Shape Parameter Estimation for Generalized Gaussian Markov Random Field Models used in MAP Image Restoration

Wai Ho Pun and Brian D. Jeffs
Department of Electrical and Computer Engineering, Brigham Young University
459 CB, Provo, UT 84602, USA. E-mail: bjeffs@ee.byu.edu

Abstract

In this paper, we propose using the Generalized Gaussian Markov Random Field (GGMRF) image model with MAP estimation to solve the problem of restoration for a blurred and noise corrupted image. The restoration algorithm is adapted to specific characteristics of the true image by estimating the GGMRF shape parameter used in computing the MAP estimation. This shape parameter, p, is estimated based on the sample kurtosis of the image. It is shown that higher quality restorations are obtained when the estimated p value is used, rather than some arbitrary choice as other investigators have used.

I. Introduction

IN this paper we introduce a Bayesian approach to image restoration that uses the observed image data to control the form of the image prior model. The image is modeled as a Generalized Gauss Markov random field (GGMRF) [1], but unlike previous approaches, we assume the corresponding shape parameter is unknown. Both the true image, and the image prior shape parameter are estimated from the degraded observation. This joint estimation is difficult because the blurring process masks the structure of the true image and can bias the shape parameter estimate towards a model that favors smoother images with less edge detail. The proposed approach overcomes this problem, and can yield better results than other Bayesian methods where the image prior model is fixed arbitrarily. Such methods cannot adapt to the wide range of image structural forms regularly encountered in restoration applications. The adaptive shape parameter represents a degree of uncertainty about the true image structure, which we resolve using information from the observed degraded image alone.

We adopt the following familiar linear observation model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{y}$ is the row scanned degraded image vector, $\mathbf{H}$ is the convolutional blur matrix, $\mathbf{x}$ is the true image vector, and $\mathbf{n}$ is additive noise. We seek an estimate of $\mathbf{x}$ that maximizes the posterior distribution.

$$\begin{aligned} \hat{\mathbf{x}} &= \arg\{\max_{\mathbf{x},p} f_x(\mathbf{x}|\mathbf{y};p)\} \\ &= \arg\{\max_{\mathbf{x},p}\{\log f_y(\mathbf{y}|\mathbf{x}) + \log f_x(\mathbf{x};p)\}\}. \end{aligned} \tag{2}$$

where $f(\cdot)$ represents a probability density function, and p is the deterministic shape parameters which affect the distribution of $\mathbf{x}$.

II. Image Prior Model

To model an image $\mathbf{x}$ as a Markov random field (MRF), the prior distribution is usually specified through the use of a Gibb's distribution, whose energy function has the general form [2]

$$U(\mathbf{x}) = \sum_{\{s,t\}\in\mathcal{C}} \phi(x_s - x_t), \tag{3}$$

where $\phi(\cdot)$ is the potential function and $\mathcal{C}$ is the set of all cliques defined for the neighborhood system of the Markov field. Many potential functions have been proposed in the literature to be applied to various problems. Table I gives a representative sample of these functions. Figure 1 shows the shape of each function.

n	$\phi_n(x)$ function	Authors
1	$\lvert x\rvert$	J. Besag [3]
2	$\begin{cases} x^2-1 & \text{if } x^2 \le 1 \\ 0 & \text{otherwise} \end{cases}$	A. Blake and A. Zisserman
3	$\frac{-1}{1+\lvert x\rvert}$	D. Geman and G. Reynolds
4	$\frac{1}{1+x^2}$	S. Geman and D. McClure
5	$\log\cosh(x)$	P. J. Green
6	$\log(1+x^2)$	T. Hebert et al.
7	$\begin{cases} 0 & \text{if } \lvert x\rvert = 0 \\ 1 & \text{if } \lvert x\rvert > 0 \end{cases}$	Y. G. Leclerc
8	x^2	P. A. P. Moran

