Prior Models and the Richardson-Lucy Restoration Method

R. Molina, J. Mateos and J. Abad

Dept. of Computer Science, Faculty of Science, University of Granada, 18071 Granada, Spain

Abstract. Following the Bayesian paradigm for image restoration we show how smoothness constraints can be incorporated into the R-L method. We also examine different noise models and study their approximation by Gaussian models that are robust to detector errors.

1. Introduction

Image restoration refers to the problem of recovering an image, $\mathbf{f}$, from its blurred and noisy observation, $\mathbf{g}$, for the purpose of improving its quality or obtaining some type of information that is not readily available from the degraded image. Blur is frequently due to the relative motion between the subject and the camera, atmospheric turbulence, out of focus lenses, and/or the image sensor. Film grain, electronic noise, and quantization are the major sources of noise in a digital image.

In this paper we describe how the Bayesian approach to image restoration can be used to incorporate prior information, in the form of smoothness constraints, to the R-L restoration method (Lucy, 1974). We also examine the noise models and study how to remove detector errors.

The work is divided as follows. In section 2 we describe briefly the Bayesian paradigm. In section 3 we study the observation process and how to remove detector errors. Section 4 is devoted to the study of prior models that take into account smoothness constraints and also to the study of the prior model used in the R-L restoration method. The next step is to calculate the estimate of the real underlying image, $\mathbf{f}$; the algorithm implementing this task and its relationship to the R-L restoration method are described in section 5. Finally, a test example is shown in section 6.

2. Bayesian Paradigm

The philosophy within statistics known as Bayesian inference has a very long history. It is distinguished from the perhaps more familiar classical statistical ideas by using prior information about the images being studied.

Bayesian methods start with a prior distribution, a probability distribution over images $\mathbf{f}$, $p(\mathbf{f})$, (it is here that we incorporate information on the expected structure within an image), it is also necessary to specify the probability distribution $p(\mathbf{g}|\mathbf{f})$, of observed images $\mathbf{g}$ if $\mathbf{f}$ were the true image. The Bayesian paradigm dictates that inference about the true $\mathbf{f}$ should be based on $p(\mathbf{f}|\mathbf{g})$ given by

$$p(\mathbf{f}|\mathbf{g}) \propto p(\mathbf{f})p(\mathbf{g}|\mathbf{f})$$

To show just one restoration, it is common to choose the mode of this posterior distribution, that is to display the image $\hat{\mathbf{f}}$ which satisfies

$$\hat{\mathbf{f}} \text{ maximizes } p(\mathbf{f})p(\mathbf{g}|\mathbf{f})$$

This is known as the maximum a posteriori (MAP) estimate of $\mathbf{f}$.

Equivalently, we can choose $\hat{\mathbf{f}}$ to minimize

$$-\log p(\mathbf{g}|\mathbf{f}) - \log p(\mathbf{f}) \tag{1}$$

the first term in (1) is the familiar log likelihood of $\mathbf{f}$. The second term can be thought of as a roughness penalty, as images $\mathbf{f}$ which do not correspond to our prior conceptions will be assigned a small $p(\mathbf{f})$ and hence a large penalty.

In statistical physics it is common to define probabilities by the energy U of a system, so that

$$p(\mathbf{f}) \propto \exp[-\beta U(\mathbf{f})] \tag{2}$$

where β is $1/kT$, T being the temperature and k Boltzmann's constant. If we adopt this notation, $\hat{\mathbf{f}}$ minimizes

$$-\log \text{likelihood} + \beta U(\mathbf{f}) \tag{3}$$

We can recognize this as a Lagrangian form, so its solution is equivalent to solving

$$\max \ \text{likelihood subject to energy} \leq \text{constraint}$$

and to

$$\min \ \text{energy subject to likelihood} \geq \text{constraint}$$

which correspond to the regularization approach to image restoration.

Many other deconvolution principles fit into one of these forms, in particular— as we will see later— R-L restoration method. Maximum entropy methods also fit into this framework (Molina et al. 1992a).

Having described the Bayesian paradigm let us move on to examine the two ingredients of this paradigm, the observation process, $p(\mathbf{g}|\mathbf{f})$, and the prior model or image model, $p(\mathbf{f})$.

3. Observation Process

The observed image $\mathbf{g}$ differs from the true brightness distribution $\mathbf{f}$ in having been blurred and encountering statistical noise in the recording process.

