NON STATIONARY BAYESIAN IMAGE RESTORATION

by

Giannis Chantas, Nikolas P. Galatsanos and Aristidis Likas
Department of Computer Science
University of Ioannina
Ioannina, Greece 45110
{chanjohn,galatsanos,arly}@cs.uoi.gr

ABSTRACT

In this paper we propose a new iterative Bayesian non stationary image restoration algorithm. The main novelty of this approach is the introduction of a hierarchical non stationary image prior. Based on this prior and the generative graphical model for the observations, Bayesian inference is performed integrating out the hidden variables. An interesting byproduct of this approach is the justification, using a Bayesian framework, of previous non stationary image restoration formulations that were based on heuristic arguments. Numerical experiments are provided that demonstrate the advantages of the proposed non stationary approach as compared with stationary approaches.

1. INTRODUCTION

Bayesian methods have been applied extensively for many signal processing problems including image restoration; see for example [1, 2]. The Bayesian formulation offers many advantages for the image restoration problem since it allows the incorporation of a priori knowledge in the form of priors about the image and the unknown parameters [3, 4].

In many Bayesian formulations for the image restoration problem Gaussian stationary models have been used for the image prior; see for example [3, 4]. A very popular model is the Simultaneously Autoregressive (SAR) in which the statistics of the image are assumed invariant for the different spatial locations; see for example [2-4]. This model greatly facilitates the parameter estimation process since only one parameter is used and thus can be easily estimated. However, it is seriously handicapped because it does not provide the flexibility to model the spatially varying correlations of the image. In other words, such prior enforces smoothness uniformly across the entire image and corresponds to uniform “regularization”.

There have been numerous efforts to ameliorate the problem of uniform regularization in image restoration. One of the most successful such effort has used spatially adaptive regularization [5-7]. The motivation and the justification for this approach is based on psycho visual arguments about the visibility of the noise in images. Furthermore, for its application the parameters used to define the noise visibility weights are selected in an ad hoc manner.

In this paper we propose a new non stationary image prior model. This model incorporates spatially varying variances for the residuals of the SAR predictor and thus provides the flexibility to model spatially varying correlations. To ameliorate the estimation problem of the spatially varying variances a Gamma hyperprior is used within and a Bayesian setting is used for inference. We propose a Bayesian methodology based on the graphical model for the observation that marginalizes the likelihood with respect to the “hidden” variables [10]. More specifically, we use a quadratic approximation of the Bayesian integral to marginalize the hidden variables.

An interesting result of the proposed approach is that the update equations for variances of the residuals of the non stationary SAR model are *identical* in form to the equations proposed for obtaining the visibility weights of the noise in images [6, 7]. In other words, using a Bayesian formulation we were able to obtain the same form of equations as in [6, 7] which were derived using heuristic arguments. In addition, for our approach all the parameters used are estimated from the observations in a systematic manner. We provide numerical experiments that demonstrate the advantages of the proposed approach and comparisons with similar in spirit Bayesian restoration methods that use stationary priors for the image.

2. IMAGING AND IMAGE MODELS

The imaging model is linear. Let $\mathbf{g}$ be a $N \times 1$ vector, representing the observed degraded image. We assume that this image is formed as

$$\mathbf{g} = \mathbf{Hf} + \mathbf{n}, \tag{1}$$

where, $\mathbf{f}$ the unknown original image to be estimated, $\mathbf{H}$ a $N \times N$ known degradation matrix, and $\mathbf{n}$ additive white noise. We assume Gaussian statistics for the noise given by $\mathbf{n} \sim N(\mathbf{0}, \beta^{-1}\mathbf{I})$ where $\mathbf{0}$ and $\mathbf{I}$ are a $N \times 1$ vector with zeros and the $N \times N$ identity matrix,

Research partially supported by the European Commission and the Hellenic Ministry of Education under EPEAK II

0-7695-2128-2/04 $20.00 (C) 2004 IEEE

respectively, and β the inverse of the noise variance is assumed unknown.