TABLE I
Some representative MRF potential functions

The potential function $\phi_8(x) = x^2$ corresponds to a Gauss-Markov random field (GMRF). When used as an image prior for MAP restoration, this model is known to produce results that are over-smoothed or excessively noisy [1]. The other listed potential functions were devised to overcome these problems, usually putting higher emphasis on edges in the image

1058-6393/96 $5.00 © 1996 IEEE
Proceedings of ASILOMAR-29

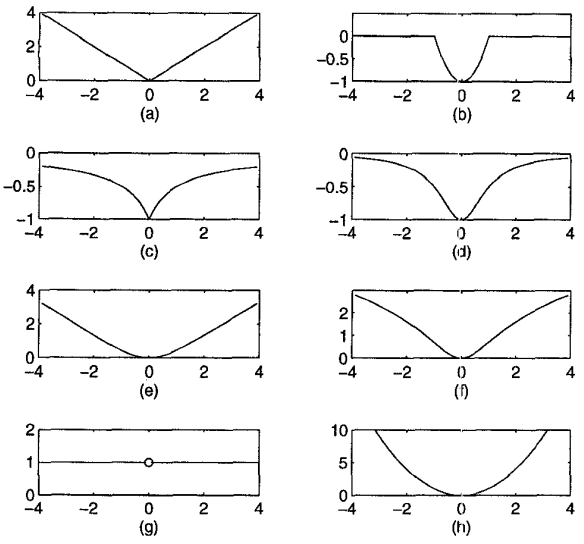

Fig. 1. Plots of different energy functions. a) $|x|$, b) $x^2 - 1$ if $x^2 \leq 1$, 0 otherwise, c) $\frac{1}{1+|x|}$, d) $\frac{1}{1+x^2}$, e) $\log\cosh(x)$, f) $\log(1+x^2)$, g) 0 if $|x| = 0$, 1 if $|x| > 0$, h) x^2.

they model to reduce over-smoothing. However, each function is tailored to fit a particular class of images and none can guarantee to give the best performance for all real-life images. Therefore, we seek a flexible prior model that will fit a variety of image structures and can adapt itself to the observed image data to produce the best possible results under the assumptions of that model. The generalized Gaussian Markov random field introduced by Bouman and Sauer [1] is a family of MRF's with a single parameter p controlling the shape of its distribution. Adopting the GGMRF as an image prior model, we have

$$f_x(\mathbf{x};p) \propto \exp\left\{-\left[\frac{A}{\sigma}\right]^p \sum_{\{s,t\}\in\mathcal{C}} |x_s - x_t|^p\right\}, p > 0. \quad (4)$$

where $A = \left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{1/2}$ and σ is the standard deviation parameter. If ∂_s represents the neighborhood of t and S represents the entire set of lattice points in the image, then the conditional distribution of x_s given all *other* pixel values in S depends only on ∂_s, and can be expressed as

$$f_{x_s}(x_s | x_t \in \partial_s) \propto \exp\left\{-\left[\frac{A}{\sigma}\right]^p \sum_{t\in\partial_s} |x_s - x_t|^p\right\}. \quad (5)$$

These priors are improper because the exponent is negative semi-definite instead of negative definite, coupled with the fact that the partition function or normalizing term is difficult to compute due to high dimensionality. However, the posterior distribution, on which we base our inferences, is typically proper for any reasonable choice of the potential function.

The potential of the GGMRF has the form $\phi(x) = |x|^p$. Figure 2 shows plots of the potential for various values of p. Notice that by changing p, the potential assumes different shapes. This dictates what the structure of the image should be. For example, small value of p models images that are dominated by abrupt edges whereas large value of p models images that are dominated by smooth areas.