If $p = m \times n$ is the size of the image, the blurring process is described by $\mathbf{Df}$ where $\mathbf{f}$ is a $p \times 1$ vector and $\mathbf{D}$ is the $p \times p$ point spread matrix defining the systematic blur and assumed to be known.

Let us describe the noise models. For each component i, $i = 1, 2, \ldots, m \times n$ of the observed vector $\mathbf{g}$ we could use a Poissonian model, thus obtaining $\mathbf{g}_i \sim \mathcal{P}((\mathbf{Df})_i)$. This model can be approximated, at least for high brightness values, by the Gaussian distribution $\mathcal{N}((\mathbf{Df})_i, (\mathbf{Df})_i)$.

An alternative model would be to assume $\mathbf{g}_i = \lambda \mathbf{z}_i$ where $\mathbf{z}_i \sim \mathcal{P}(\mathbf{Df})_i$. For this model we could use the following Gaussian approximation

$$\mathbf{g}_i \sim \mathcal{N}((\mathbf{D}(\lambda\mathbf{f}))_i, \lambda(\mathbf{D}(\lambda\mathbf{f}))_i)$$

substituting $\lambda\mathbf{f}$ by $\mathbf{f}$, we would have

$$\mathbf{g}_i \sim \mathcal{N}((\mathbf{Df})_i, \lambda(\mathbf{Df})_i)$$

Finally, a model like

$$\mathbf{g}_i = \mathcal{P}((\mathbf{Df})_i) + \mathcal{N}(0, \sigma^2)$$

and the corresponding $\lambda\mathbf{z}_i + \mathcal{N}(0, \sigma^2)$ can be approximated by normal distributions having the form $\mathcal{N}((\mathbf{Df})_i, a + b(\mathbf{Df})_i)$ for appropriate constants a and b.

If we use the Gaussian approximation we have

$$-2\,\log p(\mathbf{g}|\mathbf{f}) = \text{const} + \sum_i D_i^2 \tag{4}$$

with $D_i = [(\mathbf{g}_i - (\mathbf{Df}_i))/\sigma(\mathbf{Df}_i)]$, being $\sigma((\mathbf{Df})_i)$ the standard deviation, $\sqrt{a + b(\mathbf{Df})_i}$ in the noise model just described.

This Gaussian approximation allows the easy incorporation of robust statistics concepts to deal with detector errors (Molina and Ripley, 1989). The idea is to downweigh observations which are far away from their means. Such values are given too much weight in (4). The squared term in D_i represents the number of standard deviations that $\mathbf{g}_i$ is away from its mean. In robust statistics $\sum_i D_i^2$ is replaced by $\rho(D_i)$ for a function ρ which penalizes extreme values less severely. A typical function ρ is the Huber's 'proposal 2' function defined by

$$\rho(x) = \begin{cases} x^2 & \text{for} \quad |x| \leq c \\ 2c|x| - c^2 & \text{for} \quad |x| > c \end{cases}$$

This is quadratic in the centre, but penalizes large deviations linearly rather than quadratically. Equivalently, observations D_i are downweighted if $|D_i|$ exceeds 2. In practice c is chosen at about 2, which downweighs only those observations more than two standard deviations away from their means.

The noise model we will use in this paper is Poissonian, see Molina and Ripley (1989) for the use of robust Gaussian noise models.

4. Prior Models

Consider an image with no stars but regions of smoothly varying luminosity. We then expect $\mathbf{f}_i \geq 0$ and $\mathbf{f}$ to be spatially smooth. Probably the simplest probability models that can be used to model smoothness are spatial autoregressions (Ripley, 1981).

The conditional autoregression (CAR) model is defined by

$$p(\mathbf{f}) \propto \exp\{-\frac{1}{2}\alpha\,\mathbf{f}^t(I - \phi N)\mathbf{f}\}$$

where α is the unknown hyperparameter, matrix N is such that $N_{ij} = 1$ if cells i and j are spatial neighbours, (pixels at distance one), zero otherwise and scalar ϕ is just less than 0.25. The term $\mathbf{f}^t(I - \phi N)\mathbf{f}$ represents in matrix notation the sum of squares of the values $\mathbf{f}_i$ minus ϕ times the sum of $\mathbf{f}_i\mathbf{f}_j$ for neighbouring pixels i and j.

