Bayesian Restoration Using a New Nonstationary Edge-Preserving Image Prior

Giannis K. Chantas, Nikolaos P. Galatsanos, and Aristidis C. Likas

Abstract—**In this paper, we propose a class of image restoration algorithms based on the Bayesian approach and a new hierarchical spatially adaptive image prior. The proposed prior has the following two desirable features. First, it models the local image discontinuities in different directions with a model which is continuous valued. Thus, it preserves edges and generalizes the on/off (binary) line process idea used in previous image priors within the context of Markov random fields (MRFs). Second, it is Gaussian in nature and provides estimates that are easy to compute. Using this new hierarchical prior, two restoration algorithms are derived. The first is based on the maximum *a posteriori* principle and the second on the Bayesian methodology. Numerical experiments are presented that compare the proposed algorithms among themselves and with previous stationary and non stationary MRF-based with line process algorithms. These experiments demonstrate the advantages of the proposed prior.**

I. INTRODUCTION

IMAGE restoration is a well-known, ill-posed inverse problem that requires regularization. Due to the wide-band nature of the additive noise and the low-pass characteristics of the image and blurring operator, smoothness constraints on the restored image are used for regularization [1]. The Bayesian formulation of the image restoration problem offers many advantages since it provides a systematic and flexible way for regularization. Furthermore, it provides a rigorous framework for estimation of the model parameters; see, for example, [2] and [3].

In many Bayesian formulations, for the image restoration problem, image priors based on a Gaussian stationary assumption for the residuals of the local image differences have been used; see, for example, [3]–[6]. The most popular such model is the simultaneously autoregressive (SAR) in which the statistics of the image are assumed invariant at all spatial locations of the image; see, for example, [3]–[6]. This model greatly facilitates the parameter estimation process since few parameters are used and, thus, can be easily estimated. However, such stationary models are seriously handicapped because they do not provide the flexibility to model the spatially varying structure of images in edge and texture areas. In other words, such priors enforce smoothness uniformly across the entire image and correspond to uniform "regularization." There have been a number of efforts to ameliorate this problem. A complete survey of all of them is out of the context of this paper. In what follows, we

will discuss two such efforts in more detail that are related to the methodology proposed in this paper.

In one such effort, a class of stochastic non stationary image models, called compound Gauss–Markov models, have been proposed; see, for example, [17]–[21]. In these models, the image is assumed to be generated by a two-level process. The first level represents the correlations of adjacent pixels of the image. The second level contains a binary process used to capture the image variations (edges). In other words, when the line process between two pixels is "on" smoothness is not enforced between them while when it is "off" smoothness is enforced.

Although restoration approaches based on doubly stochastic MRF use a rigorous model for the image and a systematic approach to learn and infer the restored image they have two disadvantages. First, they are difficult to solve from a computational point of view. One reason for this difficulty is the cyclic dependency of the random variables in such models. In other words, the corresponding generative graphical model contains cycles and is well known that learning such models is difficult and computationally expensive; see, for example, [22]. Second, from an image modeling point of view, the binary (on/off) nature of the line process that is used is insufficient to capture the image variations of most natural images. More specifically, edges of different strengths and "degrees of sharpness" are present in natural images and a binary model is limited since it inevitably introduces quantization in representing them.

In another effort to avoid uniform regularization, deterministic approaches based on the visibility of the errors in images have been developed; see, for example, [8]–[11]. These methods are based on the intuitive notion that in areas of high image variation (edge and texture) errors are less visible; thus, less smoothing is necessary, while, in areas of low activity (smooth area), the errors are more visible, and more smoothing is necessary [7].

The methods based on the error visibility idea use a continuous (non binary) model to capture the visibility of the image artifacts. Since the visibility of the artifacts is related to the variation structure of the image, these methods use a continuous model for the image variations. However, their main shortcoming is that quantification of visibility is not rigorous but rather heuristic. Thus, the estimation of all the necessary parameters is not based on a systematic framework derived from a rigorous model. As a result, such models are cumbersome to use, and suboptimal.

In this paper, we introduce a methodology that ameliorates the difficulties of the above mentioned methods. In particular, we propose a new hierarchical (two-level) Gaussian nonstationary image prior. This prior assumes that the residuals of the first order differences of the image, in four *different directions*,

are Gaussian random variables with zero mean and variance that is *spatially varying*. As a result, these local directional variances capture the image discontinuities with a continuous value model and can be thought as "continuous line processes." In order to deal with the resulting over-parameterization of this model, the spatially varying variances are considered as random variables (not parameters) and a Gamma hyper-prior is imposed on them. The parameters of the imposed hyper-prior control the mean and the variance of the residual variances and in a sense control the *degree of nonstationarity* of the imposed image prior.

Another aspect of this image model is that it enforces *sparse* first order directional differences in the image using the same Bayesian mechanism as in sparse kernel-based regression and basis selection; see, for example, [12] and [29]. Sparse signal representations have found extensive applications in inverse problems and are becoming very important area of research for many signal and image processing applications [29]. This model is an effort to capture the "best of both worlds." It is edge preserving and Gaussian in nautre, thus being easy to compute.

To learn this model and infer the image we propose two iterative algorithms. The first is based on the maximum *a posteriori* estimation (MAP) principle and computes explicitly both the image and the spatially varying variances in all four directions. The second is a Bayesian algorithm that marginalizes the "hidden variables"; see, for example, [14]. At this point, we would like to make two observations. First, unlike MRF-based models, the generative graphical model that stems from the proposed prior in this paper *does not contain cycles*; thus, learning and inference based on it is easy. Second, we obtain as MAP estimates of the inverse local variances the spatially adaptive regularization weights in [8]–[11] which were previous obtained based on heuristic arguments. An identical algorithm to the herein proposed itereative MAP can also be derived using the deterministic half-quadratic regularization methodology proposed in [30]. More specifically, (5) in this paper is identical to (16) in [30] when four direction differences are used in the potential function $\varphi_{HL}(t) = C log(A + t^2)$ with the selection $A = m(l - 2)$ and $C = l - 2 + .25$. However, the main advantage of the proposed approach is that it provides statistical interpretation for all the parameters used. Thus, unlike the half-quadratic formulation in [30], it provides with guidelines for selecting good values for them in practical problems.