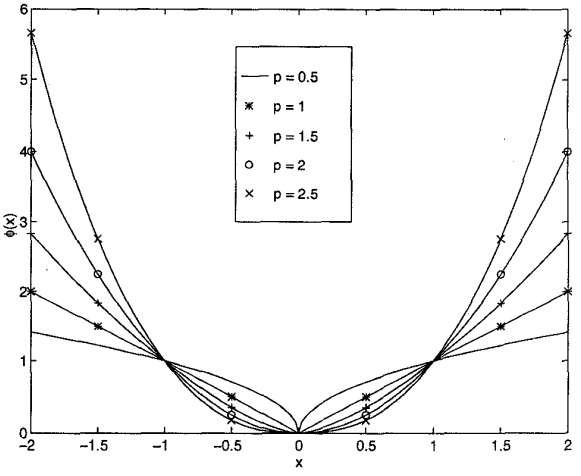

Fig. 2. Potential function of GGMRF for different values of p.

Also note that when $p = 0, 1$, or 2, the respective GGMRF's correspond to the potential functions ϕ_7, ϕ_1, ϕ_8 in Table I; and for some values of p, the potential of the GGMRF resembles the shape of other potentials in the same table. Thus the shape parameter p can be used to cover a wide range of possible distributions.

III. Solution Formulation

Given an observed image $\mathbf{y}$, and assuming the additive noise is Gaussian, a general energy function family for estimating the original image $\mathbf{x}$ is as follows

$$\Phi(\mathbf{x},\mathbf{y};p) = \sum_{s\in S}[y_s - (\mathbf{Hx})_s]^2 + \gamma\left[\frac{A}{\sigma}\right]^p \sum_{\{s,t\}\in\mathcal{C}} \phi(x_s, x_t; p), \quad (6)$$

where $\mathcal{C}$ contains all nearest neighbor pairs, and $\phi(x_s, x_t; p) = |x_s - x_t|^p$ for the GGMRF image prior model. γ is then the regularization parameter controlling the relative influence on the solution from the image prior model relative to the restoration error term. The adaptive MAP estimate of $\mathbf{x}$ is the minimizer of $\Phi(\mathbf{x},\mathbf{y};p)$; thus

$$\hat{\mathbf{x}} = \arg\{\min_{\mathbf{x},p} \Phi(\mathbf{x},\mathbf{y};p)\}, \quad (7)$$

which is essentially equation (2).

IV. Parameter Estimation

The utility of the GGMRF image prior model lies in its ability to change the shape of its distribution, and

thus the regularization structural form, by changing only the parameter p. In order to exploit this flexibility however, we must be able to estimate p from our observation. In this section, we establish a key relationship between a GGMRF and the generalized Gaussian (GG) distribution, and present a viable method based on this relationship to estimate p for any given GGMRF image.

An inspection of (4) suggests that direct estimation of p from an observed GGMRF $\mathbf{x}$ would be difficult. The distribution is highly non-linear in p and dependence on p is through the neighborhood structure of $\mathcal{C}$. A direct maximum likelihood estimate is also difficult because the partition function (the constant of proportionality not shown in (4)) is unknown, though it is also a function of p. The problem can be simplified by noting similarities between the conditional distribution of x_s in Eqn. (5), and the density function for a single zero mean GG random variable d,

$$f_d(d;p) = \frac{Ap}{2\sigma\Gamma(\frac{1}{p})} \exp\left\{-\left[\frac{A|d|}{\sigma}\right]^p\right\} \tag{8}$$

These similarities suggest the following hypothesis:

Hypothesis. Let $d_{s,a}$ be the difference between nearest neighbor pixels s and a of a GGMRF,i.e. $d_{s,a} = x_s - x_a$, where a is restricted to be a neighbor in one direction (e.g. a is always the left nearest neighbor) then (neglecting boundary cases) $d_{s,a}$ is distributed as i.i.d. generalized Gaussian $\forall s \in S$ and has the same shape parameter as the GGMRF.

