

LOCAL BAYESIAN IMAGE RESTORATION USING VARIATIONAL METHODS AND GAMMA-NORMAL DISTRIBUTIONS

Javier Mateos¹, Tom E. Bishop², Rafael Molina¹, Aggelos K. Katsaggelos³

¹Dpto. Ciencias de la Computación e I. A. University of Granada. Spain.

²Dept. of Engineering & Physical Sciences. Heriot-Watt University. United Kingdom.

³Dept. of Electrical Engineering and Computer Science, Northwestern University, Illinois, USA.

ABSTRACT

In this paper we present a new Bayesian methodology for the restoration of blurred and noisy images. Bayesian methods rely on image priors that encapsulate prior image knowledge and avoid the ill-posedness of image restoration problems. We use a spatially varying image prior utilizing a Gamma-Normal hyperprior distribution on the local precision parameters. This kind of hyperprior distribution, which to our knowledge has not been used before in image restoration, allows for the incorporation of information on local as well as global image variability, models correlation of the local precision parameters and is a conjugate hyperprior to the image model used in the paper. The proposed restoration technique is compared with other image restoration approaches, demonstrating its improved performance.

Index Terms— Image restoration, Variational methods, Bayes procedures, Gamma-Normal distributions

1. INTRODUCTION

A standard formulation of the image degradation model is given in lexicographic form by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (1)$$

where the $P \times 1$ vectors \mathbf{x} , \mathbf{y} , and \mathbf{n} represent, respectively, the original image, the observed noisy and blurred image, and the white Gaussian noise with independent elements of variance $\sigma_n^2 = \beta^{-1}$, and \mathbf{H} represents the known blurring matrix. The images are assumed to be of size $m \times n$, with $P = m \times n$. The restoration problem considered here is to find an estimate of \mathbf{x} from \mathbf{y} and \mathbf{H} and some knowledge about \mathbf{n} and possibly \mathbf{x} .

Bayesian image restoration techniques are based on the introduction of a prior image model on \mathbf{x} whose aim is to encapsulate our prior image knowledge and consequently to avoid the ill-posedness of the image restoration problem. Image prior models like the ones based on Simultaneous Auto-Regression (SAR) [1], Conditional Auto-Regression (CAR) [1] or Total Variation (TV) [2] depend on a global parameter which is related to the global variance of the image model. However, for most images it is not realistic to assume that the variance of the model is the same for the whole image and consequently it should be adapted to its local characteristics.

One such spatially varying image prior model was defined recently in [3]. This hierarchical model represents the local image

This work has been supported by the “Comisión Nacional de Ciencia y Tecnología” under contract TIN2007-65533 and the Consejería de Innovación, Ciencia y Empresa of the Junta de Andalucía under contract P07-TIC-02698.

activity and discontinuities with a continuous valued set of hidden variables. These variables represent the local precision (inverse variance) of the spatial image derivatives.

In this paper we propose to extend the model of [3], improving the estimation of the precision variables. In [3], these variables are assumed to be independent. This means that the strength of an edge or feature at one pixel in the image is not related to that at neighboring pixels, and large local restoration errors may result in the presence of noise. In fact, we can do better by using our prior knowledge that, in both smooth and edge regions, these precision variables tend to be highly correlated. As we will see, this allows for a larger degree of nonstationary without the occurrence of local singularities.

To do this, we use a Gamma-Normal hyperprior distribution on the local precision parameters. This kind of hyperprior distribution [4, 5, 6], which as far as we know has not been used before in image restoration, allows for the incorporation of information on local as well as global image variability and is a conjugate hyperprior to the image model used in the paper. The advantage of such a model over the model in [3] is to not only enforce smoothness in smooth regions, but to also impose our knowledge that where there are edges we expect to find other edges correlated with these nearby.

The paper is organized as follows. In section 2 we discuss the local image model as well as the degradation model and the hyperparameter model we are going to use. Then the inference, that is, the estimation of the restored image and the parameters, is performed in section 3. Experimental results are described in section 4 and, finally, section 5 concludes the paper.