The parameters can be interpreted by the following expressions describing the conditional distribution

$$\begin{aligned} E(\mathbf{f}_i \mid \mathbf{f}_j,\ j \neq i) &= \phi \sum_{j \text{ nhbr } i} \mathbf{f}_j \\ \text{var}(\mathbf{f}_i \mid \mathbf{f}_j,\ j \neq i) &= \alpha^{-1} \end{aligned}$$

where the suffix 'j nhbr i' denotes the four neighbour pixels at distance one from pixel i. The parameter α measures the smoothness of the 'true' image.

Assuming a toroidal edge correction, the eigenvalues of the matrix $I - \phi N$ are $\lambda_{ij} = 1 - 2\phi(\cos(2\pi i/m) + \cos(2\pi j/n)), i = 0, 1, 2, \ldots, m-1, j = 0, 1, 2, \ldots, n-1$. So, the density of $\mathbf{f}$ has the form

$$p(\mathbf{f}) = \frac{1}{Z_{prior}(\alpha)} \exp\{-\frac{1}{2}\alpha\,\mathbf{f}^t(I - C)\mathbf{f}\}$$

where $Z_{prior}(\alpha) = (\prod_{i,j} \lambda_{ij})^{-1/2}(2\pi/\alpha)^{p/2}$, $C = \phi N$.

This prior model can be easily modified to work at log scale, which is the right scale for the deconvolution of galaxies (Molina et al. 1992a). Furthermore, this model can also take into account the existence of different objects in the image (Molina et al. 1992b).

Let us now examine the prior model used in the R-L restoration method. This method aims at maximizing $p(\mathbf{g}|\mathbf{f})$ when this conditional distribution is Poissonian. Under the Bayesian framework this is the same as maximizing the posterior distribution for the prior model $p(\mathbf{f}) = \text{const}$, together with the Poissonian noise model. The meaning of this prior is simple; all possible restorations have the same probability.

5. Algorithms

Having defined the prior and observational models, let us move on to estimate the MAP.

Following the R-L method, which, as we have said, corresponds to maximum a posterior estimation with a uniform image prior, we seek to find

$$\mathbf{f}_{RL} = arg \max_{\mathbf{f}} \{ \prod_{i=1}^{p} \exp\left[-(\mathbf{Df})_i\right][(\mathbf{Df})_i]^{\mathbf{g}_i}/\mathbf{g}_i! \} \tag{5}$$

Applying logarithms and differentiating with respect to $\mathbf{f}$ we obtain the following equation

$$\mathbf{D}^t(\mathbf{g}/\mathbf{Df}_{RL}) = \mathbf{1} \tag{6}$$

where we have assumed that $\mathbf{D}^t\mathbf{1} = \mathbf{1}$. To solve (6) Lucy uses the iterative scheme

$$\mathbf{f}_j^i = \mathbf{f}_j^i[\mathbf{D}^t(\mathbf{g}/\mathbf{Df}^i)]_j$$

where i denotes iteration and j component of the vector. This iterative scheme is justified as an iterative scheme derived from EM principles. However, it can also be obtained by multiplying both sides of (6) by $\mathbf{f}$.

Let us now assume that we want to impose smoothness constraints on the solution by using a CAR prior model, we have

$$p(\mathbf{f}|\mathbf{g}) \propto \exp[-\frac{\alpha}{2}\mathbf{f}^t(I-C)\mathbf{f}] \prod_{i=1}^{p} \exp\left[-(H\mathbf{f})_i\right][(H\mathbf{f})_i]^{g_i}/g_i! \tag{7}$$

Differentiating $-\log p(\mathbf{f}|\mathbf{g})$ with respect to $\mathbf{f}$ we obtain

$$\alpha(I-C)\mathbf{f} + \mathbf{1} - H^t(\mathbf{g}/H\mathbf{f}) = 0$$

or

$$\mathbf{f} + \alpha^{-1}\mathbf{1} = C\mathbf{f} + \alpha H^t(\mathbf{g}/H\mathbf{f}) \tag{8}$$

multiplying both sides of (8) by $\mathbf{f}$ we obtain the following iterative scheme

$$\mathbf{f}_j^{i+1} = (\mathbf{f}_j^i[C\mathbf{f}^i]_j + \alpha^{-1}\mathbf{f}_j^i H^t(\mathbf{g}/H\mathbf{f}^i)]_j)/(\mathbf{f}_j^i + \alpha^{-1}) \tag{9}$$

where i denotes iteration and j component of the vector. This equation can be rewritten as

$$\mathbf{f}_j^{i+1} = \mu_j^i[C\mathbf{f}^i]_j + (1-\mu_j^i)\mathbf{f}_j^i H^t(\mathbf{g}/H\mathbf{f}^i)]_j$$

where $\mu_j^i = \mathbf{f}_j^i/(\mathbf{f}_j^i + \alpha^{-1})$, $\mu_j^i = 0$ corresponds to the classical R-L restoration method.