Observations. Assume that $x_t \in \partial_s$ where ∂_s is the set of the four nearest neighbors of a pixel x_s, i.e. $\partial_s = \{x_a, x_b, x_c, x_d\}$. Replacing $x_s - x_t$ by $d_{s,t}$ in (5), the conditional distribution of a particular site s is

$$\begin{aligned} f_{x_s}(x_s|x_t \in \partial_s) &= K \exp\left\{-\left[\frac{A}{\sigma}\right]^p \sum_{t\in\partial_s} |d_{s,t}|^p\right\} \\ &= K \exp\left\{-\left[\frac{A}{\sigma}\right]^p |d_{s,a}|^p\right\} \exp\left\{-\left[\frac{A}{\sigma}\right]^p |d_{s,b}|^p\right\} \\ &\quad \times \exp\left\{-\left[\frac{A}{\sigma}\right]^p |d_{s,c}|^p\right\} \exp\left\{-\left[\frac{A}{\sigma}\right]^p |d_{s,d}|^p\right\} \end{aligned} \tag{9}$$

where K is the partition function. Comparison of (9) with (8) reveals that $f_{x_s}(x_s|x_t \in \partial_s)$ has the same form as the joint probability distribution function of four i.i.d. generalized Gaussian random variables. Since both (9) and (8) must integrate to 1, we see that $K = \left[\frac{Ap}{2\sigma\Gamma(\frac{1}{p})}\right]^4$. If we compute $f_{x_{s1}}(x_{s1}|x_t \in \partial_{s1})$ and $f_{x_{s2}}(x_{s2}|x_t \in \partial_{s2})$ for two *neighboring* pixels $s1$ and $s2$, one of the difference terms will be common, and thus x_{s1} is clearly not independent of x_{s2}. However, if we restrict ourselves to one directional differences, e.g. $d_{s,a}$, there is no repetition and the $d_{s,a}$ are independent. Thus $\mathbf{d} = [d_{1,a}, d_{2,a}, \cdots, d_{|S|,a}]^T$ is an i.i.d. GG field. We note that Besag argued similarly that the conditional distribution of a Gaussian Markov random field is Gaussian [3]. The hypothesis has also been confirmed experimentally over a wide range of p values by accurately estimating p for GGMRF's through $\mathbf{d}$, using estimation methods designed for GG random variables. □

Given $\mathbf{d}$ we can form a simple estimate for p based on kurtosis. Several authors have used kurtosis to estimate p for GG random data [4] [5], but they have used only an approximate relationship and therefor have produced biased estimates.

It can be shown that the exact relationship between kurtosis and shape parameter p is

$$\beta_2 = \frac{\Gamma(\frac{5}{p})\Gamma(\frac{1}{p})}{\left[\Gamma(\frac{3}{p})\right]^2}. \tag{10}$$

The sample kurtosis can be computed as

$$\hat{\beta}_2 = \frac{|S| \sum_{s\in S} (d_{s,a} - \bar{d})^4}{\left[\sum_{s\in S} (d_{s,a} - \bar{d})^2\right]^2}, \tag{11}$$

where a (as above) is the one directional nearest neighbor of s, $|S|$ is the cardinal number of the set of pixels on the lattice, and $\bar{d} = \frac{1}{|S|}\sum_{s\in S} d_{s,a}$. After an estimate of kurtosis is obtained, $\hat{p}$ is computed by solving equation (10). No closed form solution for p in known, however a table look-up approximate inversion of (10) is easily computed. In practice we have done this efficiently by sampling equation (10) for a number of values of p, and then using cubic spline interpolation between the nearest table entries to approximate the value of p that corresponds to $\hat{p}$.

V. An Iterative Adaptive Algorithm

The shape parameter estimator introduced above requires that the GGMRF, $\mathbf{x}$, be uncorrupted by blur or noise. The algorithm presented in this section enables us to apply these estimation methods even in the case of a corrupted GGMRF and thus to jointly estimate $\mathbf{x}$ and p. The approach is based and an EM-algorithm-like bootstrap procedure which alternates between estimates of p and $\mathbf{x}$ until convergence is achieved. An estimate of the image is first computed as the MAP solution for a fixed starting value of p (usually chosen arbitrarily as $p = 2$, corresponding to a Gauss-Markov field). This image estimate is then used to compute a new estimate of p, which is then applied to the image prior model for the following image estimation step. These two steps are carried out repeatedly until the iteration converges. Thus, we have the following algorithm

Adaptive GGMRF Algorithm

1. Choose an initial shape parameter estimate, $\hat{p}^{(0)}$, and a image estimate, $\hat{\mathbf{x}}^{(0)}$, for the GGMRF prior model.