2. BAYESIAN MODELING

The Bayesian formulation of our image restoration problem requires the definition of the joint distribution $p(\Omega, \mathbf{x}, \mathbf{y})$ of the observation \mathbf{y} , the unknown image \mathbf{x} , and the hyperparameters Ω (which obviously depend on the model used). Then, the posterior distribution of the unknowns given the observed image $p(\Omega, \mathbf{x}|\mathbf{y})$ has to be calculated and used to estimate the unknown image and hyperparameters.

Given the degradation model of Eq. (1), the distribution of the observed image \mathbf{y} given \mathbf{x} is

$$p(\mathbf{y}|\mathbf{x}) \propto \beta^{P/2} \exp \left[-\frac{1}{2}\beta \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \right], \quad (2)$$

where we are considering that the noise parameter β is either known or can be estimated from the available data. In the experimental section we will propose a method for estimating it.

Following the approximation in [3] which extends the Conditional Autoregressive (CAR) model (see [1]) to take into account

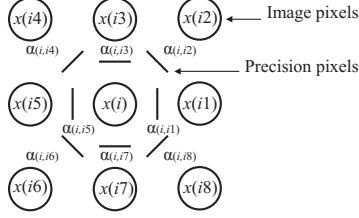


Fig. 1. Image pixel and precision pixels notation.

local variability we write

$$p(\mathbf{x}|\boldsymbol{\alpha}) \propto \prod_{i=1}^P \prod_{l=1}^4 \alpha_{(i,il)}^{1/8} \exp \left\{ -\frac{\alpha_{(i,il)}}{2} \frac{(x(i) - x(il))^2}{8} \right\}, \quad (3)$$

where $\alpha_{(i,il)}$ controls the smoothness of the restoration between image pixels i and il (see Fig. 1) and

$$\boldsymbol{\alpha} = [\alpha_{(i,il)} \mid i = 1, \dots, P, \quad l = 1, \dots, 4]. \quad (4)$$

Note that if $\alpha_{(i,il)} = \alpha, \forall(i, il)$ we have the (global) CAR prior model. The parameters $\alpha_{(i,il)}$ are precision parameters since they correspond to inverse of variances. Note that this amounts to introducing $4P$ parameters whose value has to be set or estimated.

We now proceed to define a hyperprior on the hypervector $\boldsymbol{\alpha}$ which aims to incorporate prior knowledge on the value of the precision parameters $\alpha_{(i,il)}$ and, thus, effectively reduces the number of parameters to be set. The chosen hyperprior should account for local as well as global variability. A large part of the Bayesian literature is devoted to finding hyperprior distributions $p(\boldsymbol{\alpha})$ for which $p(\boldsymbol{\alpha}, \mathbf{x}|\mathbf{y})$ can be calculated in a straightforward way or can be approximated. These are the so called conjugate priors [7]. Conjugate priors have, as we will see later, the intuitive feature of allowing one to begin with a certain functional form for the prior and end up with a posterior of the same functional form, but with the parameters updated by the sample information.

Taking the above considerations about conjugate priors into account, we propose the definition and use of the following distribution on the hyperparameter vector $\boldsymbol{\alpha}$, termed the Gamma-Normal distribution

$$p(\boldsymbol{\alpha} \mid \lambda, a, c) \propto \prod_{i=1}^P \prod_{l=1}^4 [\alpha_{(i,il)}]^a \exp[-c\alpha_{(i,il)}] \times \exp \left[-\frac{\lambda}{2} \sum_{(i,il)} \sum_{(i',il')} \frac{(\alpha_{(i,il)} - \alpha_{(i',il')})^2}{N_{il}} \right], \quad (5)$$

$\forall \alpha_{(i,il)} > 0$, where $\lambda > 0, c > 0$, and $a > 0$, and (i, il) and (i', il') are indices of neighboring precision pixels and N_{il} is the number of neighboring precision pixels of pixel i in the direction l . Different neighborhood systems can be defined but, in this paper, we use the four nearest precision pixels to the precision pixel (i, il) in its same direction (see Fig. 2).