The image $\mathbf{f}$ is assumed to be generated by an SAR prediction [4] model given by

$$\mathbf{f}(k,j)=\frac{1}{4}\sum_{l=-1}^{1}\sum_{m=-1}^{1}\mathbf{f}(k+l,j+m)+\boldsymbol{\varepsilon}(k,j),\ m\neq l$$

with $\boldsymbol{\varepsilon}(k,j)$ the prediction residual for the image location (*k, j*). The above equation can be also written in matrix vector form for the entire image as $\mathbf{Qf}=\boldsymbol{\varepsilon}$, where $\mathbf{Q}$ is a $N\times N$ matrix operator. With out loss of generality, in what follows we use for convenience one dimensional notation. We assume that the residuals have Gaussian statistics according to $\varepsilon(i)\sim N\left(0,(a_i)^{-1}\right)$, for $i=1,2\ldots N$, where N the size of the image. This induces prior for the image given by

$$p\left(\mathbf{f}\,|\mathbf{a}\right)\propto\prod_{i=1}^{N}a_i^{-0.5}\exp\left(-0.5\left(\mathbf{f}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{AQf}\right)\right),\qquad(2)$$

where the matrix $\mathbf{A}$ is a diagonal $N\times N$ matrix given by $\mathbf{A}=diag\{a_1,a_2\ldots a_N\}$, with elements inverse of the variances of the residuals, and $\mathbf{a}=\left[a_1,a_2\ldots a_N\right]^T$ a $N\times 1$ vector that also contains the same parameters.

This model is *non-stationary*, because the covariance matrix changes spatially. It introduces N parameters $a_i's$ that have to be estimated from N data points, which is clearly not a desirable situation from an estimation point of view. For this purpose we use the Bayesian paradigm to bypass this difficulty and we introduce a Gamma prior for all the $a_i's$. In the case of a stationary model all $a_i's$ are equal and it is rather straight forward to obtain good estimates using maximum likelihood (ML) for the unknown parameters.

The rational for using this Gamma prior in the non stationary case is threefold. First, it is "conjugate" for the variance of a Gaussian and ameliorates the over parameterization problem of this model. Second, similar hierarchical models have been used successfully in Bayesian formulations of other statistical learning problems; see for example [8]. Finally, as we shall see in what follows it produces update equations for the $a_i's$ previously derived using different principles.

We parameterized the Gamma hyperprior as

$$p(a_i)\propto a_i^{\left(\frac{l-2}{2}\right)}e^{-m(l-2)a_i}.\qquad(3)$$

For such a representation the mean and variance of the Gamma are given by

$E\left[a_i\right]=l\left(2m(l-2)\right)^{-1}$ and $Var\left[a_i\right]=l\left(2m^2(l-2)^2\right)^{-1}$, respectively [11]. This representation is used because the value of the parameter l can be interpreted as the level of confidence to the prior knowledge provided by the Gamma hyperprior [2, 9]. More specifically, as $l\to\infty$ $E\left[a_i\right]\to(2m)^{-1}$ and $Var\left[a_i\right]\to 0$. Thus the prior becomes very restrictive. In contrast, as $l\to 2$ both $E\left[a_i\right]\to\infty$ and $Var\left[a_i\right]\to\infty$ thus the prior becomes uninformative.

3. THE BAYESIAN ALGORITHM

The graphical model that describes the observed data generation process is shown in *Figure 5*. In this figure ellipses represent the random variables and rectangles the parameters. Thus, $\mathbf{f}$ and $\mathbf{a}$ are "hidden" (latent) variables, while m, l and β are unknown parameters. In the Bayesian inference paradigm hidden variables are marginalized while parameters are estimated [10]. Given the observations $\mathbf{g}$, the parameters are estimated by maximizing the likelihood $p\left(\mathbf{g};\beta,m,l\right)$.

Based on the graphical model in *Figure 1* the likelihood is obtained by marginalizing the joint probability density function (pdf) according to

$$\begin{aligned}&p(\mathbf{g};\beta,m,l)=\int p(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l)d\mathbf{f}d\mathbf{a}=\\&\int p(\mathbf{g}\mid\mathbf{f};\beta)p(\mathbf{f}\mid\mathbf{a})p(\mathbf{a};m,l)d\mathbf{f}d\mathbf{a}.\end{aligned},\qquad(4)$$

where, $d\mathbf{a}=da_1da_2...da_N$. The exact evaluation of this Bayesian integral is not possible thus we resort to a quadratic approximation in the region of the maximum of the integrand.