2. Fix $p = \hat{p}^k$ and compute

$$\hat{\mathbf{x}}^{(k+1)} = \arg\{\min_{\mathbf{x}} \Phi(\hat{\mathbf{x}}^{(k)}, \mathbf{y}; p)\},$$

where $\Phi(\mathbf{x}, \mathbf{y}; p)$ is the energy function of the GGMRF prior model as given in equation (6).

3. Using $\hat{\mathbf{x}}^{(k+1)}$, compute a new estimate $\hat{p}^{(k+1)}$ by forming the nearest neighbor difference image, $\mathbf{d}$ and solving equations (11) and (10).

4. If $\hat{p}^{(k+1)} \approx \hat{p}^{(k)}$, terminate, otherwise increment k and go to step 2.

The optimization called for in step 2 represents a conventional MAP restoration, and can be accomplished in a number of ways. We have used the Metropolis algorithm [6] [2], though other approaches such as the Gibbs Sampler should work as well.

VI. Results

Our experiments have shown that the adaptive algorithm works well in estimating both $\mathbf{x}$ and p when $\mathbf{x}$ is truly a (synthetic) GGMRF. Chen's algorithm [7] was used to generate GGMRF's for a variety of p values. These images were blurred and then restored using the adaptive GGMRF algorithm. Estimates of p were acceptably accurate (e.g. ± 0.2)

The more interesting question however is whether the algorithm performs well with more realistic images. The following experiments were designed to test the performance for more typical images, which are clearly not exactly modeled by a GGMRF.

A. Experiment 1

A 32-by-32 synthetic image, $\mathbf{x}$, with two different geometric shapes was created for this experiment. The synthetic image, shown in Figure 3a, is dominated by two local characteristics, flat regions and edges. This image was convolved with a 5-by-5 pixel uniform blur and Gaussian noise was added to achieve an SNR of $20dB$. Figure 3b shows the resulting observed image, $\mathbf{y}$. The Frobenius norm of the error image, $\mathbf{x} - \mathbf{y}$, was 309. Figure 3c shows the restoration result, $\hat{\mathbf{x}}_{GGMRF}$, obtained using the adaptive GGMRF method. The algorithm converged after 10 iterations and the final value for p was 1.0. The value of p estimated from Figure 3a directly using the kurtosis method outlined above was 0.7. The Frobenius norm of the error image, $\mathbf{x} - \hat{\mathbf{x}}_{GGMRF}$, was 114. For comparison purposes, Figure 3d shows the result, $\hat{\mathbf{x}}_{GMRF}$, obtained by using the GMRF prior model in the MAP restoration; the Frobenius norm of the error image, $\mathbf{x} - \hat{\mathbf{x}}_{GMRF}$, was 303. In addition to the lower error result obtained by the adaptive GGMRF method, the visual quality is also better than that of the GMRF MAP method, which is clearly over-smoothed, as shown in Figure 3d. The error images from both methods are shown in Figure 3e and 3f.

