

Variational Bayesian Sparse Kernel-Based Blind Image Deconvolution With Student's-t Priors

Dimitris G. Tzikas, Aristidis C. Likas, *Senior Member, IEEE*, and Nikolaos P. Galatsanos, *Senior Member, IEEE*

Abstract—In this paper, we present a new Bayesian model for the blind image deconvolution (BID) problem. The main novelty of this model is the use of a sparse kernel-based model for the point spread function (PSF) that allows estimation of both PSF shape and support. In the herein proposed approach, a robust model of the BID errors and an image prior that preserves edges of the reconstructed image are also used. Sparseness, robustness, and preservation of edges are achieved by using priors that are based on the Student's-t probability density function (PDF). This pdf, in addition to having heavy tails, is closely related to the Gaussian and, thus, yields tractable inference algorithms. The approximate variational inference methodology is used to solve the corresponding Bayesian model. Numerical experiments are presented that compare this BID methodology to previous ones using both simulated and real data.

Index Terms—Bayesian approach, blind image deconvolution (BID), inverse problem, kernel model, sparse prior, student-t distribution.

I. INTRODUCTION

In blind image deconvolution (BID), both the initial image and the point spread function (PSF) are unknown. Thus, for this problem, the observed data are not sufficient to uniquely specify the unknown image and PSF. In order to resolve this ambiguity, prior knowledge (constraints) has to be used for both the image and the PSF. Over the years, a number of methodologies have been employed to introduce constraints in BID. For an almost ten-year-old survey paper on this problem, the reader is referred to [1], [2]. A very recent edited book on BID methods is [3].

One category of such methods is based on regularization using the total variation (TV) principle. These methods define a distance function based on the data and use smoothness constraints on both the image and the PSF based on the TV principle [4]. A survey of recent developments on TV methods

Manuscript received March 24, 2008; revised November 11, 2008. Current version published March 13, 2009. This work was supported in part by the European Union—European Social Fund (ESF) and National Sources, in the framework of the program “Pythagoras II” of the “Operational Program for Education and Initial Vocational Training” of the 3rd Community Support Framework of the Hellenic Ministry of Education. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Stanley J. Reeves.

D. G. Tzikas and A. C. Likas are with the Department of Computer Science, University of Ioannina, Ioannina, GR 45110, Greece (e-mail: tzikas@cs.uoi.gr; arly@cs.uoi.gr).

N. P. Galatsanos is with the Department of Electrical and Computer Engineering, University of Patras, Rio, GR 26500, Greece. (e-mail: ngalatsanos@upatras.gr).

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.2008.2011757

in image recovery problems and a book containing a review of the recent developments in mathematical tools for low level image processing problems can be found in [5] and [6], respectively. Methods based on anisotropic diffusion regularization have been also proposed [7]; however, they require the choice of the diffusion operator. There are also methods based on soft constraints [8], [9], which are very flexible; however, the form and the type of the used soft constraints is ad-hoc. Methods based on sparse image representations and quasi likelihood criteria have been also suggested [10].

Another way to apply constraints to the image and the PSF, is through the use of the Bayesian methodology. In this approach the unknown quantities are assumed to be random variables and suitable prior distributions are selected to impose the desired characteristics [11]–[16]. Unfortunately, since the BID data generation model is nonlinear, the posterior distribution of the unknown image and PSF can not be computed analytically. Thus, Bayesian inference using conventional methods, such as maximum likelihood (ML) via the expectation maximization (EM) algorithm, cannot be applied.

These difficulties can be overcome using the variational Bayesian methodology [17] and [18]. To our knowledge, this methodology was first applied to the BID problem in [13]. In this paper, the PSF and the image were modeled by an exponential and a mixture of exponential distributions, respectively. Furthermore, the support of the PSF was known, and the images were line drawings which are sparse, in the sense that their intensity is zero at most locations. This work was recently extended for natural scene images in [14] with promising results. More specifically, a mixture of Gaussians for the gradient of the image, and a mixture of exponentials for the PSF were used. This PSF model allows only positive PSF intensities and encourages sparsity, all of which are desirable properties for BID. However, it does not model spatial PSF correlations. In another line of work [15], a simultaneously autoregressive (SAR) prior and a Gaussian prior with unknown mean and spherical covariance have been used for the image and PSF, respectively. This methodology was extended in [16] to account for spatial PSF correlations using SAR models for both PSF and the image. However, this approach fails to model edges in the image or PSF and does not provide a mechanism to estimate the support of the PSF.

In this paper, we propose a Bayesian model for the BID problem that allows reconstruction of image edges, models spatial PSF correlations and estimates the PSF support. The main contribution of this paper, is a model that enforces PSF smoothness and simultaneously estimates the PSF support. Specifically, we model the PSF as a linear combination of

kernel functions that are placed at all the pixels of the image. Thus, the amount of smoothness can be controlled by selecting the kernel function. The support of the PSF can be arbitrarily large, since we placed kernel functions at all image pixels. However, we assume that the distribution of the weights of the kernels that models the PSF is a heavy tailed Student's-t distribution. This distribution favors sparse models, forcing most of the weights to become zero and, therefore, limiting the support of the PSF. Furthermore, in order to promote smooth image estimates, we constrain the local image differences, by assuming that they follow a zero-mean Student's-t distribution in order to allow reconstruction of edges[19]. Finally, we model the errors of the imaging model with a Student's-t distribution. This is important, not only because the noise in the observed image may not be Gaussian, but also because inaccurate PSF estimates produce heavy tailed errors, since the BID model is nonlinear.

In the proposed model, we use extensively the Student's-t pdf. This pdf can be considered as a generalization of the Gaussian since with appropriate selection of its parameters, it can have heavy tails and in the limit it can become either Gaussian, or uninformative; see Fig. 1(a). Thus, when errors of a model are assumed Student's-t distributed, this yields robust estimators. Furthermore, when the parameters of a model are assumed Student's-t distributed, this results in sparse models i.e., models with few nonzero parameters. Such models are very attractive and are currently a hot topic of research since they automatically provide a mechanism to determine their complexity[20]. The sparsity of such models can be understood by observing the plots of the 2-D pdfs in Fig. 1(b). Most of the mass of the Student's-t pdf is concentrated along the axes and the center, unlike the Gaussian, where it is evenly distributed around ellipses, as shown in Fig. 1(c). This observation can be generalized for vectors of arbitrary dimension, where the Student's-t pdf assigns large probability mass to estimations that contain a large number of zero elements. The Laplacian pdf which has heavy tails has also been used for encoding sparse models[21]. Another very attractive property of the Student's-t pdf is that it can be generated by the superposition of an infinite number of Gaussian pdfs with common mean and precision (inverse variance) which is Gamma distributed [17]. This allows tractable Bayesian inference for models that use this pdf [22]. Specifically, if we assume a vector $\mathbf{x} = (x_1, \dots, x_N)^T$ of independent random variables x_i that follow a Student's-t distribution given by $p(\mathbf{x}) = \prod_{i=1}^N \text{St}(x_i | \mu, \lambda, \nu)$ and introduce the independent hidden variables $\boldsymbol{\tau} = (\tau_1, \dots, \tau_N)^T$, with $p(\boldsymbol{\tau}) = \prod_{i=1}^N \text{Gamma}(\tau_i | a, b)$ we can write the Student's-t pdf as $p(\mathbf{x}) = \int p(\mathbf{x} | \boldsymbol{\tau}) p(\boldsymbol{\tau}) d\boldsymbol{\tau}$, where $p(\mathbf{x} | \boldsymbol{\tau}) = \text{N}(\mathbf{x} | 0, \text{diag}\{\boldsymbol{\tau}\})$. Here

$$\text{St}(x | \mu, \lambda, \nu) = \frac{\Gamma\left(\frac{(\nu+1)}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\lambda}{\nu\pi}\right)^{1/2} \left(1 + \frac{\lambda}{\nu}(x - \mu)^2\right)^{-(\nu+1)/2}$$

is the Student's-t distribution with mean μ , precision λ and degrees of freedom ν

$$\text{N}(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-M/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{t} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{t} - \boldsymbol{\mu})\right]$$

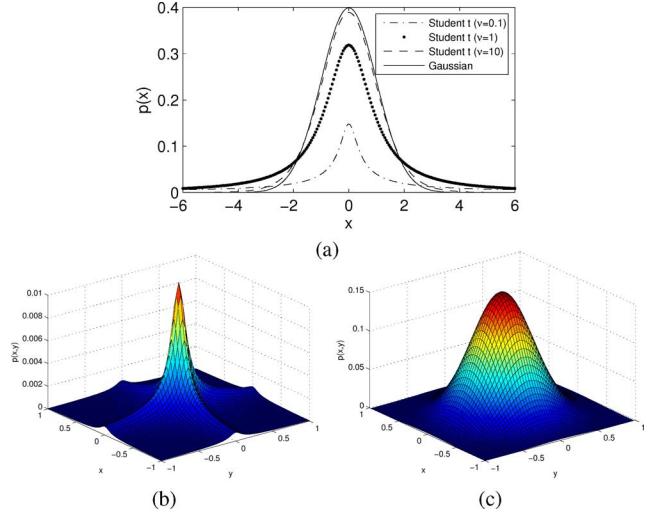


Fig. 1. (a) Student t pdf with 0.1, 1, and 10 degrees of freedom compared to the Gaussian pdf. Two-dimensional plot of (b) the Student's-t pdf with 0.1 degrees of freedom and (c) the Gaussian pdf.

is the multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\text{Gamma}(\tau | a, b) = \Gamma(a)^{-1} b^a \tau^{a-1} e^{-b\tau}$ is the Gamma distribution with parameters a and b and $\Gamma(x) = \int t^{x-1} e^{-t} dt$.

The rest of this paper is organized as follows. In Section II, the Bayesian model is presented. In Section III, a brief introduction to the variational methodology is presented and the variational methodology is applied for inference to the proposed model. In Section IV, we present experiments with artificially blurred images where the ground truth is known and with real astronomical images. In these experiments, we compare the proposed methodology with Bayesian methods that use Gaussian priors and TV based methods and the advantages of the proposed methodology are demonstrated. Finally, in Section V, we provide conclusions and directions for future work.

