14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

VARIATIONAL BAYESIAN BLIND IMAGE DECONVOLUTION BASED ON A SPARSE KERNEL MODEL FOR THE POINT SPREAD FUNCTION

Dimitris Tzikas, Aristidis Likas, and Nikolas Galatsanos

Department of Computer Science, University of Ioannina
45110 Ioannina, Greece

ABSTRACT

In this paper we propose a variational Bayesian algorithm for the blind image deconvolution problem. The unknown point spread function (PSF) is modeled as a sparse linear combination of kernel basis functions. This model offers an effective mechanism to estimate for the first time both the support and the shape of the PSF. Numerical experiments demonstrate the effectiveness of the proposed methodology.

1. INTRODUCTION

Image acquisition techniques often obtain blurred images, which can be modeled as the convolution of the initial image with some blurring point spread function (PSF) and the addition of some noise source. The blind image deconvolution (BID) problem is the problem of estimating the initial image and the blurring PSF, given only the observed image. This is a very difficult problem, because the observed data are significantly less than the unknown quantities. Indeed, methods, such as maximum likelihood, that depend only on the observed data, lead to estimates that are governed by great variance and are not usefull. Instead, most BID algorithms seek biased estimates, by assuming constraints on the image and PSF, in order to reduce the variance of the estimates.

Most existing methods apply heavy constraints on the PSF, such as assuming specific support size, symmetric shape or similarity to an existing estimate of the PSF. Such constraints can be elegantly represented under the Bayesian framework by assuming prior distributions on the unknown parameters. However, because of the complexity 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. Recently, the Bayesian framework has been applied to the BID problem by modelling the PSF as Gaussian random variable and using the variational methodology to achieve approximate inference[1][2]. In [1] a non-stationary PSF model was used, while in [2] a hierarchical stationary simultaneously autoregressive PSF model was used. However, the PSF models described in both [1] and [2] do not provide effective mechanisms to estimate, in addition to the shape, the support of the PSF.

The variational methodology[3] relies on considering a class of approximate posterior distributions and then searching to find the best approximation of the true posterior in this class. This procedure can be accomplished without computing the true posterior distribution, and thus computations can be simplified by considering an appropriate class of approximate posteriors. The variational methodology has been successfully applied to many other complex Bayesian models and although it does not benefit from proved convergence like the EM algorithm, usually this is not an issue.

In this paper we propose an alternative Bayesian method for the BID problem, which successfully estimates the PSF support and shape. The unknown PSF is modeled as a sparse linear combination of kernel functions, similarly to the Relevance Vector Machine (RVM) model [4][5]. Initially, the PSF is modeled as the superposition of one kernel at each pixel. Even though in the PSF model there are as many parameters as pixels, estimation of the PSF is very robust, because the sparse prior distribution that is imposed on the kernel weights, prunes kernels that do not fit the true PSF. The proposed algorithm does not assume any constraints on the PSF support and shape.

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 ϵ, with inverse variance β:

$$g = h * f + \epsilon, \tag{1}$$

with

$$\epsilon \sim N(\epsilon|0, \beta^{-1} I). \tag{2}$$

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 robust 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.

The PSF h is modeled as a linear combination of a fixed set of basis functions $\{\phi_i\}_{i=1}^N$:

$$h = \Phi w, \tag{3}$$

where $\Phi = (\phi_1, \ldots, \phi_N)$.

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

$$g = (\Phi w) * f + \epsilon = F\Phi w + \epsilon = \Phi W f + \epsilon. \tag{4}$$

Matrices F, W are circulant matrices whose first rows are f and w respectively, such that $Fw = f * w$ and $Wf = w * f$, where $*$ represents the circular convolution. The kernel basis function set consists of one kernel centered at each pixel and thus the matrix Φ is also circulant. Gaussian-shaped kernels are used in this paper in order to produce a smooth PSF, but any other type of kernel could be used as well. It is even possible that many different types of kernels are used with small additional computational cost.

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

$$p(w) = \prod_{i=1}^{N} N(w_i|0, \alpha_i^{-1}). \qquad (5)$$

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

$$p(\alpha) = \prod_{i=1}^{N} \Gamma(\alpha_i|a^{\alpha}, \theta^{\alpha}). \qquad (6)$$

During model learning many of these parameters α_i tend to infinity, thus the corresponding weights are constrained to zero and the corresponding kernel functions are pruned from the model. This sparse PSF model, is similar to the RVM model [4][5]. 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.