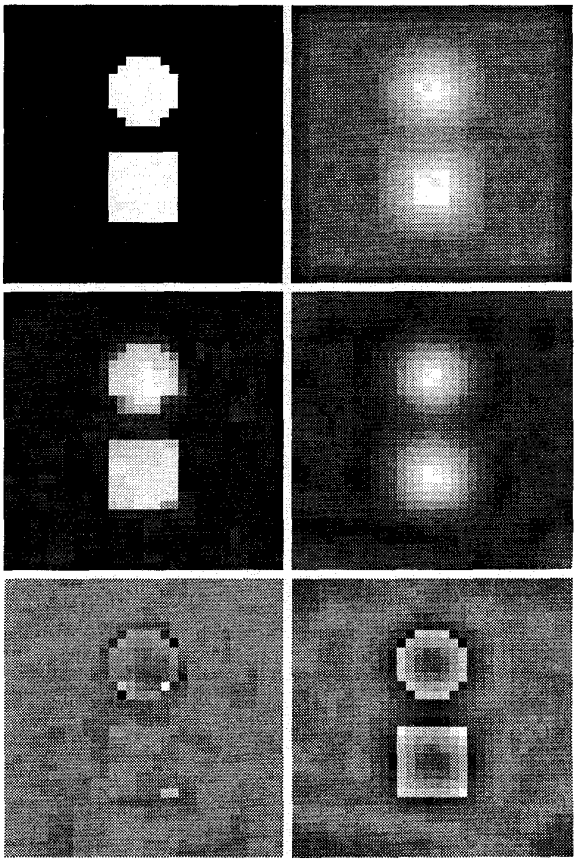

Fig. 3. Images for Experiment 1 of the fully-adaptive GGMRF method. a) [top l.] Original 32-by-32 synthetic image, b) [top r.] Blurred and noisy observation, c) [mid l.] Restoration result of (b) by the fully-adaptive GGMRF method, d) [mid r.] Restoration result of (b) by the GMRF MAP estimation, e) [bot. l.] Error image from (c), f) [bot. r.] Error image from (d). Prominent edges in the error image mean serious errors resulted form the restoration process.

B. Experiment 2

A 128-by-128 image, shown in Figure 4a, was blurred by a 25-by-1 PSF = [1 1 $\cdots$ 1], representing a 1D horizontal motion blur. White Gaussian noise at an SNR of $30dB$ was added to the blurred image to produce figure 4b. The shape parameter estimated directly from the uncorrupted Figure 4a was 0.8.

The adaptive algorithm was applied to Figure 4b to jointly estimate $\mathbf{x}$ and p. Ten iterations were used

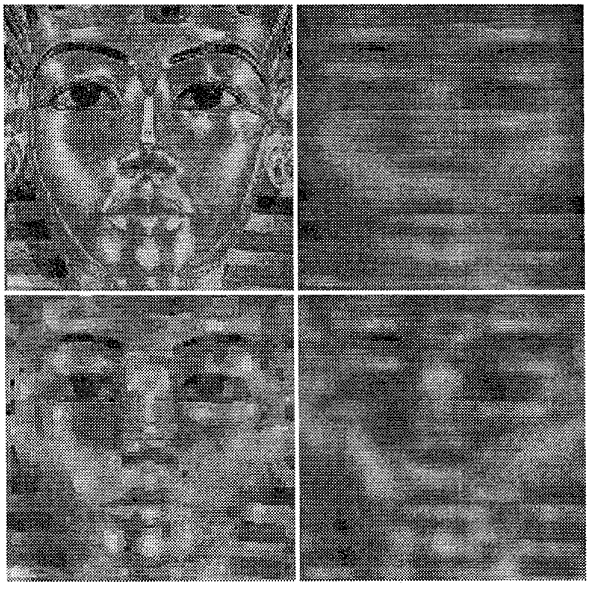

Fig. 4. Images from the experiment of the adaptive algorithm. a) [top l.] Original 128-by-128 image, b) [top r.] Blurred noisy image, c) [bot. l.] Adaptive GGMRF MAP restoration of a, d) [bot. r.] GMRF model restoration of a.

to generate Figure 4c. The starting value of p was 2; the ending estimate of p was 0.61. This is not exactly 0.8, however, the shape of the distribution at $p = 0.61$ is a good approximation for the shape at $p = 0.8$. For comparison, a MAP restoration using a GMRF ($p = 2$) was computed and shown in Figure 4d. The Frobenius norm of the corresponding error image (Figure 4a minus 4d) is twice that of the error in Figure 4c. Though some of the blurring due to horizontal motion is reduced in Figure 4d, the result still looks over-smoothed. On the other hand, Figure 4c is both visually and numerically better because the edges were preserved as a result of adapting the model to the image.