II. BID MODEL

We assume that the observed image $g(\mathbf{x})$ is given by convolving an unknown image $f(\mathbf{x})$ with an unknown PSF $h(\mathbf{x})$. To account for errors, additive, independent, identically distributed noise $n(\mathbf{x})$ is also assumed. This model is written as

$$g(\mathbf{x}) = f(\mathbf{x}) * h(\mathbf{x}) + n(\mathbf{x}) \quad (1)$$

where $\mathbf{x} = (x_1, x_2) \in \Omega_I$, $\Omega_I \subset \mathbb{R}^2$ is the support of the image and $*$ denotes 2-D circular convolution. Equivalently, this can be written in vector form as

$$\mathbf{g} = \mathbf{f} * \mathbf{h} + \mathbf{n} \quad (2)$$

where \mathbf{g} , \mathbf{f} , \mathbf{h} , and \mathbf{n} are $M \times 1$ lexicographically ordered vectors (M is the number of pixels) of the intensities of the degraded image, observed image, PSF and additive noise respectively. Here, we introduce the $M \times M$ block-circulant matrices \mathbf{F} and \mathbf{H} , which implement 2-D convolution with the vectors \mathbf{f}

and \mathbf{h} , respectively, so that $\mathbf{F}\mathbf{h} = \mathbf{H}\mathbf{f} = \mathbf{f} * \mathbf{h}$. Then, the BID model in (2) can be written as

$$\mathbf{g} = \mathbf{F}\mathbf{h} + \mathbf{n} = \mathbf{H}\mathbf{f} + \mathbf{n}. \quad (3)$$

The blind image deconvolution problem is difficult because there are too many unknown parameters that have to be estimated. More specifically, the number of unknown parameters \mathbf{h} and \mathbf{f} is larger than the number of observations \mathbf{g} , and, thus, reliable estimation of these parameters can only be achieved by exploiting prior knowledge of the characteristics of the unknown quantities. Following the Bayesian framework, the unknown parameters are treated as hidden random variables and prior knowledge is expressed by assuming that they have been sampled from specific prior distributions.

A. PSF Kernel Model

We model the PSF as the linear combination of basis functions

$$h(\mathbf{x}) = \sum_{i=1}^M w_i \phi_i(\mathbf{x}) \quad (4)$$

where $\phi_i(\mathbf{x}) = R(\mathbf{x}, \mathbf{x}_i)$ is a kernel function centered at $\mathbf{x}_i = (x_{i1}, x_{i2}) \in \Omega_I$ and $w_i \in \mathbb{R}$. We denote as $\mathbf{h} = (h(\mathbf{x}_1), \dots, h(\mathbf{x}_M))^T$ the vector of values of the PSF $h(\mathbf{x})$ at each \mathbf{x}_i and with $\boldsymbol{\phi}_i = (\phi_i(\mathbf{x}_1), \dots, \phi_i(\mathbf{x}_M))^T$ the corresponding basis vector for $\phi_i(\mathbf{x})$. Then the PSF vector \mathbf{h} is modeled as the linear combination of the basis vectors $\boldsymbol{\phi}_i$

$$\mathbf{h} = \sum_{i=1}^M w_i \boldsymbol{\phi}_i. \quad (5)$$

We further assume that the kernel is invariant to translations, i.e., $R(\mathbf{x}, \mathbf{x}_i) = R(\mathbf{x} - \mathbf{x}_i)$; thus, (5) can be written as

$$\mathbf{h} = \boldsymbol{\phi} * \mathbf{w} = \Phi \mathbf{w} = \mathbf{W} \boldsymbol{\phi} \quad (6)$$

where $\mathbf{w} = (w_1, \dots, w_M)^T$ are the weights of the linear combination and Φ, \mathbf{W} are $M \times M$ block-circulant matrices that implement 2-D convolution with $\boldsymbol{\phi} = \boldsymbol{\phi}_1$ and \mathbf{w} respectively, so that $\Phi \mathbf{w} = \mathbf{W} \boldsymbol{\phi} = \mathbf{w} * \boldsymbol{\phi}$. Thus, the BID data generation model (2) can be written as

$$\mathbf{g} = \mathbf{F}\Phi\mathbf{w} + \mathbf{n} = \Phi\mathbf{W}\mathbf{f} + \mathbf{n}. \quad (7)$$

In this paper, Gaussian kernel function of the form $R(\mathbf{x}, \mathbf{x}_0) = \exp[-(1/2\sigma_\phi^2)\|\mathbf{x} - \mathbf{x}_0\|^2]$ (RBF kernels) is considered, which produces smooth estimates of the PSF. However, any other type of kernel could be used as well. It is even possible that many different types of kernels are used simultaneously, at a small additional computational cost [23].

B. PSF Sparseness

A hierarchical prior that enforces sparsity is imposed on the weights \mathbf{w} [20]

$$p(\mathbf{w}|\boldsymbol{\alpha}) = N(\mathbf{w}|0, \mathbf{A}^{-1}) \quad (8)$$

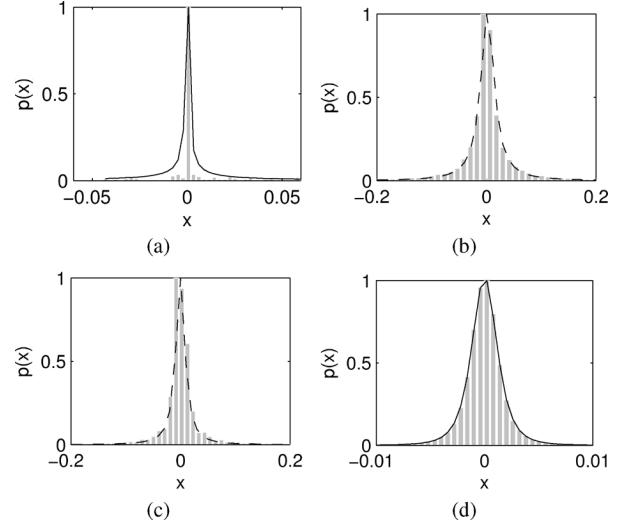


Fig. 2. Histograms of (a) estimated PSF weights assuming known PSF, (b) horizontal, and (c) vertical local differences of the "Lena" image and (d) model errors of an image restoration method using incorrect PSF estimation. Solid lines show fits by the Student's-t pdf with parameters (a) $\mu = 2.51 \times 10^{-34}$, $\lambda = 9.05 \times 10^{37}$, $\nu = 0.043$; (b) $\mu = 1.7 \times 10^{-3}$, $\lambda = 4.59 \times 10^3$, $\nu = 1.09$; (c) $\mu = -4 \times 10^{-4}$, $\lambda = 1.03 \times 10^4$, $\nu = 1.132$; and (d) $\mu = 2.39 \times 10^{-6}$, $\lambda = 6.68 \times 10^5$, $\nu = 3.12$.

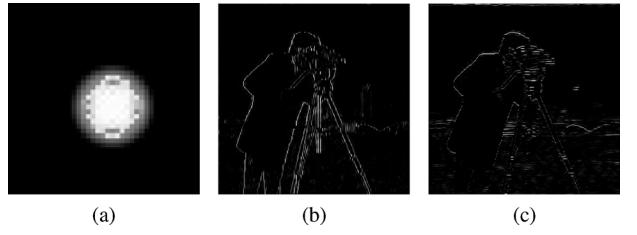


Fig. 3. Example of the estimated local variances (a) α^{-1} of the PSF weights for a uniform 7×7 square-shaped PSF, (b) and (c) $(\gamma^1)^{-1}$, and $(\gamma^2)^{-1}$ of the image model residuals.

where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_M)^T$, $\mathbf{A} = \text{diag}\{\boldsymbol{\alpha}\}$. Each weight is assigned a separate local precision parameter α_i , which is treated as a random variable that follows a Gamma distribution

$$p(\boldsymbol{\alpha}) = \prod_{i=1}^M \text{Gamma}(\alpha_i | a^\alpha, b^\alpha). \quad (9)$$

This hierarchical prior is equivalent to a Student's-t pdf. Fig. 2(a) shows a histogram of the estimated weights when the PSF is a 7×7 uniform square-shaped function and it is assumed known. It is apparent that the pdf of the weights is very heavy tailed and that there are only few nonzero weights. For this reason, we set $a^\alpha = b^\alpha = 0$ that define a very heavy tailed, uninformative Student's-t distribution. It is interesting that the hidden variables α_i of this Student's-t distribution provide an estimate of the support of the PSF. Specifically, local precision α_i that correspond to kernels outside the support of the PSF obtain very large values; therefore, those kernels are pruned by setting $w_i = 0$. This is demonstrated in Fig. 3(a) where we show the estimated local variances for a BID problem with a 7×7 uniform PSF. Notice that outside a limited area that captures the support of this PSF these variances are zero.

