ROBUST VARIATIONAL BAYESIAN KERNEL BASED BLIND IMAGE DECONVOLUTION

Dimitris Tzikas, Aristidis Likas and Nikolaos Galatsanos
Department of Computer Science, University of Ioannina, Ioannina, Greece

Keywords: Bayesian, Variational, Blind Deconvolution, Kernel Prior, Sparse Prior, Robust Prior, Student-t Prior.

Abstract: In this paper we present a new Bayesian model for the blind image deconvolution (BID) problem. The main novelties of this model are two. First, a sparse kernel based representation of the point spread function (PSF) that allows for the first time estimation of both PSF shape and support. Second, a non Gaussian heavy tail prior for the model noise to make it robust to large errors encountered in BID when little prior knowledge is available about both image and PSF. Sparseness and robustness are achieved by introducing Student-t priors both for the PSF and the noise. A Variational methodology is proposed to solve this Bayesian model. Numerical experiments are presented both with real and simulated data that demonstrate the advantages of this model as compared to previous Gaussian based ones.

1 INTRODUCTION

In blind image deconvolution (BID) both initial image and blurring point spread function (PSF), are unknown. Thus this problem is difficult, because the observed data are significantly fewer than the unknown quantities and do not specify them uniquely. For this reason, in order to resolve this ambiguity, prior knowledge (constraints) have to be used for both the image and the PSF.

BID is a problem with a long history. For a rather recent survey on this problem the reader is referred to (D.Kundur and D.Hatzinakos, 1996), (Kundur and Hatzinakos, 1996). Recently, constraints on the image and PSF have been expressed using the Bayesian methodology, by assuming the unknown quantities to be random variables, and assigning them suitable prior distributions that impose the desired characteristics (Jeffs and Christou, 1998) (Galatsanos et al., 2002). Unfortunately, because of the non-linearity of the data generation model, Bayesian inference using conventional methods, such as the Expectation Maximization (EM) algorithm, presents several computational difficulties, since the posterior distribution of the unknown parameters can not be computed in closed form. These difficulties have been overcome using the variational Bayesian methodology (Likas and Galatsanos, 2004) (Molina et al., 2006). In (Likas and Galatsanos, 2004) a non-stationary PSF model was used, while in (Molina et al., 2006) a hierarchical stationary simultaneously autoregressive PSF model was used. However, the PSF models described in both (Likas and Galatsanos, 2004) and (Molina et al., 2006) do not provide effective mechanisms to estimate, in addition to the shape, the support of the PSF. Furthermore, in previous works (Likas and Galatsanos, 2004), (Molina et al., 2006) the Bayesian models used assumed stationary Gaussian statistics for the errors in the imaging model. This is a serious shortcoming, since in the vicinity of edges the inaccurate initial estimates of the PSF that are usually available in BID, give large errors, which make the error pdf heavier tailed than Gaussian. For this reason, a few large errors can throw off the estimation algorithm when Gaussian statistics are used.

In this paper we propose a Bayesian methodology for the BID problem, which introduces two novelties to ameliorate the above mentioned shortcomings of previous Bayesian methods. First, we introduce a Bayesian model that has the ability to estimate both PSF support and shape. More specifically, a sparse kernel based prior is used for the unknown PSF, in

Tzikas D., Likas A. and Galatsanos N. (2007).
ROBUST VARIATIONAL BAYESIAN KERNEL BASED BLIND IMAGE DECONVOLUTION.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications*, pages 143-150
DOI: 10.5220/0002063601430150

a similar manner as for the Relevance Vector Machine (RVM) (Tipping, 2001)(Tipping and Lawrence, 2003). This prior prunes out kernels that do not fit the data. Second, this model is made robust to large errors of the imaging model. This is achieved by assuming the errors to be non Gaussian distributed and are modeled by a pdf with heavier tails. The Student-t pdf is used to model both PSF and image model errors. This pdf can be viewed as an infinite mixture of Gaussians with different variances (Bishop, 2006) and provides both sparse models and robust representations of large errors (Peel and Mclachlan, 2000), (Tipping and Lawrence, 2003).

