathbf{x}^t\mathbf{y}, \quad (27)$$

where the matrix $U(\mathbf{u}^k)$ is defined as

$$U(\mathbf{u}^k) = diag\left(\frac{1}{\sqrt{u_i^k}}\right).\ i=1,\ldots,N \quad (28)$$

In the third and fourth steps of the algorithm, we find the vectors $\mathbf{w}^{k+1}$ and $\mathbf{u}^{k+1}$ as

$$\mathbf{w}_i^{k+1} = \mathrm{E}_{\mathrm{q}^k(\mathbf{x})}[(\Delta_i^1(\mathbf{x}))^2 + (\Delta_i^2(\mathbf{x}))^2],\ \ i=1,\ldots,N, \quad (29)$$

$$\mathbf{u}_i^{k+1} = \mathrm{E}_{\mathrm{q}^k(\mathbf{h})}[(\Delta_i^1(\mathbf{h}))^2 + (\Delta_i^2(\mathbf{h}))^2],\ \ i=1,\ldots,N. \quad (30)$$

Finally, in the last step, we find the distributions of the hyperparameters. It can be shown that they have Gamma distributions, given by

$$\mathrm{q}^{k+1}(\alpha_{\mathrm{im}}) \propto \alpha_{\mathrm{im}}^{N/2}\exp\left[-\alpha_{\mathrm{im}}\sum_i\sqrt{w_i^{k+1}}\right],$$

$$\mathrm{q}^{k+1}(\alpha_{\mathrm{bl}}) \propto \alpha_{\mathrm{bl}}^{N/2}\exp\left[-\alpha_{\mathrm{bl}}\sum_i\sqrt{u_i^{k+1}}\right],$$

$$\mathrm{q}^{k+1}(\beta) \propto \beta^{N/2}\exp\left[-\beta\left(\frac{\mathrm{E}_{\mathrm{q}^k(\mathbf{x})\mathrm{q}^{k+1}(\mathbf{h})}\left[\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2\right]}{2}\right)\right],$$

with the corresponding means

$$(\alpha_{\mathrm{im}}^{k+1})^{-1} = \frac{\sum_i\sqrt{w_i^{k+1}}}{N/2+1}, \quad (31)$$

$$(\alpha_{\mathrm{bl}}^{k+1})^{-1} = \frac{\sum_i\sqrt{u_i^{k+1}}}{N/2+1}, \quad (32)$$

$$(\beta^{k+1})^{-1} = \frac{\mathrm{E}_{\mathrm{q}^k(\mathbf{x})\mathrm{q}^{k+1}(\mathbf{h})}\left[\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2\right]}{N+2}. \quad (33)$$

Note that in Algorithm 1 no assumptions were imposed on the posterior approximations $\mathrm{q}(\mathbf{x})$ and $\mathrm{q}(\mathbf{h})$. We can, however, assume that these distributions are *degenerate*, that is, $\mathrm{q}^k(\mathbf{x}) = \delta(\mathbf{x}-\mathbf{x}^k)$ and $\mathrm{q}^k(\mathbf{h}) = \delta(\mathbf{h}-\mathbf{h}^k)$ at each iteration k. Using this assumption we obtain Algorithm 2, which is similar to Algorithm 1 except that the cross-terms $N\beta^k\mathrm{cov}^k(\mathbf{h})$ and $N\beta^k\mathrm{cov}^k(\mathbf{x})$ in Eqs. (22) and (26) are set equal to zero, and the expectations with respect to $\mathbf{x}$ and $\mathbf{h}$ are replaced by the values of $\mathbf{x}$ and $\mathbf{h}$. Note that in Algorithm 2, if the hyperparameters Ω are assumed to be known, the estimates of $\mathbf{x}$ and $\mathbf{h}$ correspond to the *maximum a posteriori* estimates, that is,

$$(\hat{\mathbf{x}},\hat{\mathbf{h}}) = \underset{(\mathbf{x},\mathbf{h})}{\mathrm{argmax}}\,\mathrm{p}(\alpha_{\mathrm{im}},\alpha_{\mathrm{bl}},\beta,\mathbf{x},\mathbf{h}|\mathbf{y}), \quad (34)$$

which can be shown to be equivalent to Eq. (2). Therefore, the TV-based algorithms proposed in [1] and [2] are special cases of Algorithm 2, with the values of the hyperparameters assumed to be known.