C. Image Model

The image prior that we use is based on K filtered versions of the image: $\boldsymbol{\epsilon}^k = \mathbf{Q}^k \mathbf{f}$, where \mathbf{Q}^k are $M \times M$ convolutional operators of the filters ($k = 1, \dots, K$). Specifically, we use horizontal and vertical first order local differences, by defining $K = 2$, \mathbf{Q}^1 and \mathbf{Q}^2 so that

$$\epsilon^1(x, y) = f(x, y) - f(x + 1, y) \quad (10)$$

$$\epsilon^2(x, y) = f(x, y) - f(x, y + 1). \quad (11)$$

Without any changes in the method, we could also use other convolutional operators \mathbf{Q}^k [24]. In practice, we join all operators \mathbf{Q}^k in the $KM \times M$ operator $\tilde{\mathbf{Q}} = (\mathbf{Q}^{1T}, \dots, \mathbf{Q}^{KT})^T$ that produces the $KM \times 1$ vector $\tilde{\boldsymbol{\epsilon}} = (\epsilon^{1T}, \dots, \epsilon^{KT})^T$

$$\tilde{\boldsymbol{\epsilon}} = \tilde{\mathbf{Q}}\mathbf{f} = ((\mathbf{Q}^1\mathbf{f})^T, \dots, (\mathbf{Q}^K\mathbf{f})^T)^T. \quad (12)$$

We assume that ϵ_i^k is Gaussian distributed with distinct precision γ_i^k

$$p(\epsilon_i^k | \gamma_i^k) = N(\epsilon_i^k | 0, (\gamma_i^k)^{-1}). \quad (13)$$

Assuming the ϵ_i^k independent with respect to i , induces a prior for the image, which is given by

$$p_k(\mathbf{f} | \boldsymbol{\gamma}^k) = N(\mathbf{f} | 0, (\mathbf{Q}^{kT}\boldsymbol{\Gamma}^k\mathbf{Q}^k)^{-1}) \quad (14)$$

with $\boldsymbol{\gamma}^k = (\gamma_1^k \dots \gamma_M^k)^T$ and $\boldsymbol{\Gamma}^k = \text{diag}\{\boldsymbol{\gamma}^k\}$. In order to combine the information captured by each prior p_k , we define a composite prior, which is the product of them [25]

$$p(\mathbf{f} | \tilde{\boldsymbol{\gamma}}) = \frac{1}{Z} \prod_{k=1}^K p_k(\mathbf{f} | \boldsymbol{\gamma}^k) = N(\mathbf{f} | 0, (\tilde{\mathbf{Q}}^T \tilde{\mathbf{I}} \tilde{\mathbf{Q}})^{-1}) \quad (15)$$

with $\tilde{\boldsymbol{\gamma}} = (\boldsymbol{\gamma}^{1T}, \dots, \boldsymbol{\gamma}^{KT})^T$ and $\tilde{\mathbf{I}} = \text{diag}\{\tilde{\boldsymbol{\gamma}}\}$. Unfortunately, it is not possible to analytically compute the determinant $|\tilde{\mathbf{Q}}^T \tilde{\mathbf{I}} \tilde{\mathbf{Q}}|$ that is required to estimate the normalization constant Z since $\tilde{\mathbf{Q}}$ is not square. Instead, we approximate it as $|\tilde{\mathbf{Q}}^T \tilde{\mathbf{I}} \tilde{\mathbf{Q}}| \approx |\tilde{\mathbf{I}}| |\tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}}|$, giving

$$p(\mathbf{f} | \tilde{\boldsymbol{\gamma}}) \propto \prod_{k=1}^K \prod_{i=1}^M (\gamma_i^k)^{1/2} \exp\left[-\frac{1}{2} \mathbf{f}^T \tilde{\mathbf{Q}}^T \tilde{\mathbf{I}} \tilde{\mathbf{Q}} \mathbf{f}\right]. \quad (16)$$

Notice that the approximation only affects the normalizing constant of the pdf. Therefore, this is an *improper* pdf whose integral is not necessarily unity. Improper pdfs have been used in many other Bayesian methods [26]. The local precision parameters γ_i^k are assumed to be independent identically distributed, Gamma random variables

$$p(\tilde{\boldsymbol{\gamma}}) = \prod_{k=1}^K \prod_{i=1}^M \text{Gamma}(\gamma_i^k | a^\gamma, b^\gamma). \quad (17)$$

Thus, the prior on the first order local differences $\boldsymbol{\epsilon}^k$ is equivalent to a Student's-t pdf. The appropriateness of the Student's-t to model local image differences is demonstrated in Fig. 2(b) and (c), where a histogram of the horizontal and vertical local differences for the "Lena" image is shown. Notice, that most local differences, that correspond to smooth image regions, are very close to zero. However, there is also a significant number

of large values that correspond to image textured regions and image edges. A heavy tailed pdf, such as the Student's-t is necessary in order to model both smooth and textured regions or edges. At this point it is interesting to point out that the hidden variables of this Student's-t prior capture the image edge structure using a continuously valued model. This is demonstrated in Fig. 3(b) and (c), where the local variances $(\boldsymbol{\gamma}^k)^{-1}$ are shown. It is interesting to notice that the variances $(\boldsymbol{\gamma}^1)^{-1}$ and $(\boldsymbol{\gamma}^2)^{-1}$ provide the vertical and horizontal edge structure of the image, respectively.

D. Noise Model

The noise \mathbf{n} of the BID model (3) is assumed to be zero mean Gaussian distributed, given by

$$p(\mathbf{n} | \boldsymbol{\beta}) = \prod_{i=1}^M N(n_i | 0, \beta_i^{-1}) = N(\mathbf{n} | 0, \mathbf{B}^{-1}) \quad (18)$$

with $\boldsymbol{\beta} = (\beta_1, \dots, \beta_M)$ and $\mathbf{B} = \text{diag}\{\boldsymbol{\beta}\}$. The local precision parameters β_i are also assumed to be random variables with a Gamma prior

$$p(\boldsymbol{\beta}) = \prod_{i=1}^M \text{Gamma}(\beta_i | a^\beta, b^\beta). \quad (19)$$

This two-level hierarchical prior for noise is equivalent to a Student's-t pdf. Since the Student's-t distribution can have heavy tails, this prior model does not excessively penalize large errors of the BID model as the Gaussian does, and, thus, it produces robust estimators. This is a desirable feature in BID, because at least in the beginning of the algorithm, when the PSF has not yet been estimated adequately, the errors of the imaging model are heavy tailed, as shown in Fig. 2(d). This happens because an inaccurate PSF estimation introduces large errors near edges of the image and in textured regions, but only small errors in smooth regions of the image.

III. VARIATIONAL BAYESIAN INFERENCE

The observed variables of the proposed model are $\mathbf{D} = \{\mathbf{g}\}$, the hidden variables are $\boldsymbol{\theta} = \{\mathbf{w}, \mathbf{f}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}\}$, and the parameters of the model are $\boldsymbol{\xi} = \{a^\alpha, b^\alpha, a^\beta, b^\beta, a^\gamma, b^\gamma\}$. The dependencies among the random variables that define the proposed Bayesian model are shown in the graphical model of Fig. 4.

Because the BID model is nonlinear, the posterior distribution of the parameters $p(\boldsymbol{\theta} | \mathbf{D})$ cannot be computed. Thus, we can not apply exact inference methods, such as maximum likelihood via the EM algorithm. Instead, we resort to approximate inference and specifically to the variational Bayesian methodology [18], in which we assume a family of approximate posterior distributions $q(\boldsymbol{\theta})$, and then seek values for the parameters $\boldsymbol{\theta}$ that best approximate the true posterior $p(\boldsymbol{\theta} | \mathbf{D})$.

The evidence of the model $p(\mathbf{D}) = \int p(\mathbf{D}, \boldsymbol{\theta}) d\boldsymbol{\theta}$ can be decomposed as

$$\ln p(\mathbf{D}) = \mathcal{L}(\boldsymbol{\theta}) + KL(q(\boldsymbol{\theta}) || p(\boldsymbol{\theta} | \mathbf{D})) \quad (20)$$

where

$$\mathcal{L}(\boldsymbol{\theta}) = \int q(\boldsymbol{\theta}) \ln \frac{p(\mathbf{D}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \quad (21)$$

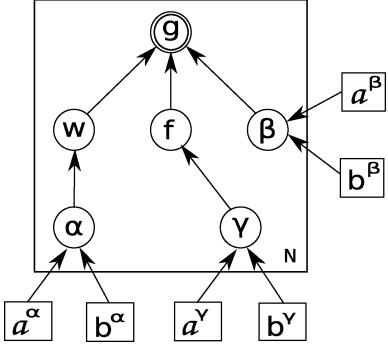


Fig. 4. Graphical model that describes the dependencies between the random variables of the proposed model. Circular nodes represent random variables, while square nodes represent parameters of the model. The observed variables are represented by double circled nodes.

is called the variational bound and

$$KL(q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\mathbf{D})) = - \int q(\boldsymbol{\theta}) \ln \frac{p(\boldsymbol{\theta}|\mathbf{D})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \quad (22)$$

is the Kullback–Leibler divergence between the approximating distribution $q(\boldsymbol{\theta})$ and the exact posterior distribution $p(\boldsymbol{\theta}|\mathbf{D})$. We find the best approximating distribution $q(\boldsymbol{\theta})$ by maximizing the variational bound \mathcal{L} , which is equivalent to minimizing the KL divergence $KL(q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\mathbf{D}))$

$$q(\boldsymbol{\theta}) = \underset{q(\boldsymbol{\theta})}{\text{argmax}} \mathcal{L}(\boldsymbol{\theta}) = \underset{q(\boldsymbol{\theta})}{\text{argmin}} KL(q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\mathbf{D})). \quad (23)$$

In order to perform the maximization of the variational bound with respect to the approximating distribution $q(\boldsymbol{\theta})$, we can assume a specific parametric form for it and then maximize with respect to the parameters. An alternative common approach is the mean field approximation, which assumes that the posterior distributions of the hidden variables are independent

$$q(\boldsymbol{\theta}) = \prod_i q(\boldsymbol{\theta}^i). \quad (24)$$

Then, the variational bound is maximized by

$$q(\boldsymbol{\theta}^i) = \frac{\exp[I(\boldsymbol{\theta}^i)]}{\int \exp[I(\boldsymbol{\theta}^i)] d\boldsymbol{\theta}^i} \quad (25)$$

where

$$I(\boldsymbol{\theta}^i) = \langle \ln p(\mathbf{D}, \boldsymbol{\theta}) \rangle_{q(\boldsymbol{\theta}^i)} = \int q(\boldsymbol{\theta}^i) \ln p(\mathbf{D}, \boldsymbol{\theta}) d\boldsymbol{\theta}^{i \setminus i} \quad (26)$$

$\boldsymbol{\theta}^{i \setminus i}$ denotes the vector of all hidden variables except $\boldsymbol{\theta}^i$ [18] and $\langle f(x) \rangle_{g(x)} = \int f(x)g(x) dx$ denotes the expected value if $f(x)$ with respect to $g(x)$.

Computation of $q(\boldsymbol{\theta}^i)$ is not straightforward, since $I(\boldsymbol{\theta}^i)$ depends on the approximate distribution $q(\boldsymbol{\theta}^i)$. Variational inference proceeds by assuming some initial distribution $q(\boldsymbol{\theta}_0)$ and then iteratively updating $q(\boldsymbol{\theta}^i)$ using (25) and (26). If the prior distribution $p(\boldsymbol{\theta})$ is defined in terms of some parameters ξ , then these can be estimated by maximizing the variational bound \mathcal{L} with respect to ξ .