We provide numerical experiments where we compared the proposed restoration algorithms with two different versions of the classical Wiener filter [24], the constrained least squares approach with spatially adaptive constraints [8]–[11], previous Bayesian algorithms based first on the stationary SAR model in [3], and second on the CGMRF that use "binary line processes" in [21]. We also compared the proposed algorithms among themselves in terms of both the bias and the variance of the inferred restored image using a Monte-Carlo simulation. Our experimental results are encouraging and demonstrate the advantages of both the proposed new prior and the Bayesian methodology.

The rest of this paper is organized as follows. In Section II, we present the imaging and the proposed image prior models. In Section III, we present the MAP-based restoration algorithm. In Section IV, we derive the Bayesian marginalization-based

algorithm. In Section V, we present our numerical experiments, and, finally, in Section VI, we provide conclusions and thoughts for future research.

II. IMAGING AND IMAGE PRIOR MODEL

A linear imaging model is assumed. The $N \times 1$ vector $\mathbf{g}$, represents the observed degraded image which is obtained by

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n} \tag{1}$$

where $\mathbf{f}$ is the *(unknown)* original image, $\mathbf{H}$ *is a* $N \times N$ known convolution matrix and $\mathbf{n}$ is additive white noise. We assume Gaussian statistics for the noise given by $\mathbf{n} \sim N\left(\mathbf{0}, \beta^{-1}\mathbf{I}\right)$ where $\mathbf{0}$ is a $N \times 1$ vector with zeros and $\mathbf{I}$ the $N \times N$ identity matrix, respectively, and β^{-1} the noise variance which is assumed unknown.

For the image prior model we assume that the first order differences of the image $\mathbf{f}$ in four directions, $0°$, $90°°$, $45°$, and $135°$, respectively, are given by

$$
\begin{aligned}
\varepsilon^1(i,j) &= f(i,j) - f(i,j+1) \\
\varepsilon^2(i,j) &= f(i,j) - f(i+1,j) \\
\varepsilon^3(i,j) &= f(i,j) - f(i+1,j+1) \\
\text{and } \varepsilon^4(i,j) &= f(i,j) - f(i-1,j+1)
\end{aligned}
\tag{2}
$$

with $\varepsilon^k(i,j)$ $k = 1, 2, 3, 4$, the difference residuals for the image location (i, j). The above equations can be also written in matrix vector form for the entire image as $\mathbf{Q}^k \mathbf{f} = \boldsymbol{\varepsilon}^k$, $k = 1, 2, 3, 4$, where $\mathbf{Q}^k$ is the $N \times N$ directional difference operators for $N \times 1$ images. Without loss of generality, in what follows, for convenience, we will use one-dimensional notation; in other words, we assume $\boldsymbol{\varepsilon}^k = \left[\varepsilon_1^k \varepsilon_2^k, \ldots \varepsilon_N^k\right]^T$. We assume that the residuals have Gaussian statistics according to $\varepsilon_i^k \sim N\left(0, \left(a_i^k\right)^{-1}\right)$, for $i = 1, 2 \ldots N$ and $k = 1, 2, 3, 4$, where a_i^k the inverse variance of ε_i^k and N the size of the image.

For the inverse variance a_i^k, we introduce the notation $\mathbf{A}^k = \text{diag}\left\{a_1^k, a_2^k, \ldots a_N^k\right\}$ an $N \times N$ diagonal matrix and $\hat{\mathbf{A}} = \text{diag}\left\{\mathbf{A}^1, \mathbf{A}^2, \mathbf{A}^3, \mathbf{A}^4\right\}$ a $4N \times 4N$ diagonal matrix and $\tilde{\mathbf{a}} = \left[\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3, \mathbf{a}^4\right]^T$ a $4N \times 1$ vector. Also, for the errors, we use the notation $\tilde{\boldsymbol{\varepsilon}} = \left[\boldsymbol{\varepsilon}^1, \boldsymbol{\varepsilon}^2, \boldsymbol{\varepsilon}^3, \boldsymbol{\varepsilon}^4\right]^T$. We assume that the errors in each direction and at each pixel location are independent. This is based on the assumption that at each pixel location an edge can occur at any direction independently of what happens in adjacent pixels. This assumption makes subsequent calculations tractable. Thus, the joint density for the errors is Gaussian and is given as

$$
\begin{aligned}
p(\tilde{\boldsymbol{\varepsilon}}; \tilde{\mathbf{a}}) &\propto \prod_{k=1}^{4} \prod_{i=1}^{N} \left(a_i^k\right)^{1/2} \exp\left(-0.5 \left(\left(\boldsymbol{\varepsilon}^k\right)^T \mathbf{A}^k \boldsymbol{\varepsilon}^k\right)\right) \\
&= \prod_{k=1}^{4} \prod_{i=1}^{N} \left(a_i^k\right)^{1/2} \exp\left(-0.5 \left(\tilde{\boldsymbol{\varepsilon}}^T \tilde{\mathbf{A}} \tilde{\boldsymbol{\varepsilon}}\right)\right).
\end{aligned}
$$

To relate $\tilde{\boldsymbol{\varepsilon}}$ with the image $\mathbf{f}$, we define the $4N \times N$ operator $\tilde{\mathbf{Q}} = \left[(\mathbf{Q}^1)^T, (\mathbf{Q}^2)^T, (\mathbf{Q}^3)^T, (\mathbf{Q}^4)^T\right]^T$. Then, the relation between

the image and the residuals is $\tilde{\boldsymbol{\varepsilon}} = \tilde{\mathbf{Q}}\mathbf{f}$. Based on this relation and $p(\tilde{\boldsymbol{\varepsilon}}; \tilde{\mathbf{a}})$, we can define an *improper prior* for the image $\mathbf{f}$. This prior is given by

$$
\begin{aligned}
p(\mathbf{f}; \tilde{\mathbf{a}}) &\propto \prod_{k=1}^{4} \prod_{i=1}^{N} \left(a_i^k\right)^{1/8} \exp\left(-0.5\left(\left(\tilde{\mathbf{Q}}\mathbf{f}\right)^T \tilde{\mathbf{A}}\tilde{\mathbf{Q}}\mathbf{f}\right)\right) \\
&= \prod_{k=1}^{4} \prod_{i=1}^{N} \left(a_i^k\right)^{1/8} \exp\left(-0.5\left(\left(\mathbf{Q}^k\mathbf{f}\right)^T \mathbf{A}^k\mathbf{Q}^k\mathbf{f}\right)\right).
\end{aligned}
\tag{3}
$$

