ΣΤΟΧΑΣΤΙΚΑ ΜΟΝΤΕΛΑ ΚΑΙ ΜΕΘΟΔΟΙ ΓΙΑ ΑΠΟΣΥΝΕΛΙΞΗ ΕΙΚΟΝΩΝ

Χάντας Ιωάννης

Ιωάννινα, Σεπτέμβριος 2004

Περίληψη

Η αποσυνέλιξη εικόνων είναι ένα δύσκολο αλλά και βασικό πρόβλημα που δεκαετίες έχει απασχολήσει τους ερευνητές επεξεργασίας σήματος. Το πρόβλημα ορίζεται ως εξής: η υποβαθμισμένη εικόνα έχει συνελιχθεί με ένα γραμμικό και χρονικά αμετάβλητο φίλτρο, ενώ μετέπειτα θόρυβος έχει προστεθεί. Το πρόβλημα είναι η αποκατάσταση (restoration) της αρχικής εικόνας, δεδομένης της υποβαθμισμένης, μέσω αποσυνέλιξης. Πρόκειται για ένα «κακώς ορισμένο» (ill-posed) πρόβλημα, οπότε η εισαγωγή εκ των προτέρων γνώσης για την εικόνας μας είναι αναγκαία.

Το στοχαστικό μοντέλο της διαδικασίας της υποβάθμισης, η κακώς ορισμένη φύση της συνέλιξης που παράγει τις παρατηρήσεις καθιστά επιτακτική την ανάγκη για εισαγωγή εκ των προτέρων γνώσης της εικόνας στο πρόβλημα της αποσυνέλιξης. Αυτό μας οδηγεί αυτόματα στη χρήση της Μπεϋζιανής μεθοδολογίας (Bayesian methodology). Η Μπεϋζιανή μεθοδολογία είναι μια πολύ ευέλικτη και εύχρηστη μεθοδολογία η οποία έχει χρησιμοποιηθεί κατά κόρο σε προβλήματα αυτής της φύσεως.

Με στόχο την αναπαράσταση χαρακτηριστικών μιας εικόνας αυτή μοντελοποιείται από μια στοχαστική διαδικασία, που πολλές φορές θεωρείται στατική (stationary). Όμως οι στατικές ιδιότητες που υποθέτονται πολύ συχνά για λόγους υπολογιστικής ευκολίας, είναι μη ρεαλιστικές μιας και δεν μπορούν να αναπαραστήσουν τις χωρικά μεταβαλλόμενες ιδιότητες των εικόνων. Ως εκ τούτου δεν δίνουν πολύ καλά αποτελέσματα όταν χρησιμοποιούνται σαν μοντέλα στην αποκατάσταση της εικόνας. Για αυτό το λόγο μη-στατικά (non-stationary) μοντέλα και μέθοδοι έχουν προταθεί.

Σε αυτήν την εργασία ένα νέο στοχαστικό μη-στατικό μοντέλο για εικόνες μελετήθηκε, και με βάση αυτό μέθοδοι για την αποκατάσταση εικόνων αναπτύχθηκαν και υλοποιήθηκαν. Στο πρώτο κεφάλαιο γίνεται η εισαγωγή στο πρόβλημα της αποσυνέλιξης εικόνων, καθώς και κάποιες κλασικές μέθοδοι στατικής αποκατάστασης. Στο δεύτερο παρουσιάζονται προηγούμενες εργασίες και μοντέλα για μη-στατικές μεθόδους. Στο τρίτο παρατίθεται γενική περιγραφή της Μπεϋζιανής μεθοδολογίας καθώς και οι αρχές της. Στο τέταρτο κεφάλαιο, παρουσιάζεται το μη-στατικό μοντέλο και οι τρόποι επίλυσης του. Στο τέλος παρουσιάζονται τα αριθμητικά πειράματα και τα συμπεράσματα που προέκυψαν από την εκπόνηση αυτής της διατριβής.

University of Ioannina
Deparment of Computer Science

STOCHASTIC MODELS AND METHODS FOR IMAGE DECONVOLUTION

Chantas Ioannis

Ioannina, September 2004

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION TO DIGITAL IMAGE RESTORATION

1.1 Imaging Models in Restoration

Image restoration is the scientific field that studies methods for recovery of the original image from observed degraded data. This field emerged as imaging systems appeared. The imperfections of the imaging systems and the artifacts that they yield when they produce images motivated the study of methods, mainly in the field of signal processing, to restore degraded images. The development also of the digital computer along with the invention of digital imaging systems has contributed even more to the development of this field.

Applications of image restoration are numerous. For example, astronomical images from spacecrafts and telescopes are usually degraded due to atmospheric turbulence and imperfect optics. Another important field of applications is medical imaging where the degradations hinder the patient's treatment. X-rays, Magnetic Resonance Imaging (MRI), f-MRI, Positron Emission Tomography (PET) are examples of imaging modalities where restoration is often necessary. Photography is another field where image restoration can be useful when the focus is not good or there is motion.

Images are usually degraded by two processes. In the first process a spatial degradation is caused from various reasons, i.e. out of focus camera, atmospheric turbulence and motion. This results in blurring the original image. The second process is a point degradation according to which noise affects the individual pixel gray level.