For this purpose we can write the joint pdf as

$$p(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l)=\exp\left(\log p(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l)\right)$$

and we define the function J as

$$J\left(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l\right)=-\log p(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l).$$

Thus

$$\begin{aligned}&J\left(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l\right)=-\log p\left(\mathbf{g}\,|\mathbf{f};\beta\right)-\log p\left(\mathbf{f}\,|\mathbf{a}\right)-\\&-\log p\left(\mathbf{a};m,l\right)=\\&=-\frac{N}{2}\log\beta+\frac{1}{2}\beta\left\|\mathbf{Hf}-\mathbf{g}\right\|^2-\frac{1}{2}\sum_{i=1}^{N}\log a_i\\&+\frac{1}{2}\mathbf{f}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{AQf}-\frac{l-2}{2}\sum_{i=1}^{N}\log a_i+m\left(l-2\right)\sum_{i=1}^{N}a_i.\end{aligned}\qquad(5)$$

Setting the gradient of $J(\ )$ with respect to $\mathbf{f}$ and $\mathbf{a}$ equal to zero gives

$$\mathbf{f}^*=\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}+\beta^{-1}\mathbf{Q}^{\mathrm{T}}\mathbf{AQ}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{g}\qquad(6)$$

$$a_i^* = \left(\tfrac{1}{2} + (l-2)/2\right)\left(.5\left([\mathbf{Qf}](i)\right)^2 + m(l-2)\right)^{-1} \quad (7)$$

The term $\left([\mathbf{Qf}](i)\right)^2$ can be viewed as the local variance of the image at the i^{th} location while $m(l-2)$ is the bias of the prior.

We use a quadratic Taylor approximation of $J(\ )$ in the vicinity of the "mode" of the joint pdf. Thus, at this location gradient equal to zero. This approximation gives

$$\begin{aligned} J(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l) &= J\left(\mathbf{g},\mathbf{f}^*,\mathbf{a}^*;\beta,m,l\right) \\ &+\frac{1}{2}\left([\mathbf{f}^*,\mathbf{a}^*]-[\mathbf{f},\mathbf{a}]\right)^{\mathbf{T}} \nabla^2 J(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l)\big|_{(\mathbf{f}^*,\mathbf{a}^*)} \\ &\left([\mathbf{f}^*,\mathbf{a}^*]-[\mathbf{f},\mathbf{a}]\right). \end{aligned} \quad (8)$$

where $\nabla^2 J$ is the Hessian of the function $J(\ )$.

Using the approximation in (7) and the integration properties of the Gaussian pdf the Bayesian integral in (4) yields

$$\begin{aligned} p(\mathbf{g};\beta,m,l) &= \int p(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l)d\mathbf{f}d\mathbf{a} \propto \\ &\exp\left(-J\left(\mathbf{g},\mathbf{f}^*,\mathbf{a}^*;\beta,m,l\right)\right)\cdot\frac{1}{2}\left|\nabla^2 J\left(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l\right)_{(\mathbf{f}^*,\mathbf{a}^*)}\right| \end{aligned} \quad (9)$$

where $|\ |$ denotes the determinant.

The goal of this Bayesian approach is to estimate the parameters m, l and β to maximize the likelihood. Assuming $\mathbf{f}^*$ and $\mathbf{a}^*$ known, the minimization of the negative logarithm of the likelihood function $-\log\left(p\left(\mathbf{g};\beta,m,l\right)\right)$, can yield the maximum likelihood estimates unknown parameters. However, their computation from (9) is very hard because of the non circulant nature of the large matrices in the determinant. Thus, in this paper we resorted to a different approach which we describe in what follows.

4. NUMERICAL EXPERIMENTS

To demonstrate our algorithm numerical experiments using images of size 256×256 are shown. In contrast to the stationary case, for the non stationary case the computation of $\mathbf{f}^*$ in equation (6) cannot be performed in closed form in the discrete Fourier Transform domain. The matrix that has to be inverted is non circulant, because of the diagonal matrix $\mathbf{A}$ which has different elements in the diagonal. For this reason we resorted to an iterative conjugate gradient approach to find $\mathbf{f}^*$ in (6).To produce the observations the original image was blurred with a Gaussian shaped point spread function with variance 2 and then Gaussian noise was added with variance 10^{-4}. The iterative restoration algorithm iterates between (6) and (7) the estimates $\mathbf{f}^*$ and $\mathbf{a}^*$. For the reasons explained above $\mathbf{f}^*$ in (6) is computed iteratively by the conjugant gradient algorithm. The parameters m and β were assumed known and were estimated from a stationary model in which all $a_i^* = a$ for all i's. More specifically, the maximum likelihood estimates of a and β were found using the expectation maximization (EM) algorithm. The estimate of a was assumed to be the mean for the Gamma hyperprior in (3) for $l\to\infty$. This defines the parameter m, which was kept fixed in our algorithm. The value of the parameter l is not kept constant. We start iterating between (6) and (7) using a large value of l and we decrease it gradually to 2. In other words, as we estimate a_i^*'s iteratively we start by placing more confidence to the hyperprior and as we obtain better and better estimates of the original image we reduce this confidence. This becomes evident by observing (7). For $l\to\infty$ $a_i^* = (2m)^{-1}$ thus all a_i^*'s are equal to the mean of the hyperprior. For $l\to 2$ the term $\left([\mathbf{Qf}](i)\right)^2$ dominates, thus the dependence on the hyperprior is diminished.