This prior is termed *improper* since it is not scaled to integrate to 1. For a proper Gaussian the normalizing constant as a function of the spatially varying variances, a_i^k cannot be of the form in (3), since $\tilde{\mathbf{Q}}$ is not a square matrix, $\mathrm{Det}\{\tilde{\mathbf{Q}}^T \tilde{\mathbf{A}}\tilde{\mathbf{Q}}\} \neq C \prod_{k=1}^{4}\prod_{i=1}^{N}(\alpha_i^k)^{1/8}$ where C *is* a constant, even though $\tilde{\mathbf{A}}$ is diagonal. However, improper priors are used on a routinely basis with success in Bayesian modeling [15]. More specifically, the prior in (3) is obtained by assuming that all the elements of the diagonal matrix $\tilde{\mathbf{A}}$ are equal to their geometric mean $\left(\prod_{i=1}^{N}\prod_{k=1}^{4}\alpha_i^n\right)^{1/4N}$. This implies, because $\tilde{\mathbf{Q}}^T\tilde{\mathbf{A}}\tilde{\mathbf{Q}} = \sum_{k=1}^{4}(\mathbf{Q}^k)^T\mathbf{A}^k\mathbf{Q}^k$ is a $N \times N$ matrix, that $\left(\det\{\tilde{\mathbf{Q}}^T\tilde{\mathbf{A}}\tilde{\mathbf{Q}}\}\right)^{1/2} \propto \left(\prod_{i=1}^{N}\prod_{k=1}^{4}\alpha_i^n\right)^{1/8}$. This assumption results in the improper prior in (3) and leads to tractable calculations.

The role of the parameters a_i^k is to capture the directional variation structure of the image. More specifically, a large variance (small a_i^k) indicates the presence of a large variation along the direction of the difference, in other words, an edge perpendicular to this direction. The introduction of the spatially varying a_i^k scales down the differences of adjacent pixels in regions of image discontinuities. As a result this prior maintains edges and suppresses noise in smooth areas of the image. This principle is identical to the one that motivated the use of the binary (0 or 1) line process idea; see for example [17]–[21]. However, since the values of a_i^k are continuous our model can be considered as *generalization* of the MRF model with the on/off binary line process.

The drawback of the proposed prior is that it introduces $4N$ parameters a_i^k that have to be estimated from N observations. This is clearly not a desirable situation from an estimation point of view. For this purpose we employ the Bayesian paradigm and consider a_i^k as random variables (instead of parameters) and introduce Gamma hyper-priors for them. In the case of a stationary model where all a_i^ks are equal the over parameterization problem does not exist, and it is rather straightforward to obtain good estimates for the unknown parameters using even maximum likelihood (ML).

The rationale for using a Gamma prior in the non stationary case is threefold. First, it is "conjugate" for the variance of a Gaussian and facilitates analysis of the Bayesian model [14]. Second, similar hierarchical models have been used successfully in Bayesian formulations of other statistical learning problems and produce *sparse* representations; see, for example, [12] and [29]. Sparse local differences encouraged by this model are a good model for image edges which are overall much less than

the pixels in the image. Finally, as we shall see in what follows, it produces update equations for a_i^ks that were previously derived empirically.

We consider the following parameterization for the Gamma hyper-prior:

$$
p\left(a_i^k; m_k, l_k\right) \propto \left(a_i^k\right)^{(l_k-2)/2} \exp\left\{-m_k\left(l_k-2\right)a_i^k\right\} \tag{4}
$$

For such a representation the mean and variance of Gamma are given by $E\left[a_i^k\right] = l_k\left(2m_k\left(l_k-2\right)\right)^{-1}$, and $Var\left[a_i^k\right] = l_k\left(2m_k^2\left(l_k-2\right)^2\right)^{-1}$, respectively; see [3] and [13]. This representation is used because the value of the parameter l_k can be also interpreted as the level of confidence to the prior knowledge provided by the Gamma hyper prior. More specifically, as $l_k \to \infty$, $E\left[a_i^k\right] \to \left(2m_k\right)^{-1}$ and $Var\left[a_i^k\right] \to 0$. In other words, the prior becomes very informative and restrictive resulting in $a_i^k = \left(2m_k\right)^{-1} \forall i$. This also implies that the image model becomes stationary. In contrast, when $l_k \to 2$, then both $E\left[a_i^k\right] \to \infty$ and $Var\left[a_i^k\right] \to \infty$; thus, the prior becomes uninformative and does not influence at all the values of the a_i^ks. In other words, the a_i^ks are free from the moderating influence of the prior and are allowed to "vary wildly" following the data. In such case the image model becomes "highly non stationary." As a result, the value of the parameter l_k can be also viewed as a way to adjust for the *degree of non stationarity* of the image model.

III. MAXIMUM *A Posteriori* (MAP) ESTIMATION

At first, we propose a MAP approach to infer $\tilde{\mathbf{a}}$ and $\mathbf{f}$. This is based on maximization of the posterior probability. Thus, we have

$$
\begin{aligned}
p(\mathbf{f}, \tilde{\mathbf{a}} \mid \mathbf{g}; \beta, \mathbf{m}, \mathbf{l}) &\propto p(\mathbf{g}, \mathbf{f}, \tilde{\mathbf{a}}; \beta, \mathbf{m}, \mathbf{l}) \\
&= p(\mathbf{g} \mid \mathbf{f}, \tilde{\mathbf{a}}; \beta)\, p(\mathbf{f} \mid \tilde{\mathbf{a}};)\, p(\tilde{\mathbf{a}}; \mathbf{m}, \mathbf{l})
\end{aligned}
$$

where

$$
\mathbf{m} = [m_1, m_2, m_3, m_4]^T, \quad \mathbf{l} = [l_1, l_2, l_3, l_4]^T.
$$