The image is assumed to have been sampled from the prior distribution:

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

where Q is the Laplacian operator. This model is equivalent to assuming a simultaneously autoregressive (SAR) model on f:

$$f = Qf + \tilde{\epsilon}, \qquad (8)$$

or

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

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

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

This image model, penalizes large differences in neighboring pixels, and thus favors smooth images. Edges and textured area, would be better modeled by a non-stationary image model, such as:

$$p(f) = N(f|0, (Q^T \Gamma Q)^{-1}), \qquad (11)$$

where $\Gamma = diag\{\gamma_1, \ldots, \gamma_N\}$, so that the smoothness parameter γ would be different at each pixel. This image model, although it has been successfully applied in the image restoration problem where the PSF is assumed known[6], it was not considered here, because it introduces several computational problems.

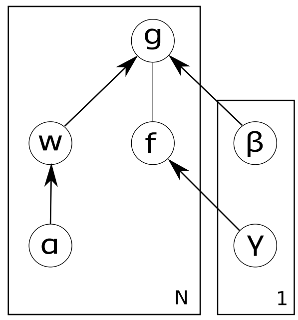

Figure 1: graphical model

The noise variance β is assumed to be a Gamma distributed random variable:

$$p(\beta) = \Gamma(\beta|a^{\beta}, \theta^{\beta}). \qquad (12)$$

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 can not be applied. Instead, we resort to approximations such as the variational Bayesian inference methodology.

The hyperprior parameters of the model are set to a small value $a^{\alpha} = \theta^{\alpha} = a^{\beta} = \theta^{\beta} = a^{\gamma} = \theta^{\gamma} = 10^{-6}$, in order to impose broad and rather uninformative hyperpriors.

3. THE VARIATIONAL METHODOLOGY FOR BAYESIAN INFERENCE

A probabilistic model consists of observed random variables D and 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[3] is an approximate inference method, which considers a family of approximate posterior distributions $q(\theta)$, and then seek 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:

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

where

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

is called the variational bound and

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

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 [3]:

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

where

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

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. Even though convergence of this iterative scheme is not proved, it is usually not a problem, especially if the approximation in 17 is not extreme.

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:

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

where

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

The required expected values are evaluated as:

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

$$\langle \|g - F\Phi w\|^2\rangle = g^T g - 2\langle w\rangle^T \Phi^T \langle F\rangle^T g + \phi^T \langle W^T W\rangle \langle F^T F\rangle \phi \quad (44)$$

where 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

$$\begin{aligned}
\langle \lambda_{w_i}^* \lambda_{w_i}\rangle &= (\mu_w * \mu_w)_i + \sum_k \Sigma_{w_{k,(k-i)}}, && (45)\\
\langle \lambda_{f_i}^* \lambda_{f_i}\rangle &= (\mu_f * \mu_f)_i + N\Sigma_{f_i}. && (46)
\end{aligned}$$

Notice that the matrix Σ_f can be easily computed, since matrices Φ, W, Q and therefore Σ_f are circulant. However, the matrix Σ_w is not circulant and thus its computation involves inverting a $N \times N$ matrix, which requires $O(N^3)$ time.

Instead, we compute the posterior weight mean μ_w, by solving the linear system:

$$\Sigma_w^{-1}\mu_w = \langle\beta\rangle \Phi^T \langle F\rangle^T g. \quad (47)$$

The conjugate gradient method was used to solve the above linear system, and in practice a good enough approximation is obtained in a few iterations.

When computing the expected value $\langle w_i^2\rangle$ and $\langle W^T W\rangle$, the matrix Σ_w in (36) and (45) is approximated by its main diagonal. This approximation is equivalent to assuming that the weights $\{w_i\}_{i=1}^N$ are independent and has been proved very efficient experimentally.

Each iteration of the variational algorithm can be splitted in two steps, first the computation of the posterior probabilities in (25) to (34) and then the computation of the expected values in (35) to (46). Convergence of the algorithm is obtained when the variational bound $\mathcal{L}$, which is computed at each iteration, stops increasing. The variational bound $\mathcal{L}$, is also used to ensure the correctness of the implementation of the algotirhm, by verifying that it increases at each iteration.

Figure 2: Gaussian PSF with large support. Images (left to right) of the true PSF that was used to generate the degraded image (top), the estimated PSF (bottom), the observed image and the estimated image.

Figure 3: Gaussian PSF with small support. Images (left to right) of the true PSF that was used to generate the degraded image (top), the estimated PSF (bottom), the observed image and the estimated image.

5. NUMERICAL EXPERIMENTS

In this section we present numerical experiments that demonstrate the ability of the algorithm to robustly estimate the support and shape of the PSF. We generated degraded images g by convolving an initial image f with a PSF h and then adding white Gaussian noise with variance $\sigma^2 = 10^{-6}$. In order to demonstrate the ability of the algorithm to estimate the support and shape of the PSF, we examined several different PSF choices.