Let us before proceeding analyze in some depth this distribution. Gamma-Normal distributions extend the definition of the Rectified Gaussian distribution [8], a modification of the standard Gaussian distribution in which the variables are constrained to be positive, by considering the distribution in Eq. (5) as the product of a Rectified Gaussian distribution

$$p_1(\boldsymbol{\alpha}|\lambda) \propto \exp \left[-\frac{\lambda}{2} \sum_{(i,il)} \sum_{(i',il')} \frac{(\alpha_{(i,il)} - \alpha_{(i',il')})^2}{N_{il}} \right], \quad (6)$$

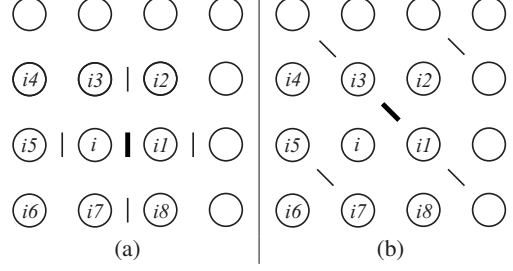


Fig. 2. Examples of neighborhood system of a precision pixels (in bold). (a) for $\alpha_{(i,i1)}$; (b) for $\alpha_{(i,i2)}$.

$\forall \alpha_{(i,il)} > 0$, and P independent Gamma distributions

$$p_2(\boldsymbol{\alpha}|a, c) \propto \prod_{i=1}^P \prod_{l=1}^4 [\alpha_{(i,il)}]^a \exp[-c\alpha_{(i,il)}] \quad (7)$$

and consequently its domain is $\boldsymbol{\alpha} > 0$.

From its definition we have

$$p(\boldsymbol{\alpha} \mid \lambda, a, c) \propto p_1(\boldsymbol{\alpha}|\lambda)p_2(\boldsymbol{\alpha}|a, c) \quad (8)$$

with

$$\int_{\boldsymbol{\alpha} > 0} p(\boldsymbol{\alpha} \mid \lambda, a, c) d\boldsymbol{\alpha} = 1. \quad (9)$$

That is, $p(\boldsymbol{\alpha} \mid \lambda, a, c)$ is not an improper distribution.

Furthermore, given the neighboring precision pixels of a given one, the mode of the corresponding conditional distribution from $p_1(\boldsymbol{\alpha}|\lambda)$ is the mean of the neighboring precision pixels which we term $\alpha_{(i,il)}$, while for a given precision pixel the mode of its distribution from $p_2(\boldsymbol{\alpha}|a, c)$ is a/c . This indicates as we will show more formally later that this prior distribution can be used to model fidelity to a (global) given value (a/c) as well as to the (local) mean values of neighboring pixels.

The new hyperprior model defines a positive-constrained Markov random field (MRF) over the precision variables, with the amount of correlation controlled by λ . Thus, even though the local image differences $(x(i) - x(il))$ are uncorrelated, which aids tractability, the continuity of edge features is maintained through the hidden variables $\boldsymbol{\alpha}$. The advantage of this model is to ensure that the values of $\alpha_{(i,il)}$ remain large in smooth regions, suppressing noise, and do not drop unboundedly at sharp features, but take on locally consistent values.

3. BAYESIAN INFERENCE

The Bayesian paradigm dictates that inference on $(\boldsymbol{\alpha}, \mathbf{x})$ should be based on

$$p(\boldsymbol{\alpha}, \mathbf{x}|\mathbf{y}) = \frac{p(\boldsymbol{\alpha}, \mathbf{x}, \mathbf{y})}{p(\mathbf{y})} = \frac{p(\boldsymbol{\alpha})p(\mathbf{x}|\boldsymbol{\alpha})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}, \quad (10)$$

where $p(\boldsymbol{\alpha})$, $p(\mathbf{x}|\boldsymbol{\alpha})$ and $p(\mathbf{y}|\mathbf{x})$ have been defined in Eqs. (5), (3), (2), respectively.

