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 two desirable features. First, it models the image discontinuities in different directions with a continuous value model. Thus, it generalizes the on/off (binary) line process idea used in previous image priors within the context of Markov Random Fields (MRF). Second, it is based on an acyclic graphical model, unlike MRFs, which greatly facilitates learning and inference. Using this new hierarchical prior two restoration algorithms are derived. The first is based on the maximum a posteriori (MAP) 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.
1. 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], [4], [5] and [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], [4], [5], and [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], [18], [19] and [20]. 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], [9], [10], and [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 thus 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 non-stationary 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 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 non-stationarity of the imposed image prior.

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 [9], [10], [11], and [8] which were previously obtained based on heuristic arguments.

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 2 we present the imaging and the proposed image prior models. In section 3 we present the MAP based restoration algorithm. In section 4 we derive the Bayesian marginalization based algorithm. In section 5 we present our numerical experiments, and finally in section 6 we provide conclusions and thoughts for future research.
2. IMAGING AND IMAGE PRIOR MODEL

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

$$g = Hf + n,$$  \hfill (1)

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

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

$$
epsilon^1(i, j) = f(i, j) - f(i, j + 1), \quad 
epsilon^2(i, j) = f(i, j) - f(i + 1, j),$$

$$
epsilon^3(i, j) = f(i, j) - f(i + 1, j + 1), \quad 
epsilon^4(i, j) = f(i, j) - f(i - 1, j + 1),$$  \hfill (2)

with $\nepsilon^k(i, j)$ $k=1, 2, 3, 4$, 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 $Q^k f = \epsilon^k$, $k=1, 2, 3, 4$ where $Q^k$ 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 $\epsilon^k = [\epsilon^k_1, \epsilon^k_2, \ldots, \epsilon^k_N]^T$. We assume that the residuals have Gaussian statistics according to $\epsilon^k_i \sim N\left(0, (a_i^k)^{-1}\right)$, for $i=1, 2 \ldots N$ and $k=1, 2, 3, 4$, where $a_i^k$ the inverse variance of $\epsilon^k_i$ and $N$ the size of the image.

For the inverse variance $a_i^k$ we introduce the notation $A^k = diag \{a_1^k, a_2^k, \ldots, a_N^k\}$ an $N \times N$ diagonal matrix and $\tilde{A} = diag \{A^1, A^2, A^3, A^4\}$ a $4N \times 4N$ diagonal matrix and $\tilde{a} = [a^1, a^2, a^3, a^4]^T$ a $4N \times 1$ vector. Also for the errors we use the notation $\tilde{\epsilon} = [\epsilon^1, \epsilon^2, \epsilon^3, \epsilon^4]^T$.

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

\[ p(\tilde{e}; \tilde{a}) \propto \prod_{k=1}^{4} \prod_{i=1}^{N} (a_i^k)^{1/2} \exp \left( -0.5 \left( \tilde{e}^k \right)^T \tilde{A}^k \tilde{e}^k \right) \]

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

\[ \tilde{e} = \tilde{Q} f . \]

Based on this relation and \( p(\tilde{e}; \tilde{a}) \) we can define an improper prior for the image \( f \). This prior is given by:

\[ p(f; \tilde{a}) \propto \prod_{k=1}^{4} \prod_{i=1}^{N} (a_i^k)^{1/8} \exp \left( -0.5 \left( \tilde{Q} f \right)^T \tilde{A} \tilde{Q} f \right) \]