A. Approximate Posterior Distributions

The approximate posterior distributions of the hidden variables can be computed using (25), as shown in Appendix A. Because we have used conjugate priors, the approximate posteriors have the same form as the priors. Specifically, the approximate posterior distributions of the PSF weights w and the image f are Gaussian and the distributions of the precision parameters α , β , and γ are Gamma

$$q(\mathbf{w}) = \mathcal{N}(\mathbf{w} | \boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w) \quad (27)$$

$$q(\mathbf{f}) = \mathcal{N}(\mathbf{f} | \boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f) \quad (28)$$

$$q(\boldsymbol{\alpha}) = \prod_{i=1}^M \text{Gamma}(\alpha_i | \tilde{a}^\alpha, \tilde{b}_i^\alpha) \quad (29)$$

$$q(\boldsymbol{\beta}) = \prod_{i=1}^M \text{Gamma}(\beta_i | \tilde{a}^\beta, \tilde{b}_i^\beta) \quad (30)$$

$$q(\boldsymbol{\gamma}) = \prod_{k=1}^K \prod_{i=1}^M \text{Gamma}(\gamma_i^k | \tilde{a}^\gamma, \tilde{b}_i^{\gamma^k}) \quad (31)$$

where

$$\boldsymbol{\mu}_w = \boldsymbol{\Sigma}_w \boldsymbol{\Phi}^T \langle \mathbf{F}^T \rangle \langle \mathbf{B} \rangle \mathbf{g} \quad (32)$$

$$\boldsymbol{\Sigma}_w = (\boldsymbol{\Phi}^T \langle \mathbf{F}^T \mathbf{B} \mathbf{F} \rangle \boldsymbol{\Phi} + \langle \mathbf{A} \rangle)^{-1} \quad (33)$$

$$\boldsymbol{\mu}_f = \boldsymbol{\Sigma}_f \boldsymbol{\Phi}^T \langle \mathbf{W}^T \rangle \langle \mathbf{B} \rangle \mathbf{g} \quad (34)$$

$$\boldsymbol{\Sigma}_f = (\boldsymbol{\Phi}^T \langle \mathbf{W}^T \mathbf{B} \mathbf{W} \rangle \boldsymbol{\Phi} + \tilde{\mathbf{Q}}^T \langle \tilde{\mathbf{F}} \rangle \tilde{\mathbf{Q}})^{-1} \quad (35)$$

$$\tilde{a}^\alpha = a^\alpha + \frac{1}{2} \quad (36)$$

$$\tilde{b}_i^\alpha = b_i^\alpha + \frac{1}{2} \langle w_i^2 \rangle \quad (37)$$

$$\tilde{a}^\beta = a^\beta + \frac{M}{2} \quad (38)$$

$$\tilde{b}_i^\beta = b_i^\beta + \frac{1}{2} \langle \mathbf{n} \mathbf{n}^T \rangle_{ii} \quad (39)$$

$$\tilde{a}^\gamma = a^\gamma + \frac{1}{2} \quad (40)$$

$$\tilde{b}_i^{\gamma^k} = b_i^{\gamma^k} + \frac{1}{2} \left(\mathbf{Q}^k \langle \mathbf{f} \mathbf{f}^T \rangle \mathbf{Q}^k \right)_{ii} \quad (41)$$

The required expected values can be computed as

$$\langle \mathbf{w} \rangle = \boldsymbol{\mu}_w \quad (42)$$

$$\langle w_i^2 \rangle = \mu_{w_i}^2 + \Sigma_{w_{ii}} \quad (43)$$

$$\langle \mathbf{f} \rangle = \boldsymbol{\mu}_f \quad (44)$$

$$\langle \mathbf{f} \mathbf{f}^T \rangle = \boldsymbol{\mu}_f \boldsymbol{\mu}_f^T + \boldsymbol{\Sigma}_f \quad (45)$$

$$\langle \alpha_i \rangle = \frac{\tilde{a}^\alpha}{\tilde{b}_i^\alpha} \quad (46)$$

$$\langle \beta_i \rangle = \frac{\tilde{a}^\beta}{\tilde{b}_i^\beta} \quad (47)$$

$$\langle \gamma_i^k \rangle = \frac{\tilde{a}^\gamma}{\tilde{b}_i^{\gamma^k}} \quad (48)$$

$$\langle \mathbf{n} \mathbf{n}^T \rangle = \mathbf{g} \mathbf{g}^T - 2 \boldsymbol{\Phi} \langle \mathbf{F} \mathbf{w} \rangle \mathbf{g}^T + \boldsymbol{\Phi} \langle \mathbf{F} \mathbf{w} \mathbf{w}^T \mathbf{F}^T \rangle \boldsymbol{\Phi}^T. \quad (49)$$

B. Parameter Estimation

The parameters a^β , b^β and a^γ , b^γ of the noise and image Gamma hyperpriors can be estimated by optimizing the variational bound \mathcal{L} (21), whose form is given in Appendix B. The derivatives of \mathcal{L} with respect to the above parameters are

$$\frac{\partial \mathcal{L}}{\partial a^\beta} = M \ln b^\beta - M\psi(a^\beta) + \sum_{i=1}^M \langle \ln \beta_i \rangle \quad (50)$$

$$\frac{\partial \mathcal{L}}{\partial b^\beta} = M \frac{a^\beta}{b^\beta} - \sum_{i=1}^M \langle \beta_i \rangle \quad (51)$$

$$\frac{\partial \mathcal{L}}{\partial a^\gamma} = MK \ln b^\gamma - MK\psi(a^\gamma) + \sum_{k=1}^K \sum_{i=1}^M \langle \ln \gamma_i^k \rangle \quad (52)$$

$$\frac{\partial \mathcal{L}}{\partial b^\gamma} = MK \frac{a^\gamma}{b^\gamma} - \sum_{k=1}^K \sum_{i=1}^M \langle \gamma_i^k \rangle \quad (53)$$

where $\psi(x)$ is the digamma function given by $\psi(x) = d \ln \Gamma(z)/dz = \Gamma'(z)/\Gamma(z)$ and $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$. We can obtain updates for these parameters by setting the above derivatives to zero. This cannot be done analytically for the parameters a^β and a^γ ; thus, we find a numerical solution using a combination of bisection, secant, and inverse quadratic interpolation methods, as implemented by matlab's fzero function.

C. Computational Issues

The computations in (32)–(49) involve matrix operations, whose dimension is $M \times M$, where M is the number of pixels in the image. Unfortunately, computation of Σ_f and Σ_w involves inversion of matrices that contain both diagonal and circulant matrices and cannot be performed explicitly for large M . However, diagonal and circulant matrices are easy to invert. For this reason, we approximate Σ_w (33) with a diagonal matrix and Σ_f (35) with a circulant matrix, as

$$\tilde{\Sigma}_w = \left(\text{diag}\{\Phi^T \langle F^T BF \rangle \Phi\} + \langle A \rangle \right)^{-1} \quad (54)$$

$$\tilde{\Sigma}_f = \left(\langle \bar{\beta} \rangle \Phi^T \langle W^T W \rangle \Phi + \langle \bar{\gamma} \rangle \tilde{Q}^T \tilde{Q} \right)^{-1} \quad (55)$$

with $\bar{\gamma} = (1/MK) \sum_{k=1}^K \sum_{i=1}^M \gamma_i^k$, $\bar{\beta} = (1/M) \sum_{i=1}^M \beta_i$ and

$$\langle W^T W \rangle = \langle W^T \rangle \langle W \rangle + I \sum_{i=1}^M \langle \tilde{\Sigma}_{wii} \rangle \quad (56)$$

$$\langle F^T BF \rangle = \langle F^T \rangle \langle B \rangle \langle F \rangle + \mathbf{E}_f \sum_{i=1}^M \langle \beta_i \rangle. \quad (57)$$

The diagonal approximation for matrix Σ_w is justified because parameters α_i that appear in the diagonal were found to dominate in (33). On the other hand, Σ_f is approximated with a circulant matrix because both the parameters β_i and γ_i^k obtain values in the same range. The above approximations are used for computation of \tilde{b}_i^α , \tilde{b}_i^β , and \tilde{b}_i^γ in (37), (39), and (41) respectively, where the elements of the matrices Σ_w and Σ_f appear directly. Furthermore, they are used for computing the expected value $\langle F w w^T F^T \rangle$ that appears in (49) as

$$\langle F w w^T F^T \rangle = \langle F \rangle \langle w w^T \rangle \langle F^T \rangle + \mathbf{E}_f \sum_{i,j} \langle w w^T \rangle_{ij}. \quad (58)$$

For the posterior image and weight means μ_f and μ_w , we do not use the above approximations, since we can exactly obtain them by solving the following linear systems:

$$\Sigma_f^{-1} \mu_f = \Phi^T \langle W \rangle^T \langle B \rangle g \quad (59)$$

$$\Sigma_w^{-1} \mu_w = \Phi^T \langle F \rangle^T \langle B \rangle g. \quad (60)$$

These linear systems are solved iteratively with the conjugate gradient method, using the approximation matrices $\tilde{\Sigma}_f$ and $\tilde{\Sigma}_w$ as preconditioners. In these iterations, products of circulant matrices are efficiently computed in the DFT domain, while products of diagonal matrices in the spatial domain. Specifically, each conjugate gradient iteration requires $O(M \log M)$ iterations. Theoretically, an exact solution of the linear system is obtained after $C = N$ iterations, however, we typically obtain a good approximation after only few iterations, e.g., $C = 20$. The overall computation cost is $O(CM \log M)$

D. Variational Optimization Algorithm

Each iteration of the optimization algorithm proceeds as follows. First we compute the parameters of the approximate posterior probabilities, as given in (32)–(41) and then we compute the expected values using (42)–(49). Finally, we may update the parameters of the noise and image prior distributions, using (50)–(53). The means of the posteriors $q(\mathbf{w})$ and $q(\mathbf{f})$ are used to obtain estimates of the PSF $\hat{\mathbf{h}}$ and the image $\hat{\mathbf{f}}: \hat{\mathbf{h}} = \Phi \mu_w$ and $\hat{\mathbf{f}} = \mu_f$.