We can not calculate this posterior distribution in closed form and therefore we apply variational methods to approximate it by the distribution $q(\boldsymbol{\alpha}, \mathbf{x})$. The variational criterion used to find $q(\boldsymbol{\alpha}, \mathbf{x})$ is the minimization of the Kullback-Leibler divergence, given by [9,

10]

$$\begin{aligned} C(q(\alpha, \mathbf{x})||p(\alpha, \mathbf{x}|\mathbf{y})) &= \int_{\alpha, \mathbf{x}} q(\alpha, \mathbf{x}) \log \left(\frac{q(\alpha, \mathbf{x})}{p(\alpha, \mathbf{x}|\mathbf{y})} \right) d\alpha d\mathbf{x} \\ &= \int_{\alpha, \mathbf{x}} q(\alpha, \mathbf{x}) \log \left(\frac{q(\alpha, \mathbf{x})}{p(\alpha, \mathbf{x}, \mathbf{y})} \right) d\alpha d\mathbf{x} + \text{const}, \end{aligned} \quad (11)$$

which is always non negative and equal to zero only when $q(\alpha, \mathbf{x})$ and $p(\alpha, \mathbf{x}|\mathbf{y})$ coincide.

We will approximate $p(\alpha, \mathbf{x}|\mathbf{y})$ by the distribution $q(\alpha, \mathbf{x}) = \mathbf{z}_D(\alpha)\mathbf{z}_D(\mathbf{x})$, where we assume that both $\mathbf{z}_D(\alpha)$ and $\mathbf{z}_D(\mathbf{x})$ are degenerate distributions (they take one value with probability one and the rest with probability zero). As we will now see, the corresponding algorithm is equivalent to maximizing $p(\alpha, \mathbf{x}, \mathbf{y})$ alternatively in the hyperparameters and image. In other words, the estimation procedure is an iterated conditional mode (ICM) algorithm [11]. The corresponding optimization algorithm is as follows

Algorithm 1 *Local image restoration algorithm*

Given α^1 , an initial estimate of the hyperparameter precision vector where the distribution $\mathbf{z}_D(\alpha)$ is degenerate,

for $k = 1, 2, \dots$ until a stopping criterion is met:

1. Find \mathbf{x}^k , an estimate of the image where the distribution $\mathbf{z}_D(\mathbf{x})$ is degenerate by solving

$$\mathbf{x}^k = \arg \max_{\mathbf{x}} p(\alpha^k, \mathbf{x}, \mathbf{y}). \quad (12)$$

2. Find α^{k+1} , an estimate of the hyperparameter precision vector where the distribution $\mathbf{z}_D(\alpha)$ is degenerate by solving

$$\alpha^{k+1} = \arg \max_{\alpha} p(\alpha, \mathbf{x}^k, \mathbf{y}). \quad (13)$$

To explicitly calculate \mathbf{x}^k we have to solve

$$\mathbf{x}^k = \arg \min_{\mathbf{x}} \hat{\beta} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \sum_{i=1}^P \sum_{l=1}^4 \alpha_{(i,il)}^k \frac{(x(i) - x(il))^2}{8}, \quad (14)$$

which can be solved, for instance, by Gradient Descent or Conjugate Gradient methods.