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 Eq. (3), since \( \tilde{Q} \) is not a square matrix, \( \text{Det} \left\{ \tilde{Q}^T \tilde{A} \tilde{Q} \right\} \neq \prod_{k=1}^{4} \prod_{i=1}^{N} (a_i^k)^{1/8} \) where \( C \) is a constant, even though \( \tilde{A} \) is diagonal. However, improper priors are used on a routinely basis with success in Bayesian modeling [15]. More specifically, the prior in equation (3) is obtained by assuming that all the elements of the diagonal matrix \( \tilde{A} \) are equal to their geometric mean \( \left( \prod_{i=1}^{N} \prod_{k=1}^{4} a_i^k \right)^{1/4N} \). This implies, because \( \tilde{Q}^T \tilde{A} \tilde{Q} = \sum_{k=1}^{4} (Q^k)^T \tilde{A}^k Q^k \) is a \( N \times N \) matrix, that \( \left( \text{det} \left\{ \tilde{Q}^T \tilde{A} \tilde{Q} \right\} \right)^{1/2} \propto \left( \prod_{i=1}^{N} \prod_{k=1}^{4} a_i^k \right)^{1/8} \). This assumption results in the improper prior in Eq. (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 $a^k_i$ 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], [18], [19],
[20] and [21]. However, since the values of $a^k_i$ 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 $a^k_i$ 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 $a^k_i$ as
random variables (instead of parameters) and introduce Gamma hyper-priors for them. In the
case of a stationary model where all $a^k_i$’s 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].
Finally, as we shall see in what follows it produces update equations for $a^k_i$’s that were
previously derived empirically.

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

$$p(a^k_i;m_k,l_k) \propto \left(a^k_i\right)^{l_k/2-1} \exp\left\{-m_k(l_k-2)a^k_i\right\},$$  \hspace{1cm} (4)

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

$$E\left[a^k_i\right]=l_k\left(2m_k(l_k-2)\right)^{-1}, \text{ and } Var\left[a^k_i\right]=l_k\left(2m_k^2(l_k-2)^2\right)^{-1}$$

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 \( l_i \to \infty \), \( E[a_i^k] \to (2m_k)^{-1} \) and \( Var[a_i^k] \to 0 \). In other words, the prior becomes very informative and restrictive resulting in \( a_i^k = (2m_k)^{-1} \forall i \). This also implies that the image model becomes stationary. In contrast, when \( l_i \to 2 \) then both \( E[a_i^k] \to \infty \) and \( Var[a_i^k] \to \infty \), thus the prior becomes uninformative and does not influence at all the values of the \( a_i^k \)'s. In other words, the \( a_i^k \)'s 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.

The graphical model corresponding to the observations generation mechanism is shown in Figure 1 below. Random variables are represented as ellipses and parameters as rectangles. The variables \( \hat{a} \) and \( f \) are hidden, since they are not observed and relate the parameters with the observations. In the MAP approach, point estimates are obtained for \( \hat{a} \) and \( f \). In the Bayesian approach, instead of using point estimates for \( \hat{a} \), their average influence is taken into consideration by integrating them out (marginalization), therefore point estimates are used only for \( f \).
3. MAXIMUM A POSTERIORI (MAP) ESTIMATION

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

\[
p(\tilde{a}, f | g, \beta, \mathbf{m}, l) \propto p(g | f, \tilde{a}, \beta, \mathbf{m}, l) = p(f | \tilde{a}, \beta) p(\tilde{a} | \mathbf{m}, l)
\]

where:

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

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

\[
J_{MAP}(f, \tilde{a} | g, \beta, \mathbf{m}, l) = -\log p(g, f, \tilde{a}, \beta, \mathbf{m}, l) = -\log p(g | f, \tilde{a}, \beta) - \log p(f | \tilde{a}) - \log p(\tilde{a} | \mathbf{m}, l)
\]

\[
\Rightarrow J(f, \tilde{a} | g, \beta, \mathbf{m}, l) = -\frac{N}{2} \log \beta + \frac{1}{2} \beta \| \mathbf{H} \mathbf{f} - \mathbf{g} \|^2 - \frac{1}{8} \sum_{i=1}^{4} \sum_{j=1}^{N} \log a_{ij}^k + \frac{1}{2} \sum_{i=1}^{4} \left( \sum_{j=1}^{N} \log a_{ij}^k \right) + \frac{1}{2} \left( \sum_{i=1}^{4} \left( m_1^i - 2 \sum_{j=1}^{N} a_{ij}^k \right) \right)
\]

To minimize the above function with respect to \( f \) and \( \tilde{a} \), we adopt an iterative scheme that sets alternatively the gradient of \( f \) and \( \tilde{a} \) equal to zero. Setting \( \nabla_f J_{MAP}(f, \tilde{a} | g, \beta, \mathbf{m}, l) = 0 \) yields
\[
(a^i) = \frac{\left(\frac{1}{8} + \frac{1}{2}(l_k - 2)\right)}{\left[\frac{1}{2}(e^i_l)^2 + m_i (l_k - 2)\right]}.
\] (6)