C. Experiment 3

For the sake of completeness, we would like to carry out an experiment on images which can be modeled by the GGMRF prior with $p > 2.0$. However, our search for such real-life images was unsuccessful. The diverse patterns in images which are of interest to humans almost always have edges which are best modeled by a shape parameter value lower than 2. Consequently, we resort to synthetic images generated as a GGMRF for this experiment.

The synthetic GGMRF image shown in Figure 5a was generated using Chen's algorithm [7] with $p = 3.0$. This image was convolved with a 3 by 3 pixel uniform blur and Gaussian noise was added to produce an SNR of $20dB$. Figure 5b shows the observed image, which has a corresponding error image with a Frobenius norm of 57. Figure 5c shows the result obtained by the fully-adaptive GGMRF method. The Frobenius norm of the error image from this result is 51. Figure 5d shows the result obtained by a GMRF MAP estimation for comparison. The Frobenius norm of the error image from this result is 60, which offers no improvement over that of the observed image.

These experiments demonstrate the effectiveness of the adaptive GGMRF algorithm and suggest it will be particularly useful when there is not enough known about the true image to choose a good prior model.

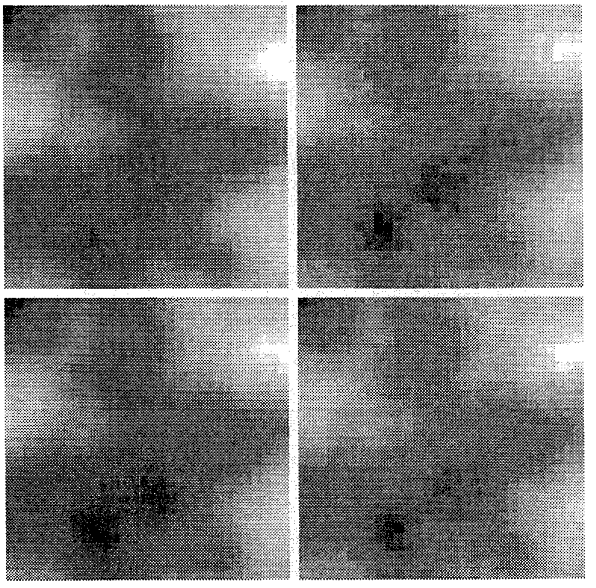

Fig. 5. Images for the Experiment 3 of the adaptive algorithm. a) [top l.] Original 128-by-128 image, b) [top r.] Blurred noisy image, c) [bot. l.] Fully-adaptive restoration of a, d) [bot. r.] GMRF model restoration of a.

References

[1] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation", *IEEE Transactions on Image Processing*, vol. 2, no. 3, pp. 296–310, July 1993.

[2] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 721–741, Nov. 1984.

[3] J. Besag, "Towards Bayesian image analysis", *Journal of Applied Statistics*, vol. 16, no. 3, pp. 395–407, 1989.

[4] A. H. Money, J. F. Affleck-Graves, M. L. Hart, and G. D. I. Barr, "The linear regression model: l_p-norm estimation and the choice of p", *Communications in Statistics - Simulation and Computation*, vol. 11, no. 1, pp. 89–109, 1982.

[5] V. A. Sposito and M. L. Hand, "On the efficiency of using the sample kurtosis in selecting optimal l_p estimators", *Communications in Statistics - Simulation and Computation*, vol. 12, no. 3, pp. 265–272, 1983.

[6] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computer machines", *Journal of Chemical Physics*, vol. 21, pp. 1087–1092, 1953.

[7] C. C. Chen, *Markov random field models in image analysis*, PhD thesis, Michigan State University, East Lansing, 1988.