The proposed algorithm was then used to obtain an estimate of the initial image $\hat{f}$ for each PSF choice. In practice, the shape and size of the kernels that are used to model the PSF should be selected according to any known PSF characteristics, e.g. uniform kernels should be used to model a uniform PSF. However, in these experiments, we always used the same kernels, in order to demonstrate the effectiveness of the algorithm without assuming any prior knowledge for the PSF type. Furthermore, the algorithm parameters were always initialized at the same values.

Specifically, the kernels ϕ that were used to model the PSF, were Gaussian-shaped with variance $\sigma_\phi^2 = 4$. Parameters α_i, β and γ were always initialized to values $\alpha_i = 10^{-5}, \beta = 10^3, \gamma = 10^2$ and the estimated image $\hat{f}$ was always initialized as the observed image g. The algorithm does not depend on an initial estimate of the PSF h, since it is initially estimated using the image initialization $\hat{f} = g$.

In order to evaluate the quality of the estimated image, we calculated the improved signal to noise ratio (ISNR), which is a measure of the improvement of the estimated image $\hat{f}$ with respect to the observed image g:

$$ISNR = 10 \log \frac{\|f - g\|^2}{\|f - \hat{f}\|^2}. \tag{48}$$

In the first experiment we used a Gaussian shaped PSF with variance $\sigma_h^2 = 20$ and additive noise with variance $\sigma^2 = 10^{-6}$. Figure 2 shows the degraded image, the true PSF and the image and PSF estimations provided by the algorithm. The ISNR of the estimated image was 0.92 and the euclidean distance between the true PSF and the estimated PSF was $\|h - \hat{h}\| = 3.5 \times 10^{-2}$.

The next experiment demonstrates the ability of the

Figure 4: Square-shaped PSF. Images (left to right) of the true PSF that was used to generate the degraded image (top), the estimated PSF (bottom), the observed image and the estimated image.

algorithm to determine correctly the support of the PSF. We used again a Gaussian shaped PSF but with variance $\sigma_h^2 = 5$ and applied the algorithm, initializing the parameters to exactly the same values as before. The results are shown in fig. 3. The ISNR of the estimated image was 2.05 and the euclidean distance between the true PSF and the estimated PSF was $\|h - \hat{h}\| = 4.3 \times 10^{-2}$.

Another experiment was performed with a uniform square-shaped 7×7 PSF . Notice in fig. 4 that though the estimated PSF $\hat{h}$ is not perfectly square shaped, it has the same support as the true PSF. The ISNR of the estimated image was 0.95 and the euclidean distance between the true PSF and the estimated PSF was $\|h - \hat{h}\| = 9.13 \times 10^{-2}$.

In several other experiments that were performed, the algorithm was found to be very insensitive to initial values for the parameters α. On the other hand, the parameters β and γ should be initialized to reasonable values in order to obtain good performance. Extremely bad initialization may lead to poor performance, for example, initializing γ at a very large value may lead to a very smooth image estimation.

6. CONCLUSIONS

We presented a Bayesian treatment to 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. Because of the complexity of the model, we used the variational framework to achieve inference. Several experiments have been carried out, that demonstrate the robustness of the method.

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 treated.

Acknowledgments

This research was co-funded by the European Union 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.

REFERENCES

[1] A. Likas, N. P. Galatsanos, A Variational Approach for Bayesian Blind Image Deconvolution, *IEEE Trans. on Signal Processing* , vol. 52, no. 8, pp. 2222-2233, August 2004.

[2] J. Mateos, R. Molina, A.K. Katsaggelos, Approximations Of Posterior Distributions In Blind Deconvolution Using Variational Methods, *IEEE International Conference on Image Processing*, vol 2, pp. 770-773, September 2005.

[3] M.I. Jordan, Z. Gharamani, T.S. Jaakkola, L.K. Saul. An introduction to variational methods for graphical models, In M.I. Jordan, editor, *Learning in Graphical Models*, pp 105-162. Kluwer, 1998.

[4] M. E. Tipping, Sparse Bayesian Learning and the Relevance Vector Machine, *Journal of Machine Learning Research*, 1:211-244, 2001.

[5] A. C. Faul, M. E. Tipping, A Variational Approach to Robust Regression, ICANN 2001: 95-102.

[6] G. Chantas, N. P. Galatsanos, A. Likas, Non Stationary Bayesian Image Restoration, *Proceedings of the IEEE International Conference on Pattern Recognition*, Cambridge, UK, August 2004.