Setting \( \nabla_i J_{MAP}(f, \tilde{a} | g; \beta, m, 1) = 0 \) yields:

\[
f^* = \left( H^T H + \beta^{-1} \sum_{k=1}^{K} \left( Q^k \right)^T A^k Q^k \right)^{-1} H^T g.
\] (7)

However, equation (7) cannot be solved in closed form since analytical inversion of
\[
\left[ H^T H + \beta^{-1} \sum_{k=1}^{K} \left( Q^k \right)^T A^k Q^k \right]
\] is not possible due to the non-circulant nature of matrices \( A^k \).

Thus, we resort to a numerical solution using a conjugate gradient algorithm. The proposed MAP algorithm iterates between equations (6) and (7) till convergence.

It is interesting to point out that a formula similar to Eq. (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], [9], [10], and [11].

In addition, the observation of the previous section that the parameters \( l_k \) control the degree of non-stationarity of the model can be verified from Eq. (6), the MAP estimates of the \( (a^i) \). More specifically, when \( l_k \to \infty \), \( (a^i) = (2m_i)^{-1} \forall i \), and the image model becomes stationary. In contrast, when \( l_k \to 2 \), \( (a^i) = \left((e^i_l)^2\right)^{-1} \forall i \), thus the \( (a^i) \)'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 Eq. (6) tend to zero, it holds that \( (a^i) \to \infty \).

4. 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 2, \( f \) and \( \tilde{a} \) are considered “hidden” (latent) variables, while \( m \), \( l \) and \( \beta \) are the unknown parameters. In the Bayesian
inference paradigm, hidden variables are marginalized while parameters are estimated by maximizing the likelihood \( p(g; \beta, m, l) \) of the observations \( g \):

\[
p(g; \beta, m, l) = \int \int p(g, f, \beta, m, l) df d\beta = \int \int p(g | f, \beta) p(f | \beta) p(\beta, m, l) df d\beta.
\]  

The exact evaluation of this Bayesian integral is not possible since we cannot integrate in closed form with respect to both \( f \) and \( \beta \). Instead we marginalize in closed form only with respect to \( \beta \). We chose to marginalize \( \beta \) for two reasons. First, because the maximization with respect to \( f \) that follows is tractable. Second, because this approach requires explicit computation of \( f \) as part of the Bayesian algorithm and we do not have to compute it separately. More specifically, we have:

\[
p(g, f; \beta, m, l) = \int p(f | \beta) p(\beta, m, l) df.
\]  

The calculation of the integral can be made as follows

\[
I = \int p(f | \beta) p(\beta, m, l) df d\beta = \prod_{i=1}^{4} \left\{ \prod_{i=1}^{N} \left( \frac{1}{\sqrt{2\pi}} \right)^{-N} \Gamma\left( \frac{l_i}{2} + \frac{1}{8} \right) \prod_{i=1}^{N} \left( m_i \left( t_i - 2 \right) \right)^{\frac{l_i}{2} + \frac{1}{8}} \exp \left\{ -\left( m_i \left( t_i - 2 \right) + \frac{1}{2} \left( \left[ Q^T f \right] (i) \right)^2 \right) \right\} \right\}.
\]  

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:

\[
I = \prod_{i=1}^{4} \left\{ \Gamma\left( \frac{l_i}{2} + \frac{1}{8} \right) \prod_{i=1}^{N} \left( m_i \left( t_i - 2 \right) \right)^{\frac{l_i}{2} + \frac{1}{8}} \Gamma\left( \frac{l_i}{2} \right) \prod_{i=1}^{N} \left( m_i \left( t_i - 2 \right) + \frac{1}{2} \left( \left[ Q^T f \right] (i) \right)^2 \right)^{\frac{l_i}{2} + \frac{1}{8}} \right\}.
\]  