IV. NUMERICAL EXPERIMENTS

Several numerical experiments have been carried out both with artificially generated observations where the ground truth is known and with real observations in order to demonstrate the properties of the proposed method. We compare the proposed method with previous Bayesian BID formulations based on Gaussian PSF and image models [15], with the TV-based blind deconvolution method in [4] and another recent variational Bayesian method in [16].

Hereafter, we will refer to the proposed method as the StStSt method, to imply that three Student's-t priors are used to model the PSF weights, the BID model errors and the image local differences. We also considered several simpler versions of this Bayesian model that use Gaussian distributions in place of the Student's-t distributions. Specifically, we consider Gaussian distributions for the PSF weights, $p(\mathbf{w}) = N(\mathbf{w}|0, \alpha^{-1} \mathbf{I})$, the additive noise, $p(\mathbf{n}) = N(\mathbf{n}|0, \beta^{-1} \mathbf{I})$, and the image local differences, $p(\mathbf{f}) = N(\mathbf{f}|0, (\gamma \mathbf{Q}^T \mathbf{Q})^{-1})$. The names of these simplified versions consist of three parts that express the distributions of the PSF weights, the additive noise and the image local differences. For example, the method that uses Gaussian distribution for the image local variances but Student's-t distributions for the PSF weights and noise is denoted as StStG.

The GGG is very similar to the VAR1 method described in [15], which also assumes that the PSF weights, the imaging model errors and the image local differences are Gaussian. The only difference between VAR1 and GGG is that VAR1 does not use a kernel model for the PSF, i.e., $\mathbf{h} = (w_1, \dots, w_M)^T$. Thus, the VAR1 method is identical to the GGG, when a Gaussian kernel of very small size is used.

In the simplified models GGG, GStG, StGG, and StStG, where Gaussian stationary image priors are used, we consider the typical simultaneously autoregressive (SAR) prior that has been used extensively in image restoration [15], [16]. This prior assumes a pdf for the image residuals $\epsilon(x, y)$ given by

$$\epsilon(x, y) = \sum_{(k,l) \in D(x,y)} (f(x, y) - f(k, l)) \quad (61)$$

where $D(x, y)$ is the set of four neighbors of (x, y) , given by $D(x, y) = \{(x+1, y), (x-1, y), (x, y-1), (x, y+1)\}$. The Bayesian method in [16] uses the SAR prior for both the image and PSF and then uses the variational methodology to achieve inference, similarly to the proposed method.

Furthermore, we provide a detailed comparison with the TV blind deconvolution method [4]. This method provides estimates of the image and PSF by solving the following minimization problem:

$$\min_{f,h} \frac{1}{2} \|h * f - g\|^2 + \alpha_f TV(f) + \alpha_h TV(h) \quad (62)$$

where $TV(x) = \int |\nabla x(z)| dz$ is a total variation regularization term.

A. Experiments on Artificially Blurred Images

In the first experiment, we compared all the methods using artificially degraded images. We generated a degraded image \mathbf{g} by blurring the true image \mathbf{f} with a known PSF \mathbf{h} and then adding Gaussian noise with variance $\sigma^2 = 10^{-6}$. The signal to noise ratio (SNR) of the observed image \mathbf{g} is $SNR = 10 \log_{10} \|\mathbf{f}\|^2 / M\sigma^2 = 45$ dB. In all methods, the initial PSF \mathbf{h}_{in} was set to a Gaussian-shaped function with variance $\sigma_{h_{in}}^2 = 3$. Since the true image is known, we can measure the quality of a recovered image $\hat{\mathbf{f}}$, by computing the improved signal to noise ratio $ISNR_f = 10 \log_{10} \|\mathbf{f} - \mathbf{g}\|^2 / \|\mathbf{f} - \hat{\mathbf{f}}\|^2$ which is a measure of the improvement of the quality of the estimated image with respect to the initial degraded image. We can also measure the quality of a PSF estimation $\hat{\mathbf{h}}$, by computing $ISNR_h = 10 \log_{10} \|\mathbf{h} - \mathbf{h}_{in}\|^2 / \|\mathbf{h} - \hat{\mathbf{h}}\|^2$.

The PSF that was used in this experiment was a 7×7 uniform, square-shaped PSF. However, we initialized the PSF as a Gaussian-shaped function with variance $\sigma_{h_{in}}^2 = 3$. The kernel function that we used was set to a Gaussian with variance $\sigma_\phi^2 = 0.1$, which is flexible enough to model the boundaries of the square. The ISNR values for the image and PSF estimates of all methods are shown in Table I. Furthermore, the degraded image and restored images for some of these methods are shown in Fig. 5 along with the restoration in [19], which was obtained by assuming that the PSF is known and a similar in spirit image prior.

Inspection of these results reveals that, in general, improvement in the accuracy of the estimated PSF implies improvement in the quality of the recovered image. Furthermore, using a Student's-t distribution to model the weights of the kernel model of the PSF gives significantly better PSF estimates as compared to using a Gaussian distribution for the same task. This demonstrates beyond any doubt the importance of this selection for

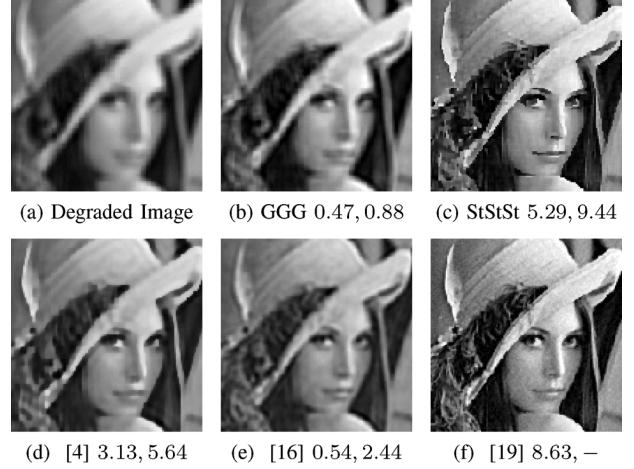


Fig. 5. Comparison of the proposed methods on the (a) Lena image degraded with a uniform, 7×7 square-shaped PSF. Estimated images using the (b) GGG method, (c) StStSt method, (d) method in [4], (e) method in [16], and (f) Known PSF restoration method in [19]. In all cases, the PSF was initialized as a Gaussian with $\sigma_{h_{in}}^2 = 3$ and the kernel was a Gaussian with variance $\sigma_\phi^2 = 0.1$. The numbers below each image are the ISNR values of the image ($ISNR_f$) and the corresponding PSF ($ISNR_h$).

TABLE I
ISNR FOR IMAGE AND PSF FOR THE EXPERIMENTS ON THE DEGRADED LENA IMAGE WITH A UNIFORM, 7×7 SQUARE-SHAPED PSF

Method	$ISNR_f$	$ISNR_h$
GGG	0.47	0.88
GGSt	0.58	0.79
GStG	0.05	1.53
GStSt	1.11	1.64
StGG	2.17	6.69
StGSt	5.87	8.12
StStG	5.57	10.91
StStSt	5.29	9.44
Method in [4]	3.13	5.64
Method in [16]	0.54	2.44
Known PSF in [19]	8.63	—

the BID problem. The image estimates are also improved when using Student's-t distributions for either the image local differences or noise. Finally, the StStSt model seems to produce visually more pleasing restored images with "sharper" edges than either the StGSt and StStG models, even though the $ISNR_f$ might be slightly lower. However, it is well known that $ISNR_f$ does not always capture accurately the human perception of image quality.

B. Comparison With Other BID Methods

In this subsection, we have conducted another experiment, where we compare the method based on the StStSt model with methods in [4] and [16]. In these experiments, we use the 256×256 "Cameraman" image, degraded with several PSFs and noise levels. Specifically, we used three different PSFs; a Gaussian-shaped PSFs with variance 5, a uniform square-shaped PSFs of size 7×7 and a rectangular nonsymmetric, accelerated motion blur [27] given by

$$h(x, y) = \begin{cases} (u_0^2 + 2a(x+s_x))^{-1/2}, & \text{if } |x| \leq s_x \text{ and } |y| \leq s_y \\ 0, & \text{otherwise} \end{cases}$$

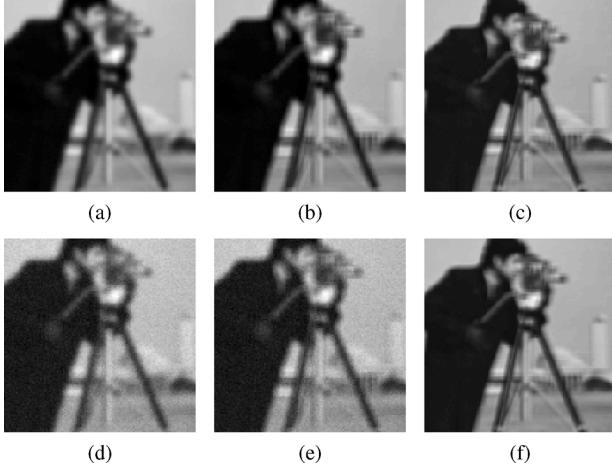


Fig. 6. Degraded cameraman images with (a)–(c) $SNR = 40$ dB and (d)–(f) $SNR = 20$ dB. The PSF was (a), (d) Gaussian-shaped with variance $\sigma_h^2 = 5$, (b), (e) uniform, square-shaped 7×7 , and (c), (f) accelerated motion blur.