Since the proposed Bayesian model cannot be solved exactly we resort to the variational approximation. This approximation methodology (Jordan et al., 1998) considers a class of approximate posterior distributions and then searches to find the best approximation of the true posterior within this class. This methodology has been used in many Bayesian inference problems with success.

The rest of this paper is organised as follows. In section 2 we explain in detail the proposed model. Then in section 3 we present a brief introduction of variational methods and in section 4 we aply the variational methodology to infer the proposed model. In section 5 we present experiments, first on artificially blurred images and then on real astronimical images. Finally, in section 6 we conclude and provide directions for future work.

2 STOCHASTIC MODEL

We assume that the observed image g has been generated by convolving an unknown image f with an also unknown PSF h and then adding independent Gaussian distributed noise n, with inverse variance β:

$$g = f * h + n. \quad (1)$$

Here, g, f, h and n are $N \times 1$ vectors of the intensities of the degraded image, observed image, blurring PSF and additive noise respectively, in lexicographical order and $*$ denotes two-dimensional circular convolution between the images.

The blind deconvolution problem is very difficult because there are too many unknown parameters that have to be estimated. In fact, the number of unknown parameters h, f is twice the number of observations 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 random variables and prior knowledge is expressed by assuming that they have been sampled from specific prior distributions.

2.1 PSF Model

We model the PSF as a linear combination of a fixed set of kernel basis functions and specifically there is one kernel function $K(x)$ centered at each pixel of the image. This kernel function is then evaluated at all the pixels of the image to give the $N \times 1$ basis vector ϕ. We denote with Φ the $N \times N$ matrix $\Phi = (\phi_1, \ldots, \phi_N)$, which is the block-circulant matrix whose first column is $\phi_1 = \phi$, so that $\Phi w = \phi * w$. Each column ϕ_i can also be considered as the kernel function shifted at the corresponding pixel $\phi_i = K(x - x_i)$. The PSF h is then modeled as:

$$h = \sum_{i=1}^{N} w_i \phi_i = \Phi w. \quad (2)$$

Thus, the data generation model (1) can be written as:

$$g = (\Phi w) * f + n = F\Phi w + n = \Phi W f + n. \quad (3)$$

Matrices F, W are defined similarly with matrix Φ, and are block-circulant matrices generated by f and w respectively, so that $Fw = f * w$ and $Wf = w * f$.

In this paper Gaussian kernel functions are considered, which produce 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, with small additional computational cost (Tzikas et al., 2006a).

A hierarchical prior that enforces sparsity is then imposed on the weigths w:

$$p(w|\alpha) = \prod_{i=1}^{N} N(w_i|0, \alpha_i^{-1}). \quad (4)$$

Each weight is assigned a separate inverse variance parameter α_i, which is treated as a random variable that follows a Gamma distribution:

$$p(\alpha) = \prod_{i=1}^{N} \Gamma(\alpha_i|a^\alpha, b^\alpha). \quad (5)$$

This two level hierarchical prior is equivalent with a Student-t prior distribution. This can be realized by integrating out the parameters α_i to compute the prior weight distribution $p(w)$:

$$p(w) = \int p(w|\alpha)p(\alpha)d\alpha = St(w|0, \frac{b^\alpha}{a^\alpha}, 2a^\alpha), \quad (6)$$

where $St(w|0, \frac{a^\alpha}{b^\alpha}, 2a^\alpha)$ denotes a zero mean Student t distribution with variance $\frac{b^\alpha}{a^\alpha}$ and $2a^\alpha$ degrees of freedom (Bishop, 2006).