Maximizing the quantity $p(\mathbf{g}, \mathbf{f}, \tilde{\mathbf{a}}; \beta, \mathbf{m}, \mathbf{l})$ with respect to $\mathbf{f}$ and $\tilde{\mathbf{a}}$ is equivalent to minimizing the negative logarithm

$$
\begin{aligned}
J_{MAP}(\mathbf{f}, \tilde{\mathbf{a}} \mid \mathbf{g}; \beta, \mathbf{m}, \mathbf{l}) &\propto -\log p(\mathbf{g}, \mathbf{f}, \tilde{\mathbf{a}}; \beta, \mathbf{m}, \mathbf{l}) \\
&= -\log p(\mathbf{g} \mid \mathbf{f}, \tilde{\mathbf{a}}; \beta) - \log p(\mathbf{f} \mid \tilde{\mathbf{a}}) - \log p(\tilde{\mathbf{a}}; \mathbf{m}, \mathbf{l}) \\
&\Rightarrow J(\mathbf{f}, \tilde{\mathbf{a}} \mid \mathbf{g}; \beta, \mathbf{m}, \mathbf{l}) \\
&= -\frac{N}{2}\log\beta + \frac{1}{2}\beta\|\mathbf{H}\mathbf{f} - \mathbf{g}\|^2 - \frac{1}{8}\sum_{k=1}^{4}\sum_{i=1}^{N}\log a_i^k \\
&\quad + \frac{1}{2}\sum_{k=1}^{4}\sum_{i=1}^{N}\left(\mathbf{Q}^k\mathbf{f}\right)^T \mathbf{A}^k\mathbf{Q}^k\mathbf{f} \\
&\quad - \sum_{k=1}^{4}\left(\frac{l_k-2}{2}\sum_{i=1}^{N}\log a_i^k\right) \\
&\quad + \sum_{k=1}^{4}\left(m_k\left(l_k-2\right)\sum_{i=1}^{N}a_i^k\right).
\end{aligned}
\tag{5}
$$

To minimize the above function with respect to $\mathbf{f}$ and $\tilde{\mathbf{a}}$, we adopt an iterative scheme that sets alternatively the gradient of $\mathbf{f}$ and $\tilde{\mathbf{a}}$ equal to zero. Setting $\nabla_{\tilde{\mathbf{a}}} J_{MAP}(\mathbf{f}, \tilde{\mathbf{a}} \mid \mathbf{g}; \beta, \mathbf{m}, \mathbf{l}) = 0$ yields

$$\left(a_i^k\right)^* = \frac{\left(\frac{1}{8} + \frac{1}{2}(l_k - 2)\right)}{\left(\frac{1}{2}\left(\varepsilon_i^k\right)^2 + m_k(l_k - 2)\right)}. \tag{6}$$

Setting $\nabla_{\mathbf{f}} J_{MAP}(\mathbf{f}, \tilde{\mathbf{a}} \mid \mathbf{g}; \beta, \mathbf{m}, \mathbf{l}) = 0$ yields

$$\mathbf{f}^* = \left(\mathbf{H}^T \mathbf{H} + \beta^{-1} \sum_{k=1}^{4} \left(\mathbf{Q}^k\right)^T \mathbf{A}^k \mathbf{Q}^k\right)^{-1} \mathbf{H}^T \mathbf{g}. \tag{7}$$

However, (7) cannot be solved in closed form since analytical inversion of $\left[\mathbf{H}^T \mathbf{H} + \beta^{-1} \sum_{k=1}^{4} \left(\mathbf{Q}^k\right)^T \mathbf{A}^k \mathbf{Q}^k\right]$ is not possible due to the noncirculant nature of matrices $\mathbf{A}^k$. Thus, we resort to a numerical solution using a conjugate gradient algorithm. The proposed MAP algorithm iterates between (6) and (7) till convergence.

It is interesting to point out that a formula similar to (6) was used in previous works to compute spatially varying regularization weights. Such a formula was derived based on heuristic arguments and empirical observations; see [8]–[10], and [11].

In addition, the observation of the previous section that the parameters l_k control the degree of nonstationarity of the model can be verified from (6), the MAP estimates of the (a_i^k). More specifically, when $l_k \to \infty$, $\left(a_i^k\right)^* = (2m_k)^{-1} \ \forall \ i$, and the image model becomes stationary. In contrast, when $l_k \to 2$, $\left(a_i^k\right)^* = \left(\left(\varepsilon_i^k\right)^2\right)^{-1} \ \forall \ i$; thus, the $\left(a_i^k\right)^*$s are completely unaffected from the moderating effect of the Gamma hyper-prior and only follow the data. For example, in smooth areas of the image where the local residual in the denominator of (6) tend to zero, it holds that $\left(a_i^k\right)^* \to \infty$.

IV. BAYESIAN ALGORITHM

In the Bayesian analysis of the proposed model, hidden variables are marginalized while parameters are estimated [14]. In our case, as explained in Section II, $\mathbf{f}$ and $\tilde{\mathbf{a}}$ are considered "hidden" (latent) variables, while $\mathbf{m}$, $\mathbf{l}$, and β are the unknown parameters. In the Bayesian inference paradigm, hidden variables are marginalized while parameters are estimated by maximizing the likelihood $p(\mathbf{g}; \beta, m, l)$ of the observations $\mathbf{g}$

$$\begin{aligned} p(\mathbf{g}; \beta, m, l) &= \int \int p(\mathbf{g}, \mathbf{f}, \tilde{\mathbf{a}}; \beta, m, l) d\mathbf{f} d\tilde{\mathbf{a}} \\ &= \int \int p(\mathbf{g} \mid \mathbf{f}; \beta) p(\mathbf{f} \mid \tilde{\mathbf{a}}) p(\mathbf{a}; m, l) d\mathbf{f} d\tilde{\mathbf{a}}. \end{aligned} \tag{8}$$

The exact evaluation of this Bayesian integral is not possible since we cannot integrate in closed form with respect to *both* $\mathbf{f}$ and $\tilde{\mathbf{a}}$. Instead, we marginalize in closed form only with respect to $\tilde{\mathbf{a}}$. We chose to marginalize $\tilde{\mathbf{a}}$ for two reasons: first, because the maximization with respect to $\mathbf{f}$ that follows is tractable; second, because this approach requires explicit computation of $\mathbf{f}$ as part of the Bayesian algorithm and we do not have to compute it separately. More specifically, we have

$$p(\mathbf{g}, \mathbf{f}; \beta, \mathbf{m}, \mathbf{l}) = \int p(\mathbf{f} \mid \tilde{\mathbf{a}}) p(\tilde{\mathbf{a}}; \mathbf{m}, \mathbf{l}) d\tilde{\mathbf{a}}. \tag{9}$$

The calculation of the integral can be made as shown in the equation at the bottom of the page. Observing the integrand in the above expression it is easy to notice it is of the form $x^a e^{-bx}$, i.e., it is similar to a Gamma PDF. Thus, its integral is given by (10), shown at the bottom of the next page. Replacing the integral of (10) in (9) gives (11), shown at the bottom of the next page.