Finally we would like to comment on the computational complexity of the algorithms. The estimates of $\mathbf{x}^k$ and $\mathbf{h}^k$ are found from Eqs. (23) and (27) (with the expectations removed in Algorithm 2). The direct solutions of these equations are hard to find due to their very high dimensionality. Therefore, we adopt a conjugate gradient (CG) algorithm with diagonal preconditioning to solve these systems numerically. Note that more advanced CG preconditioners [10], or other methods such as gradient descent can also be employed.

4. EXPERIMENTAL RESULTS

A number of experiments have been performed with the proposed algorithms. In the experiments reported below, Algorithm 1 is denoted by *TV1* and Algorithm 2 by *TV2*. We present results obtained by applying the proposed algorithms on "Lena" and "Satellite" images with an out-of-focus blur with radius 4, and white Gaussian noise is added to the blurred images to obtain degraded images with blurred-signal-to-noise ratios (BSNR) of 20 dB, 40 dB and 60 dB. The original images as well as the degraded versions are shown in Fig. 1.

In the experiments, we used $\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2 / \|\mathbf{x}^{k-1}\|^2 < 10^{-5}$ (or $\mathrm{E}^k(\mathbf{x})$ instead of $\mathbf{x}^k$) as the convergence criterion to terminate the algorithms, and a threshold of 10^{-5} is used to terminate the CG iterations. The initial values for the algorithms are chosen as follows: The observed image $\mathbf{y}$ is used as the initial estimate of

Table 1: ISNR values, number of iterations and estimated noise variances for the Lena and Satellite images degraded by an out-of-focus blur with radius 4.

		Lena			Satellite		
BSNR	Method	ISNR (dB)	iterations	$1/\beta$	ISNR (dB)	iterations	$1/\beta$
60dB	*TV1*	5.20	25	1×10^{-3}	14.37	27	0.5×10^{-3}
	TV2	5.25	23	2×10^{-3}	12.72	19	1×10^{-3}
	TV1-NB	12.66	11	6×10^{-5}	18.03	12	7×10^{-4}
	TV2-NB	12.65	11	6×10^{-5}	18.02	12	7×10^{-4}
40dB	*TV1*	7.62	58	0.08	8.54	25	0.13
	TV2	7.60	68	0.08	8.55	26	0.12
	TV1-NB	8.90	14	0.11	8.80	16	0.18
	TV2-NB	8.89	13	0.11	8.83	16	0.18
20dB	*TV1*	2.45	31	19.03	1.98	35	22.06
	TV2	2.47	29	19.00	1.93	35	22.05
	TV1-NB	3.13	5	15.96	3.28	5	18.01
	TV2-NB	3.12	6	15.96	3.28	5	18.00