Setting $a^{\alpha} = b^{\alpha} = 0$ defines an uninformative Gamma hyperprior, which corresponds to a Student t distribution for the weights w with heavy tails. Most probability mass of this distribution is concentrated in the axes of origin and among the axes of definition. For this reason, during model learning most of the weights are set to zero and the corresponding parameters α tend to infinity. Thus, the corresponding basis functions are pruned from the model, in a manner similar to the RVM model (Tipping, 2001)(Tipping and Lawrence, 2003). The importance of a sparse model is that a very wide PSF can be initially considered, e.g. by placing one kernel at each image pixel, and those kernels that do not fit the true PSF should be pruned automatically during learning. This provides a robust methodology of estimating the PSF shape and support.

2.2 Image Model

We assume a simultaneously autoregressive (SAR) model for the image:

$$p(f) = N(f|0,(\gamma Q^T Q)^{-1}), \tag{7}$$

where Q is the Laplacian operator. This model, penalizes large differences in neighbouring pixels, as can be seen by the equivalent:

$$\varepsilon = Qf \sim N(0,\gamma^{-1}), \tag{8}$$

or

$$f(x,y) = \frac{1}{4}\sum_{(k,l)\in N} f(x+k,y+l)+\varepsilon(x,y), \tag{9}$$

where $\varepsilon \sim N(0,\gamma^{-1}I)$ and $N = \{(-1,0),(1,0),(0,-1),(0,1)\}$. The variance parameter γ is assigned a Gamma distribution:

$$p(\gamma) = \Gamma(\gamma|a^{\gamma},b^{\gamma}). \tag{10}$$

We set $a^{\gamma} = b^{\gamma} = 0$ in order to obtain an uninformative Gamma prior. Since there is only one random variable γ and N observations we can efficiently estimate it without any prior knowledge.

2.3 Noise Model

The noise is assumed to be zero-mean Gaussian distributed, given by:

$$p(n|\beta) = \prod_{i=1}^{N} N(n_i|0,\beta_i^{-1}). \tag{11}$$

The parameters β_i that define the variance of the noise at each pixel, are also assumed to be random variables and they are assigned a Gamma distribution:

$$p(\beta) = \prod_{i=1}^{N} \Gamma(\beta_i|a^{\beta},b^{\beta}). \tag{12}$$

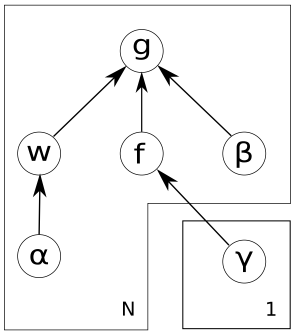


Figure 1: The graphical model that describes the dependencies between the random variables of the proposed model.

We choose values for the parameters $a^{\beta} = 10^3$ and $b^{\beta} = 10^{-3}$. This leads to a rather uninformative distribution for the noise variance, with mean value 10^{-6}.

This two level hierarchical prior is equivalent with a Student-t prior distribution, in a similar manner as in (5). The Student-t distribution is very flexible and can have heavier tails than the Gaussian distribution. Thus, it is used to achieve robust estimation. In BID this is important because given an incorrect estimation of the PSF, the distribution of the error is heavy tailed.

The relationships between the random variables that define the stochastic model are represented by the graphical model in fig. 1. Because of the complexity of the model, the posterior distribution of the parameters $p(w,f,\alpha,\beta,\gamma|g)$ cannot be computed and conventional inference methods, such as maximum likelihood via the EM algorithm, can not be applied. Instead, we resort to approximate inference methods and specifically to the variational Bayesian inference methodology.

3 THE VARIATIONAL METHODOLOGY FOR BAYESIAN INFERENCE

A probabilistic model consists of a set of observed random variables D and a set of hidden random variables $\theta = \{\theta^i\}$. Inference in such models requires the computation of the posterior distribution of the hidden variables $p(\theta|D)$, which is usually intractable. The variational methodology(Jordan et al., 1998) is an approximate inference methodology, which considers a family of approximate posterior distributions $q(\theta)$, and then seeks values for the parameters θ that best approximate the true posterior $p(\theta|D)$.