Thus, the image is estimated by its value at the mode of $p(\mathbf{g}, \mathbf{f}; \beta, \mathbf{m}, \mathbf{l})$. To compute the mode, we minimize $J_B(\mathbf{g}, \mathbf{f}; \beta, m, l)$ with respect to $\mathbf{f}$

$$\mathbf{f}^* = \arg \min_{\mathbf{f}} J_B(\mathbf{g}, \mathbf{f}; \beta, \mathbf{m}, \mathbf{l})$$

(see the equation shown at the bottom of the next page).

The mode $\mathbf{f}^*$ is found by an iterative modified Newton algorithm with the following update equation

$$\mathbf{f}^{k+1} = \mathbf{f}^k - \mu^k \nabla_{\mathbf{f}}^2 J_B(\mathbf{f}^k)^{-1} \nabla_{\mathbf{f}} J_B(\mathbf{f}^k) \tag{12}$$

where $\nabla_{\mathbf{f}} J_B$ and $\nabla_{\mathbf{f}}^2 J_B$ denote the Gradient and the Hessian matrix of the J_B function respectively (with respect to $\mathbf{f}$). The

$$\begin{aligned} I &= \int p(\mathbf{f} \mid \tilde{\mathbf{a}}) p(\tilde{\mathbf{a}}; \mathbf{m}, \mathbf{l}) d\tilde{\mathbf{a}} \\ &= \prod_{k=1}^{4} \left\{ \int_0^{+\infty} \prod_{i=1}^{N} p\left(a_i^k; m_k, l_k\right) \exp\left\{ -\frac{1}{2} \sum_{i=1}^{N} \left(\varepsilon_i^k\right)^2 a_i^k + \frac{1}{8} \sum_{i=1}^{N} \log a_i^k \right\} d\mathbf{a}_k \right\} \\ &= \prod_{k=1}^{4} \left[\prod_{i=1}^{N} \Gamma\left(\frac{l_k}{2}\right)^{-1} (m_k(l_k - 2))^{l_k/2} \int_0^{+\infty} \prod_{i=1}^{N} \left(a_i^k\right)^{(l_k/2) + (1/8) - 1} \right. \\ &\qquad \times \left. \exp\left\{ -\left(m_k(l_k - 2) + \frac{1}{2}\left(\left[\mathbf{Q}^k \mathbf{f}\right](i)\right)^2\right) a_i^k \right\} d\mathbf{a}_k \right]. \end{aligned}$$

Gradient and the Hessian can be computed analytically, and are given by

$$\nabla_{\mathbf{f}} J_B = \beta \mathbf{H}^T (\mathbf{Hf} - \mathbf{g}) + \sum_{k=1}^{4} \left(\frac{l_k}{2} + \frac{1}{8} \right) (\mathbf{Q}^k)^T \mathbf{v}_k$$

$$\nabla_{\mathbf{f}}^2 J_B = \beta \mathbf{H}^T \mathbf{H} + \sum_{k=1}^{4} \left(\frac{l_k}{2} + \frac{1}{8} \right) (\mathbf{Q}^k)^T \mathbf{\Lambda}_k \mathbf{Q}^k$$

where

$$\mathbf{v}_k (i) = \frac{[\mathbf{Q}^k \mathbf{f}] (i)}{\left(m_k (l_k - 2) + \frac{1}{2} ([\mathbf{Q}^k \mathbf{f}] (i))^2 \right)}$$
$$i = 1, 2, \ldots, N$$

and the matrix $\mathbf{\Lambda}_k$ is diagonal with diagonal elements $\lambda_k (i)$ equal to

$$\lambda_k (i) = \frac{1}{m (l - 2) + \frac{1}{2} ([\mathbf{Q}^k \mathbf{f}] (i))^2} - [\mathbf{v}_k (i)]^2$$
$$i = 1, \ldots, N.$$

To find the step μ^k in (12), we adopted a backtracking line search method [23].

V. NUMERICAL EXPERIMENTS

In this section, we present numerical experiments to evaluate our algorithms. First, we compare the proposed methods with previous methods in terms of the quality of the images provided. Second, we compare the proposed MAP and Bayesian algorithms in terms of the bias and variance of the inferred restored images.

The metrics used to quantify the quality of the degraded images, the noise levels in our degraded images and the quality of our restoration results are the peak signal to noise ratio (PSNR), the signal to noise ratio (SNR), and the improvement in signal to noise ratio (ISNR), respectively. These metrics are defined as

$$\text{PSNR} = 10 \log_{10} \frac{N (255)^2}{\|\mathbf{f} - \mathbf{g}\|^2}$$

$$\text{SNR} = 10 \log_{10} \frac{\|\mathbf{f}\|^2}{N \beta^{-1}}$$

$$\text{and ISNR} = 10 \log_{10} \frac{\|\mathbf{f} - \mathbf{g}\|^2}{\|\mathbf{f} - \hat{\mathbf{f}}\|^2}$$

where β^{-1}, $\mathbf{f}$, $\hat{\mathbf{f}}$, and $\mathbf{g}$ are the noise variance, the original, restored and degraded images, respectively.