Let us now examine the calculation of α^{k+1} . Given a precision pixel position (i, il) , differentiating $-\log p(\alpha, \mathbf{x}^k, \mathbf{y})$ with respect to $\alpha_{(i,il)}$ we obtain

$$\begin{aligned} - \frac{\partial \log p(\alpha, \mathbf{x}^k, \mathbf{y})}{\partial \alpha_{(i,il)}} &= \lambda(\alpha_{(i,il)} - \bar{\alpha}_{(i,il)}) - \frac{a}{\alpha_{(i,il)}} \\ &\quad - \frac{1}{8} \frac{1}{\alpha_{(i,il)}} + c + \frac{1}{8} \frac{(x^k(i) - x^k(il))^2}{2}, \end{aligned} \quad (15)$$

where $\bar{\alpha}_{(i,il)}$ denotes the mean of the neighboring precision pixels of $\alpha_{(i,il)}$. Setting now this derivative equal to zero and multiplying by $\alpha_{(i,il)}$ we obtain

$$\begin{aligned} \alpha_{(i,il)} &\times \left[\lambda \alpha_{(i,il)} + c + \frac{1}{8} \frac{(x^k(i) - x^k(il))^2}{2} \right] \\ &= \lambda \alpha_{(i,il)} \bar{\alpha}_{(i,il)} + a + \frac{1}{8}, \end{aligned} \quad (16)$$

or

$$\begin{aligned} \alpha_{(i,il)} &= \mu_{\mathbf{x}^k, \alpha}(i, il) \bar{\alpha}_{(i,il)} + (1 - \mu_{\mathbf{x}^k, \alpha}(i, il)) \\ &\times \left[\psi_{\mathbf{x}^k}(i, il) \frac{a}{c} + (1 - \psi_{\mathbf{x}^k}(i, il)) \frac{1}{\frac{(x^k(i) - x^k(il))^2}{2}} \right], \end{aligned} \quad (17)$$

noise variance (SNR)	method in [12]	method in [3]	proposed method
0.56 (45dB)	5.13dB	7.53dB	8.03dB
5.6 (35dB)	3.41dB	5.09dB	5.45dB

Table 1. ISNR for different restorations.

where

$$\mu_{\mathbf{x}^k, \alpha}(i, il) = \frac{\lambda \alpha_{(i,il)}}{\left[\lambda \alpha_{(i,il)} + c + \frac{1}{8} \frac{(x^k(i) - x^k(il))^2}{2} \right]}, \quad (18)$$

and

$$\psi_{\mathbf{x}^k}(i, il) = \frac{c}{c + \frac{1}{8} \frac{(x^k(i) - x^k(il))^2}{2}}. \quad (19)$$

Equation (17) suggests an iterative process where we obtain the new estimation of $\alpha_{(i,il)}$, $\alpha_{(i,il)}^{k+1}$, in the left hand side using the previous estimation, $\alpha_{(i,il)}^k$, in the right hand side. This is a very interesting equation since it tells us that $\alpha_{(i,il)}^{k+1}$ is a linear convex combination of the mean of its neighboring precision pixels $\bar{\alpha}_{(i,il)}^k$, the mode of the gamma distribution defining the Gamma-Normal distribution a/c and the maximum likelihood estimate of $\alpha_{(i,il)}$ from $p(\mathbf{x}^k|\alpha)$, that is, $1/[(x^k(i) - x^k(il))^2/2]$.

4. EXPERIMENTAL RESULTS

A number of experiments have been performed with the proposed algorithm. Here we report results on the *cameraman* image. The original image was blurred with an uniform 7×7 blur and Gaussian noise of variance $\sigma_n^2 = 5.6$ and $\sigma_n^2 = 0.56$ was added resulting in two degraded images with a SNR of 35dB and 45dB, respectively. The criterion $\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 / \|\mathbf{x}^k\|^2 \leq 10^{-6}$ was used for terminating the iteration. In all experiments, no more than 21 iterations were needed for the algorithm to converge. The performance of the restoration algorithm was evaluated by measuring the improvement in signal to noise ratio defined by $ISNR = 10 \log_{10} [\|\mathbf{x} - \hat{\mathbf{x}}\|^2 / \|\mathbf{x} - \mathbf{y}\|^2]$, where $\hat{\mathbf{x}}$ is the restored image.