Replacing the integral of Eq. (12) in Eq. (11) gives:

\[
p(g, f; \beta, m, l) = \left( \frac{\beta}{2\pi} \right)^{N/2} \prod_{i=1}^{4} \left\{ \Gamma\left( \frac{l_i}{2} + \frac{1}{8} \right) \prod_{i=1}^{N} \left( m_i \left( t_i - 2 \right) \right)^{\frac{l_i}{2} + \frac{1}{8}} \right\} \times \exp \left\{ -\frac{1}{2} \beta \left[ Hf - g \right]^2 \right\} \prod_{i=1}^{4} \left\{ \prod_{i=1}^{N} \left( m_i \left( t_i - 2 \right) + \frac{1}{2} \left( \left[ Q^T f \right] (i) \right)^2 \right)^{\frac{l_i}{2} + \frac{1}{8}} \right\}.
\]  

Thus, the image is estimated by its value at the mode of \( p(g, f; \beta, m, l) \). To compute the mode, we minimize \( J_p(g, f; \beta, m, l) \) with respect to \( f \).
\[
\mathbf{f}^* = \arg \min_{\mathbf{f}} J_{\beta}(\mathbf{g}, \mathbf{f}; \beta, \mathbf{m}, \mathbf{l}),
\]

where:

\[
J_{\beta}(\mathbf{g}, \mathbf{f}; \beta, \mathbf{m}, \mathbf{l}) = \frac{1}{2} \mathbf{H} \mathbf{f}^* - \mathbf{g}^2 - \frac{N}{2} \log \beta - \sum_{k=1}^{K} \left[ N \log \Gamma \left( \frac{l_k}{2} \right) - N \log \Gamma \left( \frac{l_k}{2} + \frac{1}{8} \right) - \frac{N}{2} \log \left( m_k (l_k - 2) \right) - N \log \Gamma \left( \frac{l_k}{2} + \frac{1}{8} \right) \right] - \sum_{k=1}^{K} \left[ \frac{l_k}{2} \sum_{i=1}^{N} \log \left( m_k (l_k - 2) + \frac{1}{2} \left[ \mathbf{Q}(\mathbf{f}^*)^2 \right] \right) \right].
\]

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

\[
\mathbf{f}^{k+1} = \mathbf{f}^k - \mu^k \nabla^2 J_{\beta}(\mathbf{f}^k)^{-1} \nabla J_{\beta}(\mathbf{f}^k), \quad (14)
\]

where \( \nabla J_{\beta} \) and \( \nabla^2 J_{\beta} \) denote the Gradient and the Hessian matrix of the \( J_{\beta} \) function respectively (with respect to \( \mathbf{f} \)). The Gradient and the Hessian can be computed analytically, see Appendix A. To find the step \( \mu^k \) in (14) we adopted a backtracking line search method [23].

5 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:

\[
PSNR = 10 \log_{10} \frac{N (255)^2}{\| \mathbf{f} - \mathbf{g} \|^2}, \quad SNR = 10 \log_{10} \frac{\| \mathbf{f} \|^2}{\| \mathbf{f} - \mathbf{g} \|^2}, \quad \text{and ISNR} = 10 \log_{10} \frac{\| \mathbf{f} - \mathbf{g} \|^2}{\| \mathbf{f} - \mathbf{f}^* \|^2},
\]
where, \( \beta^{-1} \), \( f \), \( \hat{f} \) and \( g \), are the noise variance, the original, restored and degraded images, respectively.

In our experiments we used the well-known \( 256 \times 256 \) “Lena” image, shown in Figure 2a. The image was blurred by a uniform \( 7 \times 7 \) PSF (normalized to mean equal to 1) and white Gaussian noise was added such that \( \text{SNR}=25.36 \text{dB} \). The degraded image is shown in Figure 2b with corresponding \( \text{PSNR} \) 27.08 and 23.22 dB, respectively.

In order to compare the proposed approaches with previous ones we implemented: (i) 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 Figure 2c. (ii) 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 Figure 2d. (iii) The Bayesian approach using a stationary SAR prior [3]. The corresponding image and ISNR are shown in Figure 2e. (iv) 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 Figure 2f. (v) 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 Figure 2g.