Two error metrics were used to evaluate our results. The first is the classical $\text{MSE} = \|\mathbf{f}-\hat{\mathbf{f}}\|^2$ where $\mathbf{f}$ and $\hat{\mathbf{f}}$ the original and restored images. The second metric is the weighted MSE (WMSE) that takes into account the visibility of the errors [6, 7]. This metric is defined as $\text{WMSE} = \left(\mathbf{f}-\hat{\mathbf{f}}\right)^T \mathbf{A}\left(\mathbf{f}-\hat{\mathbf{f}}\right)$.

From the restored images shown in *Figures 1-4* it is clear that the non stationary model yields visually more pleasing results. It is interesting to notice that the MSE does not always convey accurately the visual impression of the quality of the images. For example, the restored image using the stationary model in *Figure 1* has smaller MSE than the corresponding from the non stationary model in *Figure 2*. However, the WMSE metric that incorporates the visibility of the error in the images is about 50% smaller for the non stationary model in both experiments.

5. REFERENCES

[1] Ruanaidh J., and Fitzegerald W., *Numerical Bayesian Methods Applied to Signal Processing*, Springer Verlag, 1996.

[2] Molina, R., Katsaggelos, A. K., Mateos, J., "Bayesian and regularization methods for hyper-parameter estimation in image restoration," *IEEE Trans. on Image Processing*, Vol: 8 No: 2 , pp. 231 -246, Feb. 1999.

[3] Galatsanos N. P., Mesarovic V. N., Molina R. M. and Katsaggelos A. K., "Hierarchical Bayesian Image Restoration from Partially-Known Blurs," *IEEE Trans. on Image Processing*, Vol. 9, No. 10, pp. 1784-1797, October 2000.

[4] R. Molina, and B. D. Ripley, "Using spatial models as priors in astronomical images analysis", *J. Appl. Stat.*, vol.16, pp.193-206, 1989.
[5] Anderson G., and Netrevali A., "Image Restoration Based on a Subjective Criterion", *IEEE Trans. Sys. Man Cybern.*, SMC-6, 845-853, 1976.
[6] Katsaggelos A. K., Biemond J., Mersereau R. and, Schaefer R., "A Regularized Iterative Restoration Algorithm", *IEEE Trans. Signal Processing*, 39, 914-929, 1991.
[7] Katsaggelos A. K. "Iterative Image restoration", in *Handbook on Image and Video Processing*, Editor Al. Bovik, pp. 191-206, Academic Press 2000.
[8] Tipping M. E. "Sparse Bayesian Learning And The Relevance Vector Machine" *Journal Of Machine Learning Research* 1, 211-244, 2001.
[9] Galatsanos N., V. N. Mesarovic, R. M. Molina, J. Mateos, and A. K. Katsaggelos, "Hyper-parameter Estimation Using Gamma Hyper-priors in Image Restoration from Partially-Known Blurs," *Optical Engineering*, 41(8), pp. 1845-1854, August 2002.
[10] D. J. MacKay, *Information Theory, Inference and Learning Algorithms*, Cambridge University Press, 2003.

Figure 1: Stationary restored 256x256 Leena image, MSE = 70.2, WMSE = 2.74e+015.

Figure 2: Non stationary restored 256x256 Leena image, MSE = 71.2, WMSE = 1.07e+015.

Figure 3: Segment of stationary restored Cameraman 256x256 image, MSE = 1920, WMSE = 6.9e+015.

Figure 4: Segment of non stationary restored Cameraman 256x256 image, MSE = 1884, WMSE = 2.3e+015.

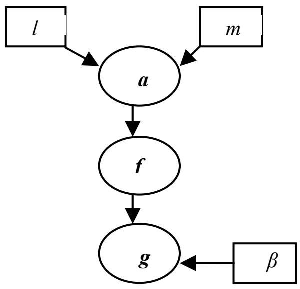

Figure 5: The graphical model of the data.