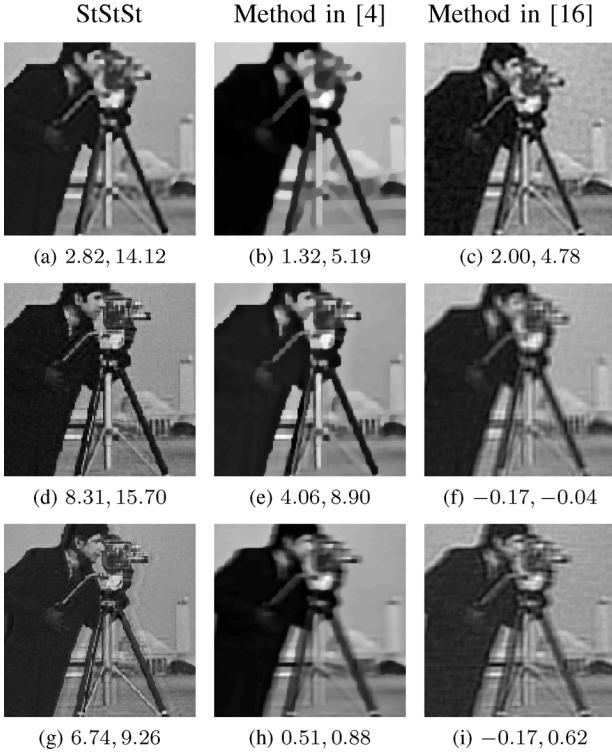


Fig. 7. Comparison on cameraman image with $SNR = 40$ dB and (a)–(c) Gaussian-shaped PSF with variance $\sigma_h^2 = 5$, (d)–(f) uniform, square-shaped 7×7 PSF, (g)–(i) motion-blur PSF. Estimates obtained with (b), (f), and (g): the proposed StStSt method; (c), (g), (k) method in [4]; and (d), (h), (l) method in [16]. The numbers below each image are the ISNR values of the image ($ISNR_f$) and the corresponding PSF ($ISNR_h$).

with $s_x = 4$, $s_y = 1$, $u_0 = 0.5$ and $a = 0.1$. We also used two levels of noise; low noise with $SNR = 40$ dB and high noise with $SNR = 20$ dB. The PSF was initialized as a Gaussian-shaped function with variance $\sigma_{h_{in}}^2 = 3$. For the StStSt method we used a Gaussian-shaped kernel function with variance $\sigma_\phi^2 = 2$, in all cases except for the case of accelerated motion PSF, where we used a Gaussian-shaped kernel with variance $\sigma_\phi^2 =$

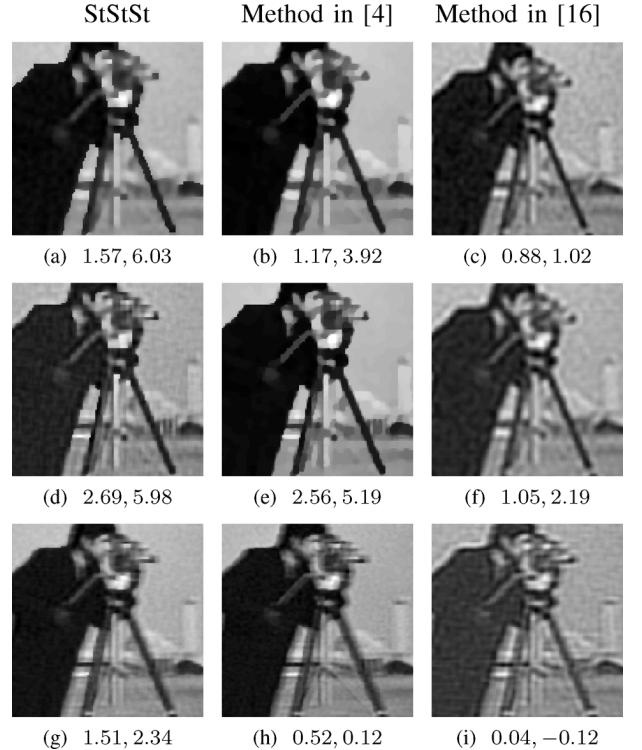


Fig. 8. Comparison on cameraman image with $SNR = 20$ dB and (a)–(c) Gaussian-shaped PSF with variance $\sigma_h^2 = 5$, (d)–(f) uniform, square-shaped 7×7 PSF, (g)–(i) motion-blur PSF. Estimates obtained with (b), (f), (g): the proposed StStSt method; (c), (g), (k) method in [4]; and (d), (h), (l) method in [16]. The numbers below each image are the ISNR values of the image ($ISNR_f$) and the corresponding PSF ($ISNR_h$).

1. The degraded images are shown in Fig. 6 and the restored images are shown in Figs. 7 and 8. The parameters of all the methods were selected in a trial and error manner in order to optimize the resulting images.

We can observe here that in all cases, the StStSt method outperforms both the methods in [4] and [16], especially in the case of low noise with $SNR = 40$ dB. Specifically, the method in [4] fails to estimate the Gaussian-shaped and motion PSFs, which is explained by the fact that the TV constraint on the PSF has the tendency to create flat areas and discontinuities, that are in contrast with the smooth PSFs that were used. In contrast, the method in [16] uses the SAR constraint for the PSF, which forces smooth PSF estimations. For this reason, it fails to estimate the square and motion blurs, which have discontinuities. Furthermore, the SAR model of the image fails to preserve the structure of edges. On the other hand, the StStSt method estimates successfully all types of PSF, because it uses distinct mechanisms to estimate the support and enforce smoothness to the PSF, via the sparse Student's-t prior and the kernel model, respectively.

In terms of computational cost, the method in [16] is the most efficient, since each iteration involves $O(M \log M)$ operations. On the other hand, each iteration of both the proposed method and the method in [4] require the solution of a $M \times M$ linear system that is solved using the conjugate gradient method and require $O(CM \log M)$ computations, where C is the number of conjugate gradient iterations.

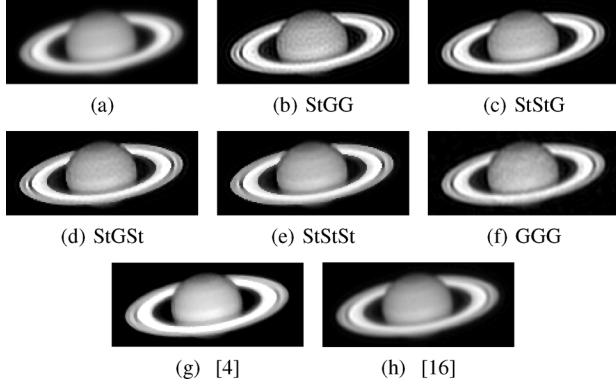


Fig. 9. (a) Degraded image. Estimated images using the methods (b) StGG, (c) StStG, (d) StGSt, (e) StStSt, (f) GGG, and the methods in (g) [4] and (h) [16]. The PSF was initialized as a Gaussian with $\sigma_{h,\text{in}}^2 = 3$ in all cases and the kernel was a Gaussian with variance $\sigma_\phi^2 = 1$.

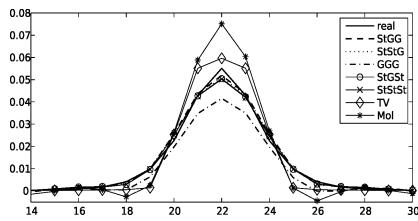


Fig. 10. One-dimensional slice of the true and estimated PSFs for the images of Fig. 9. The true PSF has been estimated as $h(r) \propto (1 + r^2/R^2)^{-\delta}$, with $\delta \approx 3$ and $R \approx 3.4$. The kernel was Gaussian with variance $\sigma_\phi^2 = 1$.

C. Experiments on Real Astronomical Images

We also applied the proposed methodology on a real astronomical image of the Saturn planet, which has previously been used in [16]. Astronomical measurements suggest the following PSF model for ground based telescopes

$$h(r) \propto \left(1 + \frac{r^2}{R^2}\right)^{-\delta}. \quad (63)$$

The parameters δ and R can be measured [16] and $\delta \approx 3$ and $R \approx 3.4$. The recovered images by the different methods are shown in Fig. 9 and the resulting PSFs in Fig. 10.

From these images, it is clear again that the models with two or more Student's-t priors give visually superior results. In these images, there is less ringing at the edges, noise in flat areas and the Saturn bands are better separated. Furthermore, the StStSt model produces again “sharper” images. It is interesting to notice that the StGG model does not yield good recovered images although it estimates well the measured PSF. This demonstrates the inappropriateness of the Gaussian to model the errors of the BID model and the image model. Notice also, that again, the TV-based methodology fails to estimate the smooth PSF and creates edges in areas where they do not exist in the original PSF; see Fig. 10.

D. Selecting the Kernel Width and Initial Values for the Parameters

The proposed method uses a sparse kernel model to estimate the PSF. The significance of the kernel model is that it favors

TABLE II
ISNR FOR IMAGE AND PSF FOR VARIOUS VALUES OF THE KERNEL WIDTH FOR THE CASE OF GAUSSIAN-SHAPED PSF WITH $\sigma_h^2 = 5$

Method	$\sigma_\phi^2 = 0.1$	$\sigma_\phi^2 = 1$	$\sigma_\phi^2 = 2$	$\sigma_\phi^2 = 3$	$\sigma_\phi^2 = 0.1$	$\sigma_\phi^2 = 1$	$\sigma_\phi^2 = 2$	$\sigma_\phi^2 = 3$
	$ISNR_f$	$ISNR_h$	$ISNR_f$	$ISNR_h$	$ISNR_f$	$ISNR_h$	$ISNR_f$	$ISNR_h$
GGG	1.62	-0.57	1.92	0.47	2.57	2.62	2.90	5.12
StGG	3.53	6.58	3.53	7.49	3.47	7.95	2.39	1.78
StStG	3.19	7.15	3.21	7.40	3.77	10.55	2.33	0.36
StGSt	3.69	8.86	3.96	10.33	4.24	12.30	1.55	2.88
StStSt	4.00	11.32	3.98	11.36	3.94	12.31	2.48	0.71

TABLE III
ISNR FOR IMAGE AND PSF FOR VARIOUS VALUES OF THE KERNEL WIDTH FOR THE CASE OF UNIFORM, 7×7 SQUARE-SHAPED PSF

Method	$\sigma_\phi^2 = 0.1$	$\sigma_\phi^2 = 1$	$\sigma_\phi^2 = 2$	$\sigma_\phi^2 = 3$	$\sigma_\phi^2 = 0.1$	$\sigma_\phi^2 = 1$	$\sigma_\phi^2 = 2$	$\sigma_\phi^2 = 3$
	$ISNR_f$	$ISNR_h$	$ISNR_f$	$ISNR_h$	$ISNR_f$	$ISNR_h$	$ISNR_f$	$ISNR_h$
GGG	0.70	-4.71	0.64	-3.41	0.12	-0.64	0.13	-3.87
StGG	2.17	6.69	1.20	9.09	-0.37	1.67	-2.31	-0.43
StStG	5.57	10.91	5.45	9.27	-0.29	1.87	-2.12	-0.20
StGSt	5.87	8.12	5.62	7.80	4.22	6.72	0.20	-0.12
StStSt	5.29	9.44	4.56	8.17	-0.51	2.01	-1.58	0.09

smooth estimations of the PSF, by forcing neighboring pixels to have similar values. This is important in order to enforce PSF smoothness and prevent the noise in the observed image to corrupt the PSF estimate. However, selecting an appropriate kernel is not straightforward. Here, we have considered several Gaussian kernels of different widths in order to determine how the proposed method is affected by the width of the Gaussian kernel. We have applied the proposed method on the artificially blurred images of the first experiment and considered degradation with Gaussian PSF or uniform-square shaped PSF. Tables II and III show the ISNRs of the image and PSF for several values of the kernel width, for the case where the true PSF is Gaussian-shaped and square-shaped, respectively. Notice that in all cases, selecting a very large kernel leads to very smooth estimates of the PSF that provide poor results. In case of uniform square true PSF (Table III) the best results are obtained when using a very small kernel. This is because the square PSF is not smooth at the edges of the rectangle. On the other hand, in the case of Gaussian-shaped true PSF (Table II), it is favorable to select a kernel that produces smooth PSF estimation.