The evidence of the model $p(D) = \int P(D,\theta)d\theta$

can be decomposed as:

$$\ln p(D) = \mathcal{L}(\theta) + KL(q(\theta)\|p(\theta|D)), \quad (13)$$

where

$$\mathcal{L}(\theta) = \int q(\theta) \ln \frac{p(D,\theta)}{q(\theta)} d\theta \quad (14)$$

is called the variational bound and

$$KL(q(\theta)\|p(\theta|D)) = -\int q(\theta) \ln \frac{p(\theta|D)}{q(\theta)} d\theta \quad (15)$$

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

$$\theta = \underset{q(\theta)}{argmax}\, \mathcal{L}(\theta) = \underset{q(\theta)}{argmin}\, KL(q(\theta)\|p(\theta|D)) \quad (16)$$

In order to be able to perform the maximization of the variational bound with respect to the approximating distribution $q(\theta)$, 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, where we assume that the posterior distributions of the hidden variables are independent, and thus:

$$q(\theta) = \prod_i q(\theta^i). \quad (17)$$

Then, the variational bound is maximized by (Jordan et al., 1998):

$$q(\theta^i) = \frac{\exp[I(\theta^i)]}{\int \exp[I(\theta^i)] d\theta^i}, \quad (18)$$

where

$$I(\theta^i) = \langle \ln p(D,\theta) \rangle_{q(\theta^{\setminus i})} = \int q(\theta^{\setminus i}) \ln p(D,\theta) d\theta^{\setminus i}. \quad (19)$$

and $\theta^{\setminus i}$ denotes the vector of all hidden variables except θ^i.

Computation of $q(\theta^i)$ is not straightforward, since $I(\theta^i)$ depends on the approximate distribution $q(\theta^{\setminus i})$. Variational inference proceeds by assuming some initial parameters θ_0 and iteratively updating $q(\theta^i)$ using 18.

4 VARIATIONAL BLIND DECONVOLUTION ALGORITHM

In this section we apply the variational methodology to the stochastic BID image model we described in section 2. The observed variable of the model is g and the hidden variables are $\theta = (w, f, \alpha, \beta, \gamma)$. The approximate posterior distributions of the hidden variables can be computed from (18), as:

$$\begin{aligned}
q(w) &= N(w|\mu_w, \Sigma_w), && (20)\\
q(f) &= N(f|\mu_f, \Sigma_f), && (21)\\
q(\alpha) &= \prod_i \Gamma(\alpha_i|\tilde{a}^\alpha, \tilde{b}_i^\alpha), && (22)\\
q(\beta) &= \prod_i \Gamma(\beta_i|\tilde{a}^\beta, \tilde{b}_i^\beta), && (23)\\
q(\gamma) &= \Gamma(\gamma|\tilde{a}^\gamma, \tilde{b}^\gamma), && (24)
\end{aligned}$$

where

$$\begin{aligned}
\mu_w &= \Sigma_w \Phi^T \langle F \rangle^T \langle B \rangle g, && (25)\\
\Sigma_w &= \left(\Phi^T \langle F^T \langle B \rangle F \rangle \Phi + diag\{\langle \alpha \rangle\}\right)^{-1}, && (26)\\
\mu_f &= \Sigma_f \Phi^T \langle W \rangle^T \langle B \rangle g, && (27)\\
\Sigma_f &= \left(\Phi^T \langle W^T \langle B \rangle W \rangle \Phi + \langle \gamma \rangle Q^T Q\right)^{-1}, && (28)\\
\tilde{a}^\alpha &= a^\alpha + 1/2, && (29)\\
\tilde{b}_i^\alpha &= b^\alpha + \frac{1}{2}\langle w_i^2 \rangle, && (30)\\
\tilde{a}^\beta &= a^\beta + N/2, && (31)\\
\tilde{b}_i^\beta &= b^\beta + \frac{1}{2}\langle nn^T \rangle_{ii}, && (32)\\
\tilde{a}^\gamma &= a^\gamma + N/2, && (33)\\
\tilde{b}^\gamma &= b^\gamma + \frac{1}{2} trace\{Q^T Q \langle ff^T \rangle\}. && (34)
\end{aligned}$$