In order to run the proposed method, a set of parameters need to be set. More specifically, we need to provide values for β , a , c and λ . Since the proposed model is derived from the global CAR image model, the method using the CAR image model in [12] using a flat hyperprior on the hyperparameters was used for estimating the noise parameter β . This method amounts to estimating the restored image by taking the MAP of the posterior distribution while the parameters are estimated by maximum likelihood, which provides a very precise estimation for the noise parameter β . This method, also provides a value for the global image model parameter α that can be used as an estimate for a/c in the proposed model. Initially, we used $\alpha_{(i,il)} = \alpha, \forall i, il$. Similarly to [3], we used $a = 0.01$ and $c = a/\alpha$. This means that the values of α should be quite similar to the global image parameter α except of the objects boundaries. To enforce continuity of the boundaries we chose $\lambda = 1e5$. We note that choosing $\lambda = 0$ corresponds to the method described in [3].

We compared the proposed method with the method using a CAR image model in [12] and the method in [3]. Results are summarized in Table 1. From this table, it is clear that the proposed method provides much better results than the method in [12], resulting in an increase of the ISNR for up to 2.9dB, and better results than the method in [3]. To visually illustrate the performance of the method, the experiment just described in the case of $\sigma_n^2 = 0.56$ is shown in Fig. 3, comparing the three methods.

From the images it is clear that the proposed method provides a better restoration with crisper edges, no noticeable ringing and

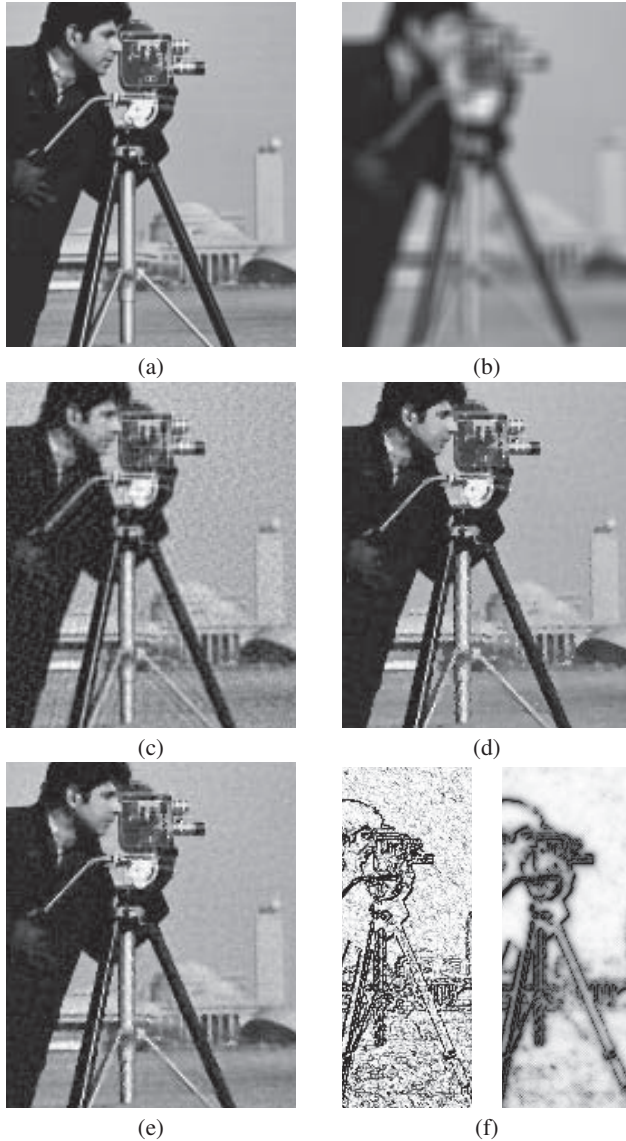


Fig. 3. a) Original cameraman image; (b) Degraded image with 7×7 uniform blur and additive noise of variance 0.56; (c) Restoration using the global image model based method in [12], $ISNR = 5.16\text{dB}$; (d) Restoration using the method in [3], $ISNR = 7.53\text{dB}$; (e) Restoration using the proposed method, $ISNR = 8.03\text{dB}$; (f) Detail of α in the direction $l = 4$ for the method in [3] (left) and the proposed method (right).