To facilitate learning the proposed image model we used the \( \beta^{-1} \) (additive noise variance) and equal \( m_k \) that was obtained by learning a stationary SAR model assuming a Laplacian operator \( Q \) for the residuals [3]. The parameters \( m_k \) were obtained as 

\[
m_k = 1/(2a_{STAT})
\]

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 non-stationarity 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 minutes) 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 Figure 2h. The resulting restored images using the proposed methods where all the unknowns are estimated from the observations are shown in Figure 2i for MAP and Figure 2j 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 non-stationary 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 1 and 2, 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 9x9 blur shown in Figure 3a. We also show and provide ISNRs for the following cases: i) the stationary restored image assuming an SAR prior [3], in Figure 3b. ii) The restored image obtained by the CLS spatially adaptive approach [8] [11], in Figure 3c and iii) the restored image by the proposed MAP algorithm in Figure 3d. In the third experiment described in [25] the 512×512 “Lena” image was degraded $\beta^{-1} = 49$ or SNR=16.62 dB, and separable 5x5 blur implemented by blurring with a PSF given by [1, 4, 6, 4, 1]/16 in each direction. For the second experiments 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 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

$$BIAS = \| f - \bar{f} \|, \quad VAR = \frac{1}{M} \sum_{k=1}^{M} \| \bar{f} - \hat{f}_k \|^2,$$

where $f$ is the original and $\hat{f}_k$ for $k=1,...,M$, 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 3 different noise levels are shown in Tables 3 and 4, respectively. The blur used here was circular Gaussian shaped with shape parameter $\sigma^2 = 2$ (normalized to mean equal to 1).

<table>
<thead>
<tr>
<th>Method</th>
<th>ISNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed MAP approach, $l=2.1$</td>
<td>9.42dB</td>
</tr>
<tr>
<td>Rule (22), UDWT [25]</td>
<td>7.47dB</td>
</tr>
<tr>
<td>Rule (22), Random shifts [25]</td>
<td>7.59dB</td>
</tr>
<tr>
<td>Modified Laplacian, UDWT [25]</td>
<td>7.26dB</td>
</tr>
<tr>
<td>Modified Laplacian, random shifts [25]</td>
<td>7.34dB</td>
</tr>
<tr>
<td>Shoft-threshold, UDWT [25]</td>
<td>7.26dB</td>
</tr>
<tr>
<td>Shoft-threshold, random shifts [25]</td>
<td>6.33dB</td>
</tr>
<tr>
<td>Result by Neelamani et al [28]</td>
<td>7.3dB</td>
</tr>
<tr>
<td>Result by Banham and Katsaggelos [26]</td>
<td>6.7dB</td>
</tr>
</tbody>
</table>

Table 1: ISNR comparisons with the experiments in Table I of [25].

<table>
<thead>
<tr>
<th>Method</th>
<th>ISNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed MAP approach, $l=2.1,$</td>
<td>3.55dB</td>
</tr>
<tr>
<td>Rule (22), UDWT [25]</td>
<td>2.94dB</td>
</tr>
</tbody>
</table>
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 2: ISNR comparison with experiments in Table III of [25].

<table>
<thead>
<tr>
<th>Noise Level</th>
<th>SNR=65.34 dB</th>
<th>SNR =25.36 dB</th>
<th>SNR =14.91 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm</td>
<td>Bayesian MAP</td>
<td>Bayesian MAP</td>
<td>Bayesian MAP</td>
</tr>
<tr>
<td>Cameraman</td>
<td>9.36 9.15</td>
<td>12.23 11.57</td>
<td>15.22 14.08</td>
</tr>
<tr>
<td>Lena</td>
<td>5.31 5.31</td>
<td>6.92 7.44</td>
<td>9.29 10.71</td>
</tr>
<tr>
<td>Barbara</td>
<td>5.33 6.97</td>
<td>13.42 11.94</td>
<td>16.42 14.25</td>
</tr>
</tbody>
</table>

Table 3: Bias metric for the MAP and the Bayesian algorithms.