In our experiments we used the well-known 256×256 "Lena" image, shown in Fig. 1(a). The image was blurred

$$I = \prod_{k=1}^{4} \left\{ \Gamma \left(\frac{l_k}{2} \right)^{-N} (m_k (l_k - 2))^{N(l_k/2)} \Gamma \left(\frac{l_k}{2} + \frac{1}{8} \right)^{N} \right. $$
$$\left. \times \prod_{i=1}^{N} \left(m_k (l_k - 2) + \frac{1}{2} ([\mathbf{Q}^k \mathbf{f}] (i))^2 \right)^{-((l_k/2) + (1/8))} \right\} \tag{10}$$

$$p (\mathbf{g}, \mathbf{f}; \beta, \mathbf{m}, \mathbf{l}) = \left(\frac{\beta}{2\pi} \right)^{N/2}$$
$$\times \prod_{k=1}^{4} \left\{ \Gamma \left(\frac{l_k}{2} \right)^{-N} \Gamma \left(\frac{l_k}{2} + \frac{1}{8} \right)^{N} (m_k (l_k - 2))^{N l_k/2} \right\}$$
$$\times \exp \left\{ -\frac{1}{2} \beta \|\mathbf{Hf} - \mathbf{g}\|^2 \right\}$$
$$\times \prod_{k=1}^{4} \left\{ \prod_{i=1}^{N} \left(m_k (l_k - 2) + \frac{1}{2} ([\mathbf{Q}^k \mathbf{f}] (i))^2 \right)^{-(l_k/2 + 1/8)} \right\} \tag{11}$$

$$J_B (\mathbf{g}, \mathbf{f}; \beta, \mathbf{m}, \mathbf{l}) = -\frac{1}{2} \beta \|\mathbf{Hf} - \mathbf{g}\|^2 - \frac{N}{2} \log \beta$$
$$- \sum_{k=1}^{4} \left[N \log \Gamma \left(\frac{l_k}{2} \right) - \frac{N l_k}{2} \log (m_k (l_k - 2)) - N \log \Gamma \left(\frac{l_k}{2} + \frac{1}{8} \right) \right]$$
$$- \sum_{k=1}^{4} \left(\left(\frac{l_k}{2} + \frac{1}{8} \right) \sum_{i=1}^{N} \log \left(m_k (l_k - 2) + \frac{1}{2} |[\mathbf{Q}^k \mathbf{f}] (i)|^2 \right) \right)$$

Fig. 1.   (a) Original "Lena" image. (b) Degraded "Lena" image with 7×7 uniform blur and additive noise SNR $= 25.36$ dB.

by a uniform 7×7 PSF (normalized to mean equal to 1) and white Gaussian noise was added such that SNR $= 25.36$ dB. The degraded image is shown in Fig. 1(b) with corresponding PSNR 27.08 and 23.22 dB, respectively.

In order to compare the proposed approaches with previous ones, we implemented 1) the classical Wiener filter in the DFT domain [24] using the degraded image to estimate the image power spectrum assuming that the additive noise variance is known. The resulting image is shown in Fig. 1(c); 2) the classical Wiener filter in the DFT domain [24] using the original image to estimate the image power spectrum and assuming that the additive noise variance is known. Clearly, this is not a realistic scenario; however, it compares our algorithm to the performance limit of the Wiener filter. The resulting image is shown in Fig. 1(d). 3) The Bayesian approach using a stationary SAR prior [3]; the corresponding image and ISNR are shown in Fig. 1(e). 4) The iterative constrained least squares (CLS) approach with spatially adaptive regularization [8], [11]. The optimal parameters for this model were found in a trial and error fashion. The resulting image is shown in Fig. 1(f). 5) The non stationary CGMRF-based approach that uses a binary line process to model the image edges in [21]. The resulting image is shown in Fig. 1(g).

To facilitate learning the proposed image model we used the β^{-1} (additive noise variance) and equal m_k that was obtained by learning a stationary SAR model assuming a Laplacian operator $\mathbf{Q}$ for the residuals [3]. The parameters m_k were obtained as $m_k = 1/\left(2a_{\mathrm{STAT}}\right)$ where a_{STAT} the image model parameter of the stationary SAR model. The parameters l_k were selected to be equal to a value denoted by l. Since, as explained previously they can be used to adjust, the degree of nonstationarity of the image model, values in the interval $l = [2.1 - 2.5]$ were found using trial and error to provide the best restored images based on both visual criteria and the ISNR metric. Since both algorithms

run very fast (1–3 min) and only one parameter is adjusted the trial and error procedure is feasible.

In order to test the performance limits of the proposed model we implemented the MAP approach estimating the model parameters from the original image. The resulting image is shown in Fig. 1(h). The resulting restored images using the proposed methods where all the unknowns are estimated from the observations are shown in Fig. 1(i) for MAP and Fig. 1(j) for the Bayesian approach. From the restored images shown in these figures is clear that the proposed non stationary restoration algorithms provide *both* higher ISNR and visually more pleasing results than all previous stationary and nonstationary-based methods. It is interesting to point out that even when the original image is used to estimate the image statistics, as in the case of the Wiener filter, both proposed approaches outperformed it.

We also tested the proposed algorithms with wavelet-based approaches with respect to the ISNR metric using the three experiments described in [25]. Although the ISNR metric is not always an accurate measure of visual impression it is an objective metric of estimation performance. Our MAP algorithm for the first and third set of experiments in [25] gave better ISNR, as shown in Tables I and II, respectively. In the first experiment, the 256×256 "Cameraman" image was degraded by additive noise with $\beta^{-1} = 0.308$ or SNR $= 38.64$ dB, and uniform 9×9 blur shown in Fig. 2(a). We also show and provide *ISNRs* for the following cases: 1) the stationary restored image assuming an SAR prior [3], in Fig. 2(b); 2) the restored image obtained by the CLS spatially adaptive approach, [8], [11], in Fig. 2(c); and 3) the restored image by the proposed MAP algorithm in Fig. 2(d). In the third experiment described in [25], the 512×512 "Lena" image was degraded $\beta^{-1} = 49$ or SNR $= 16.62$ dB, and separable 5×5 blur implemented by blurring with a PSF given by $[1, 4, 6, 4, 1]/16$ in each direction. For the second experiments

Fig. 1. *(Continued.)* (c) Wiener filter restoration, ISNR = 3.2 dB. (d) "Optimal" Wiener filter restoration, ISNR = 4.40 dB.