The required expected values are evaluated as:

$$\begin{aligned}
\langle w \rangle &= \mu_w && (35)\\
\langle w_i^2 \rangle &= \mu_{w_i}^2 + \Sigma_{w_{ii}} && (36)\\
\langle W^T W \rangle &= U^{-1} \langle \Lambda_w^* \Lambda_w \rangle U && (37)\\
\langle f \rangle &= \mu_f && (38)\\
\langle ff^T \rangle &= \mu_f \mu_f^T + \Sigma_f && (39)\\
\langle F^T F \rangle &= U^{-1} \langle \Lambda_f^* \Lambda_f \rangle U && (40)\\
\langle \alpha_i \rangle &= \tilde{a}^\alpha / \tilde{b}_i^\alpha && (41)\\
\langle \beta_i \rangle &= \tilde{a}^\beta / \tilde{b}_i^\beta && (42)\\
\langle \gamma \rangle &= \tilde{a}^\gamma / \tilde{b}^\gamma && (43)
\end{aligned}$$

$$\langle nn^T \rangle = gg^T - 2g(\langle F \rangle \Phi \langle w \rangle)^T + \Phi \langle Fww^T F^T \rangle \Phi^T \quad (44)$$

$$\langle Fww^T F^T \rangle \approx \langle F \rangle \langle ww^T \rangle \langle F^T \rangle + \bar{\Sigma}_f \sum_{i,j} \langle ww^T \rangle_{ij} \quad (45)$$

where $\bar{\Sigma}_f = \frac{1}{N}\sum_{i=1}^N (\Sigma_f)_{ii}$, U is the DFT matrix such that Ux is the DFT of x, $\Lambda_w = diag\{\lambda_{w_1} \ldots \lambda_{w_N}\}$ and $\Lambda_f = diag\{\lambda_{f_1} \ldots \lambda_{f_N}\}$ are diagonal matrices with the eigenvalues of W and F respectively, and

$$\begin{aligned}
\langle \lambda_{w_i}^* \lambda_{w_i} \rangle &= (\mu_w * \mu_w)_i + \sum_k \Sigma_{w_{k,(k-i)}}, && (46)\\
\langle \lambda_{f_i}^* \lambda_{f_i} \rangle &= (\mu_f * \mu_f)_i + N\Sigma_{f_i}. && (47)
\end{aligned}$$

Table 1: ISNR of the image and PSF.

PSF	Method	$ISNR_f$	$ISNR_h$
Gaussian	RVMBID	3.44	7.9
	rRVMBID	3.71	10.24
	VAR1	2.2	0.2
square	RVMBID	2.59	9.14
	rRVMBID	5.55	9.86
	VAR1	0.04	-1.4

Notice that computation of matrices Σ_f and Σ_w involves inverting $N \times N$ matrices, which requires $O(N^3)$ time. Instead we approximate Σ_f with a circulant matrix and Σ_w with a diagonal matrix.

When computing the posterior image and weight mean μ_f and μ_w, we do not use the above approximations, since we can obtain these by solving the following linear systems:

$$\Sigma_f^{-1}\mu_f = \Phi^T \langle W \rangle^T \langle B \rangle g, \qquad (48)$$

$$\Sigma_w^{-1}\mu_w = \Phi^T \langle F \rangle^T \langle B \rangle g. \qquad (49)$$

These linear systems are solved efficiently, using the conjugate gradient method.