<table>
<thead>
<tr>
<th>Noise Level</th>
<th>SNR =65.34 dB</th>
<th>SNR =25.36 dB</th>
<th>SNR =14.91 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm</td>
<td>Bayesian MAP</td>
<td>Bayesian MAP</td>
<td>Bayesian MAP</td>
</tr>
<tr>
<td>Cameraman</td>
<td>1.77e-004 1.77e-004</td>
<td>1.64e-004 3.60e-004</td>
<td>2.37e-004 6.49e-004</td>
</tr>
<tr>
<td>Lena</td>
<td>6.47e-005 9.16e-005</td>
<td>1.04e-004 1.38e-004</td>
<td>1.71e-004 5.25e-004</td>
</tr>
<tr>
<td>Barbara</td>
<td>1.69e-004 2.89e-004</td>
<td>1.62e-004 6.02e-004</td>
<td>1.88e-004 8.95e-004</td>
</tr>
</tbody>
</table>

Table 4: Variance metric for the MAP and the Bayesian algorithms.

The above experiments demonstrate that the Bayesian approach has a lower variance than the MAP approach, as expected 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 minutes on a Pentium 4 at 2.8 GHz personal computer for $256 \times 256$ images. In contrast, a C implementation of the deterministic relaxation MAP algorithm in [21] required 10-15 minutes 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.

6. CONCLUSIONS 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 Figure 2h, in terms of ISNR it outperforms by almost 5dB the one obtained by the Wiener filter shown in Figure 2d, when the original image was also used to estimate the power spectrum. Furthermore, the 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 non-stationarity 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.

ACKNOWLEDGEMENT

We would like to acknowledge Dr. Rafael Molina and Dr. Javier Mateos from Departamento de Ciencias de la Computación e I. A.E.T.S. Ing. Informática Universidad de Granada, for the fruitful discussions in the course of this work and for providing the restored images by the method in [21].
Figure 2a: Original “Lena” image.

Figure 2b: Degraded “Lena” image with $7 \times 7$ uniform blur and additive noise SNR=$25.36$ dB.

Figure 2c: Wiener filter restoration, ISNR = $3.2$ dB.

Figure 2d: “Optimal” Wiener filter restoration, ISNR = $4.40$ dB.
Figure 2c: Stationary restoration, ISNR = 4.25 dB.

Figure 2f: CLS method (adaptive smoothness constraint) restoration, $\theta=1000$, $a=0.01$, ISNR = 4.65 dB.

Figure 2g: Restoration with GMRF algorithm [21], ISNR = 3.46 dB.

Figure 2h: MAP “optimal” non stationary restoration, ISNR=10.43 dB, $l=2.01$. 
Figure 2i: MAP non stationary restoration, $\text{ISNR} = 5.63\ dB, l=2.2$.

Figure 2j: Bayesian non stationary restoration, $\text{ISNR} = 5.22\ dB, l=2.2$. 
Figure 3a: Degraded “cameraman” image with 9x9 uniform blur and additive noise SNR=38.64 dB, ($\beta^1=0.308$).

Figure 3b: Stationary restoration, ISNR = 6.88 dB.

Figure 3c: CLS method (adaptive smoothness constraint) restoration, $\theta=0.05$, $a=0.003$, ISNR = 7.38 dB.

Figure 3d: MAP non stationary restoration, ISNR = 9.42 dB, $l=2.1$. 
APPENDIX: Gradient and Hessian Calculations.

In this Appendix the equations for computing the gradient vector and the Hessian matrix of the function $J_g$ are derived. Without loss of generalization the analytical derivation for the gradient and Hessian is given only for one orientation in Eq. (A.3) and (A.4) (hence the relevant notation has been omitted). Using these equations the analytical formulas for all the directions are given in Eq. (A.5) and (A.6). The gradient of $J_g$ with respect to $f$ is by:

\[
\frac{\partial}{\partial f(k)} \left( \log p(g | f, \beta) - \frac{4}{N} \sum_{n=1}^{N} \left( \frac{l_n}{2} + \frac{1}{8} \right) \log \prod_{i=1}^{N} \left( m_n (l_n - 2) + \frac{1}{2} \left[ \frac{Q_f}{Q_f} \right](i)^2 \right) \right) = 0 \tag{A.1}
\]

Evaluation of the first term is straightforward, since the objective function is quadratic:

\[
\nabla \log p(g | f, \beta) = -\beta H^T (Hf - g) \tag{A.2}
\]

The second term needs more elaboration. Let $h_Q$ be a function $R^N \rightarrow R$, defined by any $N \times N$ matrix $Q$, with element $q_{ij}$, such that:

\[
h_Q(f) = \sum_{i=1}^{N} g(\left[Q_f\right](i)) ,
\]

where $g(\cdot)$ is any $R \rightarrow R$ function, and $\left[Q_f\right](i)$ is the $i$-th element of the $N \times 1$ vector $Q_f$.