It must be also noted that the performance of all the variational algorithms generally depends on the initialization of the parameters. This happens because the variational bound is a nonconvex function and, therefore, depending on the initialization, a different local maximum may be attained. In order to apply the proposed method, the following parameters have to be initialized:

1) *Weights w of the Kernel Model That Define the PSF:* In BID, having a good estimate of the PSF is usually very important and many BID methods fail when they are badly initialized. This is a significant limitation, because in many situations there is no available estimate of the PSF. The proposed method does

not rely on a good initial PSF estimation. Instead, the sparse kernel based PSF model, can successfully estimate the PSF from the observed image. This is demonstrated in the previous experiments, where we successfully estimated Gaussian-shaped, square-shaped and accelerated motion PSFs using an initial PSF that was Gaussian-shaped with variance $\sigma_{h_{in}}^2 = 3$. The weights w were initialized by solving the PSF model given in (6), which gives $w = \beta \Sigma_w \Phi^T h$ with $\Sigma_w = (\beta \Phi^T \Phi + \alpha I)^{-1}$.

2) *Weight Normalization Parameters α_i of the PSF Model and the Hyperparameters a^α, b^α :* Initially, we set all these parameters to very small values, e.g., $\alpha_i = 10^{-16}$, which corresponds to a very flexible linear model. This is desirable in order to obtain an initial estimate of the support of the PSF using all the available kernels. The hyperparameters a^α and b^α are set to zero, thus assuming a uninformative distribution for the parameters α . During inference, the parameters α_i for most kernels tend to infinity; thus, the support of the PSF is limited.

3) *Noise Precision β and the Hyperparameters a^β, b^β :* The noise precision β is initially set to $\beta = 10^3$. The hyperparameters a^β, b^β are initially set to values that define a Gamma distribution with mean 10^3 and variance 10^2 , which is a flat and rather uninformative distribution. Their values are then updated using (50) and (51).

4) *Strength of the Image Prior γ and the Hyperparameters a^γ, b^γ :* The parameter γ that defines the strength of the image prior is initially set to $\gamma = 10^2$. The hyperparameters a^γ and b^γ are set to values that define a Gamma distribution with mean 10^2 and variance 10^4 . Updating a^γ and b^γ (Section III-B), usually improves the performance of the algorithm, at least in the first few iterations. However, we have empirically found that at convergence, these hyperparameters attain very small values, thus defining an uninformative distribution. This leads to very noisy image estimates and for this reason we do not update the hyperparameters a^γ, b^γ but keep them fixed to their initial values. An explanation for the failure to estimate these parameters is that we use an improper prior for the image (16). Although selecting values for these parameters may seem arbitrary they actually depend on the characteristics of the image. Specifically, small values of the parameter b^γ lead to very smooth solutions, while small values of the parameter a^γ allow few hard edges by defining a heavy tailed distribution for the image local differences.

V. CONCLUSIONS AND FUTURE WORK

We presented a Bayesian approach to the BID problem where the PSF is modeled as a superposition of kernel functions. We assumed a suitable heavy tailed prior distribution on this kernel model, in order to obtain a sparse estimate of the support and shape of the PSF. We also used a heavy tailed pdf both for the noise, in order to achieve robustness to BID model errors and for the local image differences, in order to allow the reconstruction of edges. The Student's-t pdf was our choice as a heavy tailed pdf, due to its close relationship with the Gaussian. Because of the complexity of this model, the variational framework was used for approximate Bayesian inference.

Several experiments were carried out, to test the proposed methodology. These experiments indicated beyond doubt that the use of a Student's-t pdf to model the weights of the PSF

kernel-based model is crucial to the success of this approach. Furthermore, Bayesian BID models that use at least two Student's-t priors, one for the PSF, are clearly superior to BID models that use two or more Gaussian priors. Furthermore, it is interesting to notice that the StStSt model that uses only Student's-t priors seems to produce visually superior images as compared to models that use a combination of two Student's-t and Gaussian priors.

We also compared this methodology with TV-based and Bayesian as implemented in [16] BID in a number of different scenarios. From these comparisons it is clear that the proposed methodology is always superior to the Gaussian model based methodology in [16]. As far as TV-based BID is concerned the proposed method is clearly superior for scenarios with small sized PSFs and low noise. In the case of large PSFs and high noise the two methods produce different in nature results. The proposed methodology produces image where image details were better preserved. It also yields better ISNR values. However, it produces "ringing" artifacts in image edges. TV-based BID gave no "ringing," however, many image details were eliminated.

In the future, we plan to explore the possibility of learning the filters Q^k in a manner analogous to [25]. Furthermore, we plan to explore extending the *constrained variational* methodology in [24] to BID to avoid using the approximation of the partition function in (16).

APPENDIX A COMPUTATION OF THE POSTERIOR PARAMETER DISTRIBUTIONS

Using the mean field approximation (24), the posterior distribution of the parameters is given by (25) and (26). In order to find the posterior distribution $q(\mathbf{w})$, we start from (26) and compute only the terms of $I(\mathbf{w})$ that depend on \mathbf{w}

$$\begin{aligned} I(\mathbf{w}) &= \langle \ln p(\mathbf{g}|\boldsymbol{\beta}, \mathbf{w}, \mathbf{f}) p(\mathbf{w}|\boldsymbol{\alpha}) \rangle_{q(\mathbf{f}) q(\boldsymbol{\alpha}) q(\boldsymbol{\beta}) q(\boldsymbol{\gamma})} \\ &= -\frac{1}{2} \langle \mathbf{n}^T \mathbf{B} \mathbf{n} \rangle - \frac{1}{2} \sum_{i=1}^M \langle \alpha_i \rangle w_i^2 \\ &= -\frac{1}{2} \mathbf{w}^T \left(\Phi^T \langle \mathbf{F}^T \mathbf{B} \mathbf{F} \rangle \Phi + \langle \mathbf{A} \rangle \right) \mathbf{w} \\ &\quad - \mathbf{w}^T \Phi^T \langle \mathbf{F}^T \mathbf{B} \rangle \mathbf{g} + \text{const} \end{aligned}$$

where $\mathbf{n} = \mathbf{g} - \Phi \mathbf{w}$ and \mathbf{B} is diagonal and, therefore, symmetric. Then, from (25) we see that $q(\mathbf{w})$ is proportional to $\exp[-I(\mathbf{w})]$, and, thus, it is a Gaussian distribution given by (27). Similarly, we can obtain the posterior $q(\mathbf{f})$ which is also a Gaussian distribution given by (28).

The posterior $q(\boldsymbol{\alpha})$ is similarly obtained by computing the terms of $I(\boldsymbol{\alpha})$ that depend on $\boldsymbol{\alpha}$

$$\begin{aligned} I(\boldsymbol{\alpha}) &= \langle \ln p(\mathbf{w}|\boldsymbol{\alpha}) p(\boldsymbol{\alpha}) \rangle_{q(\mathbf{f}) q(\mathbf{w}) q(\boldsymbol{\beta}) q(\boldsymbol{\gamma})} \\ &= \frac{1}{2} \sum_{i=1}^M \ln \alpha_i - \sum_{i=1}^M \alpha_i \langle w_i^2 \rangle \\ &\quad + (a^\alpha - 1) \sum_{i=1}^M \ln \alpha_i - b^\alpha \sum_{i=1}^M \alpha_i \\ &= \left(a^\alpha - \frac{1}{2} \right) \sum_{i=1}^M \ln \alpha_i - \sum_{i=1}^M \left(\frac{1}{2} \langle w_i^2 \rangle + b^\alpha \right) + \text{const}. \end{aligned}$$