Fig. 1. *(Continued.)* (e) Stationary restoration, ISNR = 4.25 dB. (f) CLS method (adaptive smoothness constraint) restoration, $\theta = 1000$, $a = 0.01$, ISNR = 4.65 dB.

in [25], the ISNR obtained by the proposed here MAP algorithm was approximately equal to the best one obtained by the methods presented in [25]. In all experiments, the same termination criterion was used as in [25]. The proposed Bayesian algorithm in this set of experiments was not as competitive and gave slightly lower ISNR than the best case of the results reported in [25].

Finally, in order to compare the statistical properties of the proposed MAP and Bayesian algorithms, we considered two metrics, the bias (BIAS) and the variance (VAR) of the restored images. These metrics were estimated by Monte-Carlo simulations using the following equations:

$$\text{BIAS} = \left\| \mathbf{f} - \bar{\mathbf{f}} \right\|$$

$$\text{VAR} = \frac{1}{M} \sum_{k=1}^{M} \left\| \bar{\mathbf{f}} - \hat{\mathbf{f}}_k \right\|^2$$

$$\text{with } \bar{\mathbf{f}} = \frac{1}{M} \sum_{k=1}^{M} \hat{\mathbf{f}}_k$$

Fig. 1. *(Continued.)* (g) Restoration with GMRF algorithm [21], ISNR $= 3.46$ dB. (h) MAP "optimal" non stationary restoration, ISNR $= 10.43$ dB, $l = 2.01$.

Fig. 1. *(Continued.)* (i) MAP non Stationary restoration, ISNR $= 5.63$ dB, $l = 2.2$. (j) Bayesian non stationary restoration, ISNR $= 5.22$ dB, $l = 2.2$.

where, $\mathbf{f}$ is the original and $\hat{\mathbf{f}}_k$ for $k = 1, \ldots, M$, the restored image, obtained from $M = 10$ different restoration runs in which the degraded images were corrupted with different noise realizations. The results for three images (in addition to "Lena" and "Cameraman" a 256×256 segment of the "Barbara" image was also used) at two different noise levels are shown in Tables III and IV, respectively. The blur used here was circular Gaussian shaped with shape parameter $\sigma^2 = 2$ (normalized to mean equal to 1).

The above experiments demonstrate that the Bayesian approach has a lower variance than the MAP approach, as ex-

pected since it marginalizes the directional variances and does not use point estimates. However, in terms of bias, both MAP and Bayesian algorithms give comparable results.

In terms of computational cost both proposed algorithms were very fast. Typically, our algorithms required about 20 iterations to converge using as criterion the change of the likelihood between successive iterations to be less than 0.1%. Our algorithms were implemented in MATLAB and take about 1–4 min on a Pentium 4 at 2.8 GHz personal computer for 256×256 images. In contrast, a C implementation of the deterministic relaxation MAP algorithm in [21] required 10–15

Fig. 2. (a) Degraded "cameraman" image with 9×9 uniform blur and additive noise SNR $= 38.64$ dB, $(\beta^{-1} = 0.308)$. (b) Stationary restoration, ISNR $= 6.88$ dB.

Fig. 2. *(Continued.)* (c) CLS method (adaptive smoothness constraint) restoration, $\theta = 0.05$, $a = 0.003$, ISNR $= 7.38$ dB. (d) MAP non stationary restoration, ISNR $= 9.42$ dB, $l = 2.1$.

min on a Xeon 3.2-GHz machine. The constrained least squares method with spatially adaptive constraints [8]–[11] was implemented using a conjugate gradient algorithm and is of the same computational complexity, given that the correct parameters have been found, to the proposed methods. The Wiener filter and the Bayesian approach with the stationary SAR model are much faster since all calculations are done in the DFT domain and require 5"–10" using MATLAB on a Pentium 4 at 2.8 GHz personal computer.

VI. CONCLUSION AND FUTURE RESEARCH

The power of the proposed image prior model was clearly demonstrated with the MAP approach when the original image was used to estimate the model parameters. Apart from the very high visual quality of the restored images, shown in Fig. 1(h), in terms of ISNR it outperforms by almost 5 dB the one obtained by the Wiener filter shown in Fig. 1(d), when the original image was also used to estimate the power spectrum. Furthermore, the

TABLE I
ISNR COMPARISONS WITH THE EXPERIMENTS IN [25, TABLE I]

Method	ISNR
Proposed MAP approach, l=2.1	8.91dB
Rule (22), UDWT [25]	7.47dB
Rule (22), Random shifts [25]	7.59dB
Modified Laplacian, UDWT [25]	7.26dB
Modified Laplacian, random shifts [25]	7.34dB
Shoft-threshold, UDWT [25]	7.26dB
Shoft-threshold, random shifts [25]	6.33dB
Result by Neelamani et al [28]	7.3dB
Result by Banham and Katsaggelos [26]	6.7dB

TABLE II
ISNR COMPARISONS WITH THE EXPERIMENTS IN [25, TABLE III]

Method	ISNR
Proposed MAP approach, l=2.1,	3.77dB
Rule (22), UDWT [25]	2.94dB
Rule (22), random shifts [25]	1.71dB
Modified Laplacian, UDWT [25]	2.75dB
Modified Laplacian, random shifts [25]	1.77dB
Shoft-threshold, UDWT [25]	2.75dB
Shoft-threshold, random shifts [25]	1.61dB
Best result by Liu and Moulin [27]	1.078dB

TABLE III
BIAS METRIC FOR THE MAP AND THE BAYESIAN ALGORITHMS

Noise Level	SNR=65.34 dB		SNR =14.91 dB	
Algorithm	Bayesian	MAP	Bayesian	MAP
Cameraman	9.36	9.15	15.22	14.08
Lena	5.31	5.31	9.29	10.71
Barbara	5.33	6.97	16.42	14.25

TABLE IV
VARIANCE METRIC FOR THE MAP AND THE BAYESIAN ALGORITHMS

Noise Level	SNR =65.34 dB		SNR =14.91 dB	
Algorithm	Bayesian	MAP	Bayesian	MAP
Cameraman	1.77e-004	1.77e-004	2.37e-004	6.49e-004
Lena	6.47e-005	9.16e-005	1.71e-004	5.25e-004
Barbara	1.69e-004	2.89e-004	1.88e-004	8.95e-004

proposed methods compared favorably with recent CGMRF and wavelet-based methods [21] and [25], respectively.