Taking the derivative with respect to $f(k)$, yields:

\[
\frac{\partial}{\partial f(k)} \sum_{i=1}^{N} h_Q(f) = \sum_{i=1}^{N} g(\left[Q_f\right](i)) \frac{\partial}{\partial f(k)} \left[Q_f\right](i) = \sum_{i=1}^{N} \frac{\partial}{\partial f(k)} \left[Q_f\right](i) \frac{\partial}{\partial f(k)} = \sum_{i=1}^{N} \frac{\partial}{\partial f(k)} \left[Q_f\right](i) \frac{\partial}{\partial f(k)} \left[Q_f\right](i)
\]

\[
\sum_{i=1}^{N} \frac{\partial}{\partial f(k)} \left[Q_f\right](i) = \sum_{i=1}^{N} Q(i, k) g^{(i)}(\left[Q_f\right](i)) \Rightarrow \nabla h_Q(f) = Q^T \mathbf{v}
\]

where $\mathbf{v}$ an $N \times 1$ vector with elements $v(i) = g^{(i)}(\left[Q_f\right](i))$, and $g^{(i)}(\cdot)$ the first derivative of $g(\cdot)$. When this formula is applied to the second term of (A.1) we find that:

\[
\nabla h_Q(f) = Q^T \mathbf{v}
\]
\[
v(i) = \frac{[Qf](i)}{m(l-2) + \frac{1}{2}([Qf](i))^2}, \quad i = 1, 2, \ldots, N.
\] (A.4)

Similarly, using the same technique that was used to derive (A.4), it can be shown that the \((k,n)\) element of the Hessian, is given by:

\[
\left[ \nabla^2 h_Q (f) \right]_{k,n} = \frac{\partial^2 h_Q (f)}{\partial f(k) \partial f(n)} = \sum_{i=1}^{N} \frac{Q(i,k)Q^{(i)}([Qf](i))}{\partial f(n)} = \sum_{i=1}^{N} \frac{Q(i,k)g^{(i)}([Qf](i))}{\partial f(n)} = \sum_{i=1}^{N} Q(i,k)Q(i,n)g^{(2)}([Qf](i)) = Q^T \Lambda Q,
\]

where \(\Lambda\) a diagonal matrix with elements equal to the second derivative of \(g(\cdot)\):

\[
\Lambda(i,i) = g^{(2)}([Qf](i)).
\]

In our case the \(g(\cdot)\) function is:

\[
g([Qf](i)) = \log\left(m_i (l_i - 2) + \frac{1}{2}([Qf](i))^2\right).
\]

This result can be applied for each of the four directional difference operators. Thus, from (A.2) and (A.3) the Gradient and Hessian, are given by:

\[
\nabla_i J_\beta = \beta H^T (Hf - g) + \sum_{i=1}^{4} \left( \frac{l_i}{2} + \frac{1}{8} \right) (Q^i)^T v_k
\] (A.5)

\[
\nabla_i^2 J_\beta = \beta H^T H + \sum_{i=1}^{4} \left( \frac{l_i}{2} + \frac{1}{8} \right) (Q^i)^T \Lambda_i Q^i
\] (A.6)

where:

\[
v_k(i) = \frac{[Q^i f](i)}{m_i (l_i - 2) + \frac{1}{2}([Q^i f](i))^2}, \quad i = 1, 2, \ldots, N,
\] (A.7)

and the matrix \(\Lambda_i\) is diagonal with diagonal elements \(\lambda_i(i)\) equal to:

\[
\lambda_i(i) = g^{(2)}([Q^i f](i)) = \frac{1}{m(l-2) + \frac{1}{2}([Q^i f](i))^2} - [v_k(i)]^2, \quad i = 1, \ldots, N.
\]
REFERENCES