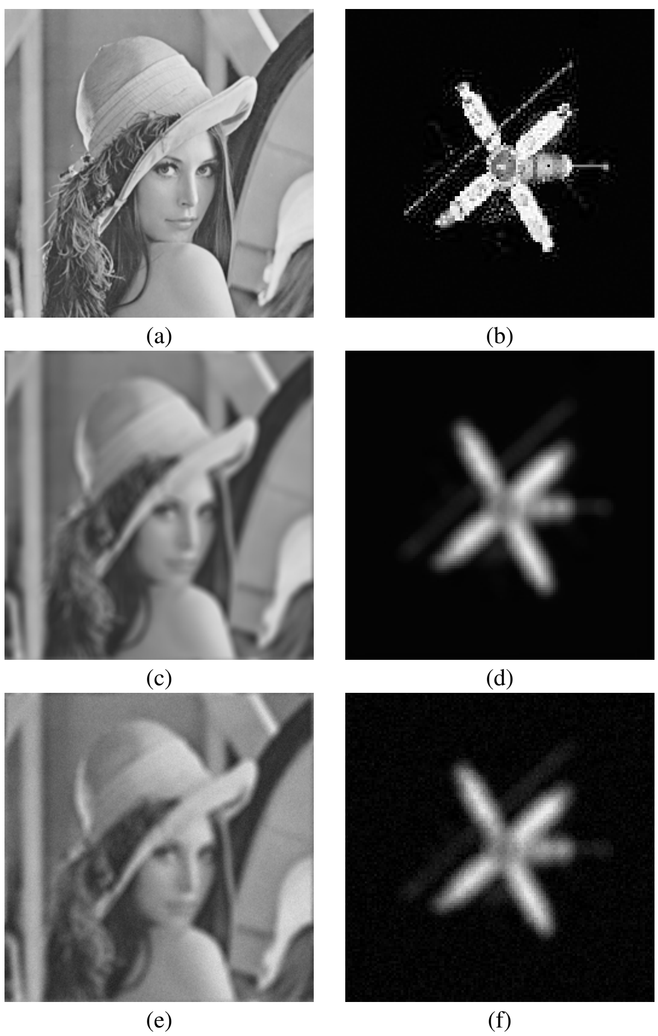


Figure 1: *Left column:* (a) Original Lena image; degraded versions with an out-of-focus blur with radius 4 and Gaussian noise of variances (c) 2×10^{-3} (BSNR = 60 dB), and (e) 17.81 (BSNR = 20 dB). *Right column:* (b) Original Satellite image; degraded versions with an out-of-focus blur with radius 4 and Gaussian noise of variances (d) 2×10^{-3} (BSNR = 60 dB), and (f) 19.44 (BSNR = 20 dB).

$\mathbf{x}^1$, and an out-of-focus blur with radius 8 as the initial estimate of $\mathbf{h}^1$. The covariance matrices $\text{cov}^1(\mathbf{h})$ and $\text{cov}^1(\mathbf{x})$ are set equal to zero, and the initial hyperparameter values are calculated from Eqs. (31)-(33). Note that except the initial blur, all parameters are calculated automatically using the observed image. In each iteration, to achieve physically meaningful solutions, we imposed positivity and symmetry constraints on the estimated blur, as in [1].

The quantitative results in terms of estimated noise variances $1/\beta$ and improvement-in-signal-to-noise ratios (ISNR) are presented in Table 1. where ISNR is defined as $10\log_{10}(\parallel x-y\parallel^2 / \parallel x-\hat{x}\parallel^2)$, with x, y, and $\hat{x}$ the original, observed, and estimated images, respectively. The ISNR results by the non-blind versions of the algorithms, where the PSF is known, are also included in this table, and they are denoted by *TV1-NB* and *TV2-NB*. It is clear from Table 1 that the proposed algorithms are very successful in estimating the noise variance. Restored images are shown in Figs. 2 and 3. Although the algorithms provide similar restoration performance in terms of ISNR, visually the restored images by *TV1* exhibit less ringing than the ones by *TV2*. Another important remark is that although the non-blind algorithms clearly outperform the blind versions in terms of ISNR, the blindly restored images are visually almost as good as the non-blind restoration results in the BSNR = 40dB case, and at acceptable levels in the BSNR = 20dB case, as can be seen from Fig. 2.

We note here that the proposed algorithms are very robust to the initial selected value of the blur. When an out-of-focus blur with radius 10 is chosen as $\mathbf{h}^1$, which is very different than the true blur, the ISNR values in the restoration of the Lena image are 5.07 dB for *TV1* and 6.07 dB for *TV2* for 60 dB BSNR, and 2.45 dB for *TV1* and 2.30 dB for *TV2* for 20 dB BSNR, similarly to the results in Table 1.

One dimensional slices through the origin of the estimated blurs corresponding to the restorations of the Lena image are shown in Fig. 4. It is clear that the algorithms provide accurate estimates of the true PSF at both noise levels. It should also be emphasized that if the TV-prior on blur is replaced by a SAR prior, the algorithms can not provide accurate estimates of the PSF, and therefore the resulting restorations are blurry. Moreover, if the noise level is high, e.g., BSNR = 20dB, in most images the algorithms with a SAR prior for the blur converge to flat images.

It has been reported [11] that the original TV-based algorithms proposed in [1] and [2] fail to provide satisfactory performances in images where the background is not black. Semi-blind restoration approaches are proposed to alleviate this problem [11]. However, as is clear from Fig. 2, the performance of the proposed algorithms does not decrease even in this case and high quality restored images are obtained. This is most probably due to the adaptive updating of the model parameters and in the case of *TV1*, utilizing the uncertainty of the variables in the estimation of other unknowns.