$$\begin{aligned}
\langle \ln p(\mathbf{g}|\mathbf{w}, \boldsymbol{\beta}, \mathbf{f}) \rangle &= -\frac{M}{2} \ln(2\pi) + \frac{1}{2} \sum_{i=1}^M \langle \ln \beta_i \rangle - \frac{1}{2} \langle (\mathbf{g} - \mathbf{F}\Phi\mathbf{w})^T \mathbf{B}(\mathbf{g} - \mathbf{F}\Phi\mathbf{w}) \rangle \\
\langle \ln p(\mathbf{w}|\boldsymbol{\alpha}) \rangle &= -\frac{M}{2} \ln(2\pi) + \frac{1}{2} \sum_{i=1}^M \langle \ln \alpha_i \rangle - \frac{1}{2} \sum_{i=1}^M \langle \alpha_i \rangle \langle w_i^2 \rangle \\
\langle \ln p(\mathbf{f}|\boldsymbol{\gamma}) \rangle &= -\frac{M}{2} \ln(2\pi) + \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^M \langle \ln \gamma_i^k \rangle - \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^M \langle \gamma_i^k \rangle \langle \mathbf{Q}^k \mathbf{f} \rangle_i^2 \\
\langle \ln p(\boldsymbol{\alpha}) \rangle &= Ma^\alpha \ln b^\alpha + (a^\alpha - 1) \sum_{i=1}^M \langle \ln \alpha_i \rangle - b^\alpha \sum_{i=1}^M \langle \alpha_i \rangle - M \ln \Gamma(a^\alpha) \\
\langle \ln p(\boldsymbol{\beta}) \rangle &= Ma^\beta \ln b^\beta + (a^\beta - 1) \sum_{i=1}^M \langle \ln \beta_i \rangle - b^\beta \sum_{i=1}^M \langle \beta_i \rangle - M \ln \Gamma(a^\beta) \\
\langle \ln p(\boldsymbol{\gamma}) \rangle &= MKa^\gamma \ln b^\gamma + (a^\gamma - 1) \sum_{k=1}^K \sum_{i=1}^M \langle \ln \gamma_i^k \rangle - b^\gamma \sum_{k=1}^K \sum_{i=1}^M \langle \gamma_i^k \rangle - MK \ln \Gamma(a^\gamma) \\
\langle \ln q(\mathbf{w}) \rangle &= -\frac{M}{2}(\ln(2\pi) + 1) - \frac{1}{2} \ln |\Sigma_w| \\
\langle \ln q(\mathbf{f}) \rangle &= -\frac{M}{2}(\ln(2\pi) + 1) - \frac{1}{2} \ln |\Sigma_f| \\
\langle \ln q(\boldsymbol{\alpha}) \rangle &= \sum_{i=1}^M [\tilde{a}^\alpha \ln \tilde{b}_i^\alpha + (\tilde{a}^\alpha - 1) \langle \ln \alpha_i \rangle - \tilde{b}_i^\alpha \langle \alpha_i \rangle - \ln \Gamma(\tilde{a}^\alpha)] \\
\langle \ln q(\boldsymbol{\beta}) \rangle &= \sum_{i=1}^M [\tilde{a}^\beta \ln \tilde{b}_i^\beta + (\tilde{a}^\beta - 1) \langle \ln \beta_i \rangle - \tilde{b}_i^\beta \langle \beta_i \rangle - \ln \Gamma(\tilde{a}^\beta)] \\
\langle \ln q(\boldsymbol{\gamma}) \rangle &= \sum_{k=1}^K \sum_{i=1}^M [\tilde{a}^\gamma \ln \tilde{b}_i^{\gamma^k} + (\tilde{a}^\gamma - 1) \langle \ln \gamma_i^k \rangle - \tilde{b}_i^{\gamma^k} \langle \gamma_i^k \rangle - \ln \Gamma(\tilde{a}^\gamma)]
\end{aligned}$$

This is the exponent of a Gamma distribution, and, therefore, $q(\boldsymbol{\alpha})$ is a Gamma distribution given by (29). The posterior distributions $q(\boldsymbol{\beta})$ and $q(\boldsymbol{\gamma})$ are also Gamma distributions given by (30) and (31) and their computation is very similar.

APPENDIX B COMPUTATION OF THE VARIATIONAL BOUND

The variational bound is given by

$$\begin{aligned}
\mathcal{L}(\theta) &= \langle \ln p(\mathbf{g}|\mathbf{w}, \boldsymbol{\beta}, \mathbf{f}) \rangle + \langle \ln p(\mathbf{w}|\boldsymbol{\alpha}) \rangle + \langle \ln p(\mathbf{f}|\boldsymbol{\gamma}) \rangle \\
&\quad + \langle \ln p(\boldsymbol{\alpha}) \rangle + \langle \ln p(\boldsymbol{\beta}) \rangle + \langle \ln p(\boldsymbol{\gamma}) \rangle - \langle \ln q(\mathbf{w}) \rangle \\
&\quad - \langle \ln q(\mathbf{f}) \rangle - \langle \ln q(\boldsymbol{\alpha}) \rangle - \langle \ln q(\boldsymbol{\beta}) \rangle - \langle \ln q(\boldsymbol{\gamma}) \rangle
\end{aligned}$$

and the required expected values can be computed as the equations shown at the top of the page.

REFERENCES

- [1] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Process. Mag.*, vol. 13, no. 3, pp. 43–64, Mar. 1996.
- [2] D. Kundur and D. Hatzinakos, "Blind image deconvolution revisited," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 61–63, Jun. 1996.
- [3] P. Campisi and K. Egiazarian, Eds., *Blind Image Deconvolution: Theory and Applications* CRC Boca Raton, FL, 2007.
- [4] T. F. Chan and C. K. Wong, "Total variation blind deconvolution," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 370–375, Mar. 1998.
- [5] T. F. Chan, S. Esedoglu, F. Park, and M. H. Yip, "Recent developments in total variation image restoration," in *Handbook of Mathematical Models in Computer Vision*. New York: Springer, 2005, vol. 3.
- [6] T. F. Chan, S. Esedoglu, F. Park, and M. H. Yip, *Image Processing and Analysis—Variational, PDE, Wavelet, and Stochastic Methods*. Philadelphia, PA: SIAM, 2005.
- [7] Y. L. You and M. Kaveh, "Blind image restoration by anisotropic regularization," *IEEE Trans. Image Process.*, vol. 8, no. 3, pp. 396–407, Mar. 1999.
- [8] K. H. Yap, L. Guan, and W. Liu, "A recursive soft-decision approach to blind image deconvolution," *IEEE Trans. Image Process.*, vol. 14, no. 5, pp. 624–633, May 2005.
- [9] L. Chen and K. H. Yap, "A soft double regularization approach to parametric blind image deconvolution," *IEEE Trans. Image Process.*, vol. 14, no. 5, pp. 624–633, May 2005.
- [10] M. M. Bronstein, A. M. Bronstein, M. Zibulevsky, and Y. Zeevi, "Blind image deconvolution of images using optimal sparse representations," *IEEE Trans. Image Process.*, vol. 14, no. 6, pp. 726–736, Jun. 2005.
- [11] B. D. Jeffs and J. C. Christou, D. Bonacini and R. K. Tyson, Eds., "Blind Bayesian restoration of adaptive optics telescope images using generalized Gaussian markov random field models," in *Proc. SPIE. Adaptive Optical System Technologies*, 1998, vol. 3353, pp. 1006–1013.
- [12] N. P. Galatsanos, V. Z. Mesarovic, R. Molina, A. K. Katsaggelos, and J. Mateos, "Hyperparameter estimation in image restoration problems with partially-known blurs," *Opt. Eng.*, vol. 41, no. 8, pp. 1845–1854, 2002.
- [13] J. Miskin and D. MacKay, *Advances in Independent Component Analysis*, M. Girolami, Ed. New York: Springer-Verlag.
- [14] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 787–794, 2006.

- [15] A. Likas and N. P. Galatsanos, "A variational approach for Bayesian blind image deconvolution," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2222–2233, Aug. 2004.
- [16] R. Molina, J. Mateos, and A. 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.
- [17] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
- [18] M. Jordan, Z. Gharamani, T. Jaakkola, and L. Saul, "An introduction to variational methods for graphical models," in *Learning in Graphical Models*, M. Jordan, Ed. Boston, MA: Kluwer, 1998, pp. 105–162.
- [19] G. Chantas, N. P. Galatsanos, and A. Likas, "Bayesian restoration using a new nonstationary edge preserving image prior," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2987–2997, Oct. 2006.
- [20] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.
- [21] M. A. T. Figueiredo, "Adaptive sparseness for supervised learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1150–1159, Sep. 2003.
- [22] C. Liu and D. Rubin, "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Statist. Sin.*, vol. 5, pp. 19–39, 1995.
- [23] D. Tzikas, A. Likas, and N. Galatsanos, "Large scale multikernel RVM for object detection," in *Proc. Hellenic Conf. Artificial Intelligence*, Heraklion, Greece, 2006, pp. 389–399.
- [24] G. Chantas, N. P. Galatsanos, and A. Likas, "Bayesian image restoration based on variational inference and a product of student-t priors," presented at the IEEE Int. Workshop MLSP, Thessaloniki, Greece, Aug. 2007.
- [25] M. Welling, G. E. Hinton, and S. Osindero, "Learning sparse topographic representations with products of student-t distributions," in *Proc. Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2003, vol. 15.
- [26] J. Bernardo and A. Smith, *Bayesian Theory*. New York: Wiley, 1994.
- [27] Y. Yitzhaky, I. Mor, A. Lantzman, and N. S. Kopeika, "Direct method for restoration of motion-blurred images," *J. Opt. Soc. Amer. A*, vol. 15, pp. 1512–1519, Jun. 1998.



Dimitris G. Tzikas received the B.Sc. degree in informatics and telecommunications from the University of Athens, Greece, in 2002, and the M.Sc. degree from the University of Ioannina, Greece, in 2004. He is currently pursuing the Ph.D. degree in the Department of Computer Science, University of Ioannina.

His research interests include machine learning, Bayesian methods, and statistical image processing.



Aristidis C. Likas (SM'03) received the Diploma degree in electrical engineering and the Ph.D. degree in electrical and computer engineering from the National Technical University of Athens, Greece.

Since 1996, he has been with the Department of Computer Science, University of Ioannina, Greece, where he is currently an Associate Professor. His research interests include machine learning, neural networks, statistical signal processing, and bioinformatics.



Nikolas P. Galatsanos (SM'95) received the Diploma of Electrical Engineering from the National Technical University of Athens, Greece, in 1982, and the M.S.E.E. and Ph.D. degrees from the Electrical and Computer Engineering Department, University of Wisconsin, Madison, in 1984 and 1989, respectively.

He was on the faculty of the Electrical and Computer Engineering Department, Illinois Institute of Technology, Chicago, from 1989 to 2002. Presently, he is with the Department of Computer Science, University of Ioannina, Ioannina, Greece. He coedited a book titled *Image Recovery Techniques for Image and Video Compression and Transmission* (Kluwer, 1998). His research interests center around image processing and statistical learning problems for medical imaging and visual communications applications.

Dr. Galatsanos has served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and the IEEE Signal Processing Magazine, and he currently serves as an Associate Editor for the *Journal of Electronic Imaging*.