The parameters a^β and b^β of the noise Gamma hyperprior can be estimated by optimizing the variational bound $\mathcal{L}$ given by (14). We compute its derivatives with respect to the parameters a^β and b^β:

$$\frac{\theta \mathcal{L}}{\theta a^\beta} = N \log d - N\psi(a^\beta) + \sum_{n=1}^{N} \langle \log \beta_n \rangle, \qquad (50)$$

$$\frac{\theta \mathcal{L}}{\theta b^\beta} = N\frac{a^\beta}{b^\beta} - \sum_{n=1}^{N} \langle \beta_n \rangle, \qquad (51)$$

where $\psi(x)$ is the digamma function. Then, we can update these parameters by setting the above derivatives to zero. This cannot be done analytically for the parameter a^β and we use a numerical method instead.

Each iteration of the optimization algorithm proceeds as follows. First we compute the approximate posterior probabilities, as given in (25) to (34) and then we compute the expected values in (35) to (47). Finally, we update the parameters of the noise prior distribution, by solving the equations in (50) and (51).

5 NUMERICAL EXPERIMENTS

5.1 Experiments on Artificially Blurred Images

Several experiments have been carried out, in order to demostrate the practical use of the proposed method. First we demonstrate the effectiveness of the proposed method on artificially blurred images. We generate a degraded image by blurring the true image with some PSF h and then adding Gaussian noise with variance $\sigma^2 = 10^{-6}$. We consider two cases for the PSF, a Gaussian function with variance $\sigma_h^2 = 5$ and a 7×7 square-shaped function. Since the true image is known we can measure the performance of the method by computing the improved signal to noise ratio, $ISNR_f = 10 \log \frac{\|f-g\|^2}{\|f-\hat{f}\|^2}$, which is a measure of the improvement of the quality of the estimated image generated by the algorithm with respect to the initial degraded image. We also measure the improvement on the PSF with respect to the PSF that was used to initialize the algorithm, by computing $ISNR_h = 10 \log \frac{\|h-h_{in}\|^2}{\|h-\hat{h}\|^2}$.

We present the results of three different variational methods. The first is the method that we described in this paper and we will call it rRVMBID. The second is a very similar but simplified method, which assumes that the noise is Gaussian distributed (Tzikas et al., 2006b) and we will call it RVMBID. The last is a variational method that is based on a much simpler model for the PSF (Likas and Galatsanos, 2004) and we will call it VAR1.

In the first artificial experiment the blurring PSF was set to a Gaussian function, with variance $\sigma_h^2 = 5$. We initialized all methods, using a Gaussian PSF with variance $\sigma_{h_in}^2 = 3$. The kernel function that was used by the kernel based methods was again a Gaussian function with variance $\sigma_\phi^2 = 2$. The estimated images of the compared algorithms are shown in fig. 2 and the estimated PSFs in fig. 3. The corresponding ISNRs are shown in table 1.

In the next artificial experiment we considered a 7×7 square-shaped PSF. This type of PSF is very difficult to estimate because of the discontinuities at the edges of the rectangle. We again initialize the PSF as a Gaussian shaped function with variance $\sigma_{h_in}^2 = 3$. The kernel function was set to a Gaussian with variance $\sigma_\phi^2 = 1$ in order to be flexible enough to model the boundaries of the square. The estimated images of the compared algorithms are shown in fig. 4 and the PSFs in fig. 5. The corresponding ISNRs are shown in table 1.

5.2 Experiments on Real Astronomical Images

We also applied the methodology on a real astronomical image of the Saturn planet, which has previously been used in (Molina et al., 2006). Previous studies have suggested the following symmetric approximation for the PSF of images taken from ground based

Figure 2: Degraded image (a) generated with a Gaussian PSF with $\sigma_h^2 = 5$. Estimated image of the (b)RVMBID, (c) rRVMBID and (d) VAR1 algorithms. The PSF was initialized as a Gaussian with $\sigma_{hin}^2 = 3$ in all cases.