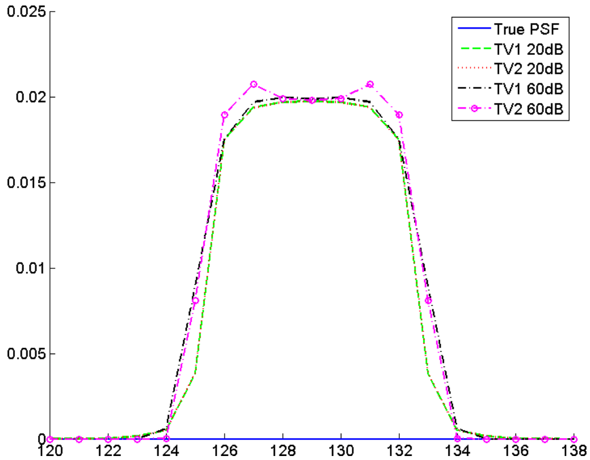
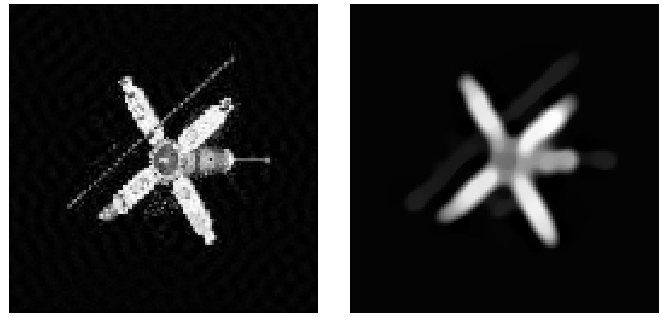


Figure 4: One-dimensional slices through the origin of the original and estimated PSFs in the restoration of the Lena image with algorithms *TV1* and *TV2*.

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

Figure 2: Restorations of the Lena image blurred with an out-of-focus blur with radius 4, and BSNR = 60 dB with (a) *TV1*, (b) *TV2*, and BSNR = 20 dB with (c) *TV1* and (d) *TV2*. The non-blind restoration results by *TV1-NB* are shown in (e) for BSNR = 60 dB and (f) for BSNR = 20dB.



(a) (b)

Figure 3: *TV1* restorations of the Satellite image blurred with an out-of-focus blur with radius 4, and (a) BSNR = 60 dB and (b) BSNR = 20 dB. The corresponding degraded images are shown in Fig. 1(d) and Fig. 1(f), respectively.

5. CONCLUSIONS

In this paper we represented a novel methodology for parameter estimation in TV-based blind deconvolution. Using a hierarchical Bayesian model and variational distribution approximations, the posterior distributions of the reconstructed image, blur and hyperparameters are simultaneously estimated. We have provided two different algorithms that resulted from this formulation, both of which are fully-automated. We have also shown that some TV-based approaches are special cases of our formulation with known hyperparameter values. Experimental results demonstrated the performance of the proposed algorithms.

REFERENCES

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

[2] Y. L. You and M. Kaveh, "Blind image restoration by anisotropic regularization," *IEEE Trans. Image Processing*, vol. 8, no. 3, pp. 396–407, 1999.

[3] S. D. Babacan, R. Molina, and A.K. Katsaggelos, "Total variation blind deconvolution using a variational approach to parameter, image, and blur estimation," in *EUSIPCO 2007*, Poznan, Poland, Sept. 2007.

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

[5] S. D. Babacan, R. Molina, and A.K. Katsaggelos, "Parameter estimation in TV image restoration using variational distribution approximation," *IEEE Trans. on Image Processing*, , no. 3, pp. 326–339, 2008.

[6] A. C. Likas and N. P. Galatsanos, "A variational approach for Bayesian blind image deconvolution," *IEEE Trans. Signal Processing*, vol. 52, no. 8, pp. 2222–2233, 2004.

[7] S. Kullback, *Information Theory and Statistics*, New York, Dover Publications, 1959.

[8] Y. L. You and M. Kaveh, "A regularization approach to joint blur and image restoration," *IEEE Trans. Image Processing*, vol. 5, no. 3, pp. 416–428, 1996.

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

[10] R. H. Chan, T. F. Chan, and C-K Wong, "Cosine transform based preconditioners for total variation deblurring," *IEEE Trans. Image Processing*, vol. 8, no. 10, pp. 1472–1478, Oct 1999.

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