Before examining an example, let us briefly comment on the problem of estimating α. Although it is possible to estimate α on a trial and error basis we are currently working on the use of the hierarchical Bayesian approach to image restoration (Molina, 1993). The idea is to use the joint distribution defined as

$$p(\alpha, \mathbf{f}, \mathbf{g}) = p(\alpha)p(\mathbf{f}|\alpha)p(\mathbf{g}|\mathbf{f}) \tag{10}$$

integrate (10) on α to obtain $p(\mathbf{f}, \mathbf{g})$ and then find $\underline{\mathbf{f}}$ which satisfies

$$\underline{\mathbf{f}} \text{ maximizes } p(\mathbf{f}, \mathbf{g})$$

For the problem we have at hand we can use improper noninformative priors, $p(\alpha) \propto const$, and also gamma distributions. The relationship between this hierarchical model and the method developed by Katsaggelos (1993) are under investigation.

6. Examples

We shall apply the algorithm just developed to the restoration of a WF/PC image of Saturn. The distorted and restoration images obtained by our method after 50 iterations, using (9), with $\alpha^{-1} = 106,000$ are shown in figure 1.

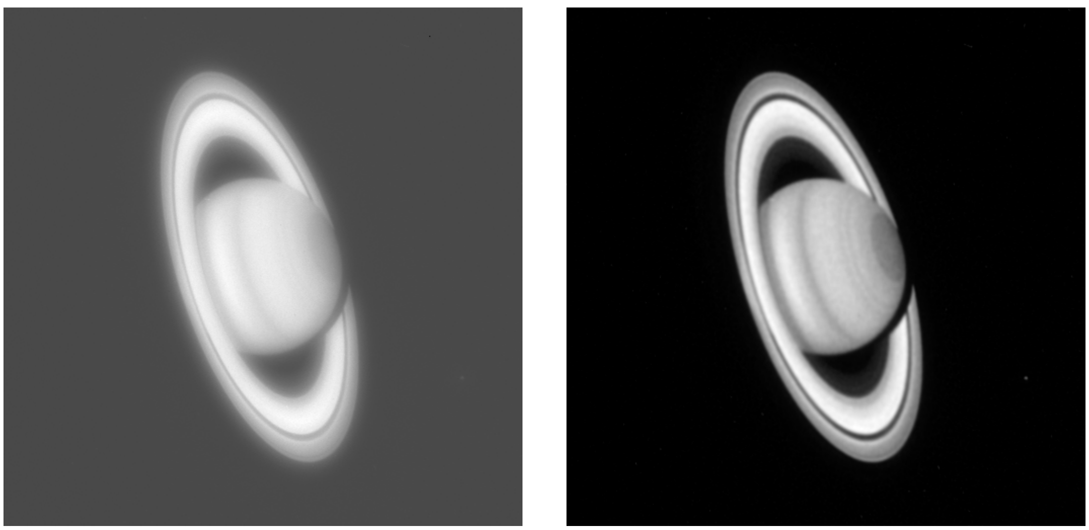

Figure 1. (a) Observed Saturn image. (b) Restoration with $\alpha^{-1} = 106,000$ after 50 iterations.

Acknowledgments. This work has been supported by the "Comisión Nacional de Ciencia y Tecnología" under contract TIC91-0660.

References

Lucy, L. 1974, AJ, 79, 745

Katsaggelos, A. 1993, in this volume

Molina, R. and Ripley, B.D. 1989, J. of Appl. Statist., 16, 193

Molina, R., Olmo, A., Perea, J and Ripley B.D. 1992a, AJ, 103, 666

Molina, R., Ripley B.D., Molina, A., Moreno, F. and Ortiz, J.L. 1992b, AJ, 104, 1662

Molina, R. 1993, Submitted to PAMI

Ripley, B.D. 1981, Spatial Statistics, (Wiley, New York)