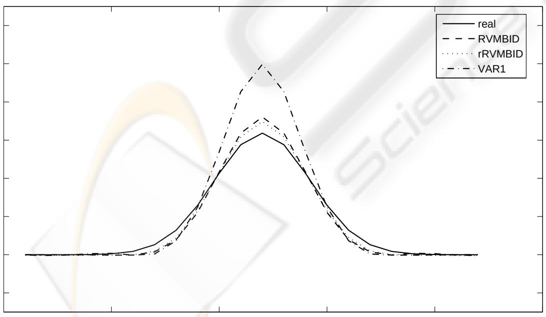


Figure 3: True PSF and PSF estimations for the case of Gaussian PSF with $\sigma_h^2 = 5$.

telescopes:

$$h(r) \propto (1 + \frac{r^2}{R^2})^{-\delta} \tag{52}$$

The parameters δ and R can be estimated (Molina et al., 2006) and are $\delta \approx 3$ and $R \approx 3.4$. The image estimations of the RVMBID, rRVMBID and VAR1 algorithms are shown in fig. 6 and the PSFs in fig. 7.

The performance of all the variational algorithms generally depends on the initialization of the parameters. This happens because the variational bound is a non-convex function and therefore depends on the initialization a different local maximum may be achieved. In all the above experiments, we sought the initialization that gave the best results.

6 CONCLUSIONS

We presented a Bayesian treatment of the BID problem in which the PSF was modeled as a superposition of kernel functions. We then applied a sparse prior distribution on this kernel model in order to estimate the support and shape of the PSF. Furthermore, we assumed a heavy tailed pdf for the noise in or-

(a) (b)

(c) (d)

Figure 4: Degraded image (a) generated with a rectangular 7×7 square PSF (e). Estimated image of the (b) RVMBID, (c) rRVMBID and (d) VAR1 algorithms. The PSF was initialized as a Gaussian with $\sigma^2_{hin} = 3$ in all cases.

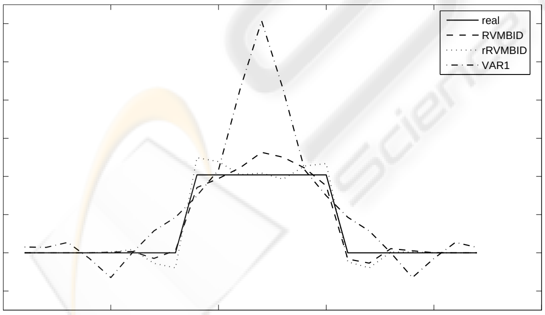


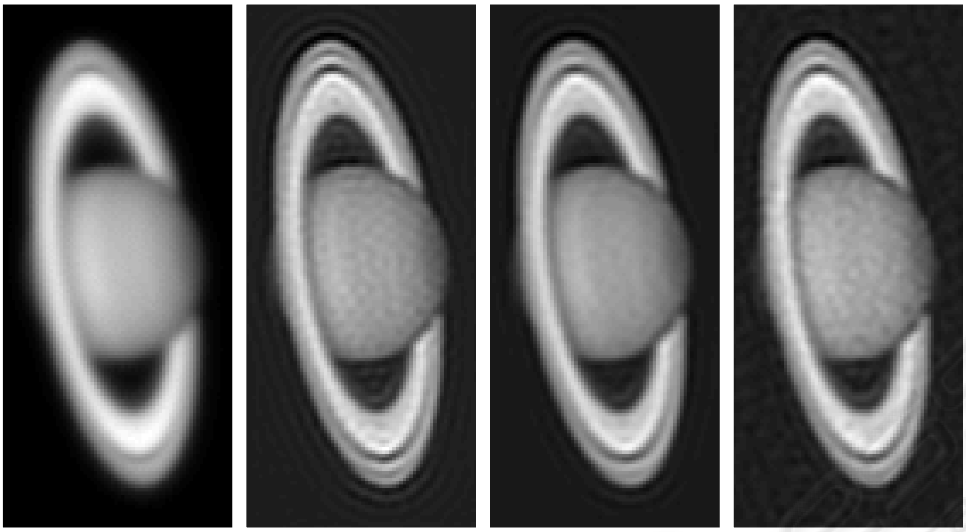
Figure 5: True PSF and PSF estimations for the case of rectangular 7×7 square PSF.

der to achieve robustness. Because of the complexity of the model, we used the variational framework to achieve inference. Several experiments have been carried out, that demonstrate the superior performance of the method with respect to another variational approach proposed in (Likas and Galatsanos, 2004).