much less noise artifacts than the method in [12]. Compared to the method in [3] we can see that enforcing continuity removes the over-smoothing into piecewise constant regions in smooth areas (such as the face), better defines faint edges (e.g., in the buildings), and also ameliorates the presence of artifacts we term “hot” and “cold” pixels. These are isolated black or white pixels that appear in the restorations, uncorrelated with their neighbors, at locations where $\alpha_{(i,il)}$ is estimated uncharacteristically low. The appearance of these pixels can be self-reinforcing in the ICM procedure; a hot or cold pixel will cause a large local variance (small $\alpha_{(i,il)}$), which can then increase the hot or cold pixels’ visibility. Using a larger value of λ means

that the process is more stable as neighbors of each $\alpha_{(i,il)}$ are taken into account. We think that these pixels can be removed by a better selection of the parameters and we are working in that direction. Figure 3f illustrates the difference in the precision parameters map. Clearly, the proposed method (right) obtains a much less noisy estimate and better continuity of the objects edges than the method in [3] (left), due to the enforcement of the correlation between precision pixels. The proposed algorithm took 280 seconds to run in a Pentium IV 3400, a time similar to the one employed by the method in [3]. The method in [12] took only 1.9 seconds.

5. CONCLUSIONS

In this paper we have presented a new Bayesian restoration method which takes into accounts the local features in the image by the mean of the prior distribution. This prior distribution depends on a large set of parameters and a Gamma-Normal distribution, that allows for the incorporation of information on local as well as global image variability, is used as hyperprior on them, effectively reducing the number of parameters that need to be set. The proposed restoration technique has been compared with other image restoration approaches and its performance has been experimentally demonstrated. The proposed method provides better visual and numerical results and ameliorates the presence of hot and cold pixels. We think those pixels can be completely removed by a better selection of the parameters and we are currently working on this problem.

6. REFERENCES

- [1] B. D. Ripley, *Spatial Statistics*, pp. 88–90, John Wiley, 1981.
- [2] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D*, pp. 259–268, 1992.
- [3] G. K. Chantas, N. P. Galatsanos, and A. C. Likas, “Bayesian restoration using a new nonstationary edge-preserving image prior,” *IEEE Trans. Image Processing*, vol. 15, pp. 2987–2997, 2006.
- [4] F. I. Toranzos, “An asymmetric bell-shaped frequency curve,” *The Annals of Math. Stats.*, vol. 23, no. 3, pp. 467–469, 1952.
- [5] J. Del Castillo and P. Puig, “Testing departures from Gamma, Rayleigh and truncated normal distributions,” *Annal. Inst. Stat. Math.*, vol. 49, no. 2, pp. 255–269, 1997.
- [6] N. V. Hartvig, “Simulation of the Gamma-Normal distribution,” 2000, part of PhD thesis, Aarhus Universitet.
- [7] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, chapter 3 and 4, New York, Springer Verlag, 1985.
- [8] N. D. Socci, D. D. Lee, and H. S. Seung, “The rectified Gaussian distribution,” in *Advances in Neural Information Processing Systems 10*, 1998, pp. 350–356.
- [9] S. Kullback and R. A. Leibler, “On information and sufficiency,” *Annals of Math. Statistics*, vol. 22, pp. 79–86, 1951.
- [10] S. Kullback, *Information Theory and Statistics*, New York, Dover Publications, 1959.
- [11] J. Besag, “On the statistical analysis of dirty pictures,” *Journal of the Royal Stat. Soc. Series B*, vol. 48, pp. 259–302, 1986.
- [12] R. Molina, A. K. Katsaggelos, and J. Mateos, “Bayesian and regularization methods for hyperparameter estimation in image restoration,” *IEEE Trans. Image Processing*, vol. 8, no. 2, pp. 231–246, 1999.