Since the parameters l_k of the proposed hyperpriors can be viewed as quantifying the degree of nonstationarity of the image model, developing an image prior with spatially varying parameters of the hyperprior seems a natural extension to this work. More specifically, we will focus on developing hierarchical priors with an additional hidden layer that allows spatially varying hyperparameters for the hyperpriors. We also believe that the proposed image prior can be used in other related image processing problems, for example, image reconstruction from projections, motion field estimation, and denoising and restoration of video.

REFERENCES

[1] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," IEEE Trans. Image Process., vol. 1, no. 7, pp. 322–336, Jul. 1992.
[2] J. Ruanaidh and W. Fitzegerald, Numerical Bayesian Methods Applied to Signal Processing. New York: Springer Verlag, 1996.
[3] R. Molina, "On the hierarchical Bayesian approach to image restoration. Applications to astronomical images," IEEE Trans. Pattern Anal. Mach. Intell., vol. 16, no. 11, pp. 1122–1128, Nov. 1994.
[4] R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyper-parameter estimation in image restoration," IEEE Trans. Image Process., vol. 8, no. 2, pp. 231–246, Feb. 1999.
[5] N. P. Galatsanos, V. N. Mesarovic, R. M. Molina, and A. K. Katsaggelos, "Hierarchical Bayesian image restoration from partially-known blurs," IEEE Trans. Image Process., vol. 9, no. 10, pp. 1784–1797, Oct. 2000.
[6] R. Molina and B. D. Ripley, "Using spatial models as priors in astronomical images analysis," J. Appl. Statist., vol. 16, pp. 193–206, 1989.
[7] G. Anderson and A. Netrevali, "Image restoration based on a subjective criterion," IEEE Trans. Syst., Man, Cybern., no. 6, pp. 845–853, Dec. 1976.
[8] A. Katsaggelos, J. Biemond, R. M. Mersereau, and R. W. Schafer, "Non stationary iterative image restoration," in Proc. IEEE Int. Conf. Acoustics Speech and Signal Processing, Tampa, FL, 1985, pp. 696–699.
[9] R. L. Lagendijk, J. Biemond, and D. E. Boekee, "Regularized iterative image restoration with ringing reduction," IEEE Trans. Acoust., Speech, Signal Process., vol. 36, no. 12, pp. 1874–1887, Dec. 1988.
[10] A. K. Katsaggelos, J. Biemond, R. Mersereau, and R. Schaefer, "A regularized iterative restoration algorithm," IEEE Trans. Signal Process., vol. 39, no. 4, pp. 914–929, Apr. 1991.
[11] A. K. Katsaggelos, , A. Bovik, Ed., "Iterative image restoration," in Handbook on Image and Video Processing. New York: Academic, 2000, pp. 191–206.
[12] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," J. Mach. Learn. Res., vol. 1, pp. 211–244, 2001.
[13] N. Galatsanos, 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," Opt. Eng., vol. 41, no. 8, pp. 1845–1854, Aug. 2002.
[14] J. Berger, Statistical Decision Theory and Bayesian Analysis. New York: Springer, 1993.
[15] J. M. Bernardo and A. F. M. Smith, Bayesian Theory. New York: Wiley, 2000.

[16] D. J. MacKay, "Bayesian methods for backpropagation networks," in *Models of Neural Networks III*, E. Domany, J. L. Van Hammen, K. Schulten, and editors, Eds. New York: Springer-Verlag, 1994, ch. 6, pp. 211–254.

[17] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, pp. 228–238, Nov. 1984.

[18] F. C. Jeng and J. W. Woods, , A. K. Katsaggelos, Ed., "Compound Gauss-Markov models for image processing," in *Digital Image Restoration*. Berlin, Germany: Springer, 1991, vol. 23.

[19] F.-C. Jeng and J. W. Woods, "Compound Gauss-Markov fields for image estimation," *IEEE Trans. Signal Process.*, vol. 39, no. 3, Mar. 1991.

[20] R. Chellapa, T. Simchony, and Z. Lichtenstein, "Image estimation using 2D noncausal Gaussian Markov random field models," in *Digital Image Restoration*, A. K. Katsaggelos, Ed. Berlin, Germany: Springer, 1991, vol. 23.

[21] R. Molina, A. K. Katsaggelos, J. Mateos, A. Hermoso, and C. A. Segall, "Restoration of severely blurred high range images using stochastic and deterministic relaxation algorithms in compound Gauss-Markov random fields," *Pattern Recognit.*, vol. 33, pp. 555–571, 2000.

[22] A. Wilsky, "Multiresolution Markov models for signal and image processing," *Proc. IEEE*, vol. 90, no. 8, pp. 1397–1458, Aug. 2002.

[23] S. G. Nash and A. Sofer, *Linear and Nonlinear Programming*. New York: Mc Graw Hill, 1996.

[24] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[25] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 866–881, Aug. 2003.

[26] M. Banham and A. Katsaggelos, "Spatially adaptive wavelet-based multiscale image restoration," *IEEE Trans. Image Process.*, vol. 3, no. 6, pp. 619–634, Nov. 1996.

[27] J. Liu and P. Moulin, "Complexity-regularized image restoration," in *Proc. IEEE Int. Conf. Image Processing*, Chicago, IL, 1998, vol. 1, pp. 555–559.

[28] R. Neelamani, H. Choi, and R. Baraniuk, "ForWaRD: Fourier-wavelet regularized deconvolution for ill-conditioned systems," *IEEE Trans. Image Process.*, vol. 52, no. 2, pp. 418–433, Feb. 2004.

[29] D. Wipf and B. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[30] P. Charbonnier, L. Bland-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge preserving regularization in computed imaging," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–311, Feb. 1997.

Giannis K. Chantas received the Diploma degree in computer science and the M.S. degree from the University of Ioannina, Ioannina, Greece, in 2002 and 2004, respectively, where he is currently pursuing the Ph.D. degree in the Department of Computer Science.

His research interests are in the areas of statistical image processing and include image restoration, blind deconvolution, and super-resolution.

Nikolaos P. Galatsanos received the Diploma of Electrical Engineering from the National Technical University of Athens, 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*.

Aristidis C. Likas received the Diploma degree in electrical engineering and the Ph.D. degree in electrical and computer engineering, both from the National Technical University of Athens, Athens, Greece.

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