An improvement to the proposed method would be to allow many different types of kernels at each pixel. Thus, one could consider, for example, both rectangular and Gaussian kernels and the best one depending on the true PSF would be selected automatically. Another interesting enhancement to the method would be to consider a non-stationary prior model for the image, which would contain a different γ_i parameter for each pixel. This image prior, would model better edge and textured area, however, there are several computational difficulties to be overcome for its implementation.

ACKNOWLEDGEMENTS

This research was co-funded by the program 'Pythagoras II' of the 'Operational Program for Edu-



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

Figure 6: Degraded image (a). Estimated image the (b) RVMBID, (c) rRVMBID and (d) VAR1 algorithms. The PSF was initialized as a Gaussian with $\sigma^2_{hin} = 3$ in all cases.

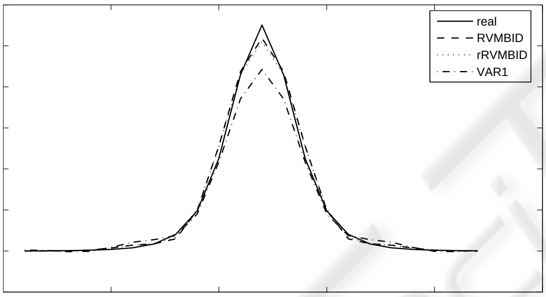


Figure 7: True PSF and PSF estimations for the real saturn image.

cation and Initial Vocational Training' of the Hellenic Ministry of Education.

REFERENCES

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

D.Kundur and D.Hatzinakos (1996). Blind image deconvolution. *IEEE Signal Processing Mag.*, 13(3):43–64.

Galatsanos, N. P., Mesarovic, V. Z., Molina, R., Katsaggelos, A. K., and Mateos, J. (2002). Hyperparameter estimation in image restoration problems with partially-known blurs. *Optical Eng.*, 41(8):1845–1854.

Jeffs, B. D. and Christou, J. C. (1998). Blind Bayesian restoration of adaptive optics telescope images using generalized Gaussian markov random field models. In Bonaccini, D. and Tyson, R. K., editors, *in Proc. SPIE. Adaptive Optical System Technologies*, volume 3353, pages 1006–1013.

Jordan, M., Gharamani, Z., Jaakkola, T., and Saul, L. (1998). An introduction to variational methods for graphical models. In Jordan, M., editor, *Learning in Graphical Models*, pages 105–162. Kluwer.

Kundur, D. and Hatzinakos, D. (1996). Blind image deconvolution revisited. *IEEE Signal Processing Mag.*, 13(6):61–63.

Likas, A. and Galatsanos, N. P. (2004). A variational approach for Bayesian blind image deconvolution. *IEEE Trans. on Signal Processing*, 52:2222–2233.

Molina, R., Mateos, J., and Katsaggelos, A. (2006). Blind deconvolution using a variational approach to parameter, image, and blur estimation. *IEEE Trans. on Image Processing, to appear*.

Peel, D. and Mclachlan, G. J. (2000). Robust mixture modelling using the t distribution. *Statistics and Computing*, 10(4):339–348.

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244.

Tipping, M. E. and Lawrence, N. D. (2003). A variational approach to robust Bayesian interpolation. In *IEEE Workshop on Neural Networks for Signal Processing*, pages 229–238.

Tzikas, D., Likas, A., and Galatsanos, N. (2006a). Large scale multikernel RVM for object detection. In *SETN*, pages 389–399.

Tzikas, D., Likas, A., and Galatsanos, N. (2006b). Variational Bayesian blind image deconvolution based on a sparse kernel model for the point spread function. In *in Proc. EUSIPCO*.