This is caused by various types of noise (shot, thermal) in the detector systems and errors in the recording process due to quantization. In this work we assume, as in many other works, the space-invariant blurring and the white noise. Formally, the imaging model is linear and is given by:

$$g(i,j)=\sum_{k=1}^{N}\sum_{l=1}^{N}h(i-k,j-l)f(k,l)+n(i,j),$$

which, can also be written as:

$$\mathbf{g}=\mathbf{h}**\mathbf{f}+\mathbf{n}. \qquad (1.1)$$

The observed image $\mathbf{g}$, is an $N\times N$ matrix, and is produced first by a convolution of the original image $\mathbf{f}$, also an $N\times N$ matrix, with $\mathbf{h}$ a linear shift invariant (LSI) low-pass filter. This is also called the point spread function (PSF), since it spreads an impulse to many pixels, and models the spatial degradation mechanism. Then, noise is added, where and are the zero and the identity matrices respectively. The operation ** is the two dimensional convolution. Noise variance is often assumed unknown, in contrast with the PSF, which will be assumed known in this thesis.

To use a more convenient notation we express the convolution by a matrix-vector multiplication by an $N^2\times N^2$ matrix $\mathbf{H}$ (representing linear convolution operator) by the vector $\mathbf{h}$. The equivalent equation is:

$$\mathbf{g}=\mathbf{H}\mathbf{f}+\mathbf{n}, \qquad (1.2)$$

where $\mathbf{g},\mathbf{f}$ and $\mathbf{n}$ are $N^2\times 1$ vectors ordered lexicographically by the previous matrix formation.

Due to the LSI nature of the PSF the matrix $\mathbf{H}$ is block-Toeplitz, but can be efficiently approximated by a block circulant matrix [3]. The larger the dimension N

is, the better the approximation becomes [4]. Using this approximation is the same as assuming that the convolution in Eq.(1.1) is circulant. One can pad the vectors in the convolution with zeros, and convert any linear convolution to a circulant one. The matrix **H** is formed as:

$$\mathbf{H}=\begin{bmatrix} H(0) & H(N-1) & \cdots & H(1) \\ H(1) & H(0) & \cdots & H(N-1) \\ \vdots & \vdots & \ddots & \vdots \\ H(N-1) & H(N-2) & \cdots & H(0) \end{bmatrix},$$

where each sub-matrix is a circulant matrix, formed by **h**, the LSI filter:

$$H(i)=\begin{bmatrix} h(i,0) & h(i,N-1) & \cdots & h(i,1) \\ h(i,1) & h(i,0) & \cdots & h(i,N-1) \\ \vdots & \vdots & \ddots & \vdots \\ h(i,N-1) & h(i,N-2) & \cdots & h(i,0) \end{bmatrix}.$$

The circulant form leads to more tractable equations and an easy to handle model in the discrete Fourier transform (DFT) as will be explained in more detail latter on. This is because the eigenvalues of all circulant matrices are obtained by DFT of the generating filter **h** [3]. In other words the eigenvectors of circulant matrices are the complex exponentials basis functions of the DFT. Formally, a circulant $N \times N$ matrix **A** is diagonalized as follows:

$$\mathbf{A} = N^{-1}\mathbf{W}\mathbf{\Lambda}\mathbf{W}^{-1}$$

where $\mathbf{\Lambda}$ is diagonal with eigenvalues (the DFT coefficients), $N^{-1/2}\mathbf{W}$ is the DFT transform operator matrix and $N^{-1/2}\mathbf{W}^{-1}$ the inverse [3]. In the case of a block circulant matrix the same equations hold. The only difference is that instead of the one dimensional DFT transform, two dimensional DFT is used. In this thesis the notation that will be used for simplicity is one dimensional.

1.2 Restoration Algorithms

Now that the imaging model has been defined, the task that remains is to find a method to obtain the original image $\mathbf{f}$ (or an estimate $\hat{\mathbf{f}}$) from the degraded observations. The most and obvious and naïve way is to take the inverse of the convolution matrix and apply it to the observations. This results in the estimate:

$$\hat{\mathbf{f}} = \mathbf{H}^{-1}\mathbf{g}.$$

In the DFT domain, using the previous diagonalization we get:

$$\hat{\mathbf{f}} = \mathbf{H}^{-1}\mathbf{g} \Rightarrow \hat{\mathbf{f}} = \frac{1}{N}\mathbf{W}\Lambda_h^{-1}\mathbf{W}^{-1}\mathbf{g} \Rightarrow \mathbf{W}^{-1}\hat{\mathbf{f}} = \Lambda_h^{-1}\mathbf{W}^{-1}\mathbf{g} \Rightarrow \hat{\mathbf{F}} = \Lambda_h^{-1}\mathbf{G},$$

where, Λ_h is a diagonal matrix with the eigenvalues of $\mathbf{H}$ (the DFT coefficients), and $\hat{\mathbf{F}}, \mathbf{G}$ are the images in the DFT domain. Unfortunately, noise has been added to the observed image, so by Eq.(1.2) the estimated image will be:

$$\hat{\mathbf{f}} = \mathbf{f} + \mathbf{H}^{-1}\mathbf{n}. \qquad (1.3)$$

Since matrix $\mathbf{H}$ has certain small eigenvalues, or even worse equal to zero, the noise will be greatly amplified. Furthermore, the zero eigenvalues make the invertion impossible. To make this more clear Eq.(1.3) is written in the frequency domain, this gives:

$$\hat{\mathbf{F}}(i) = \frac{\mathbf{G}(i)}{\Lambda_h(i,i)} = \mathbf{F}(i) + \frac{\mathbf{N}(i)}{\Lambda_h(i,i)}, \quad i = 1...N, \qquad (1.4)$$

where $\mathbf{F}$ and $\mathbf{N}$ are the DFT of the image and noise vectors respectively. Notice that the eigenvalues $\Lambda_h(i,i)$ must not be zero for the inversion to be possible. However, a very disastrous and usual situation is when large levels of noise correspond to high frequencies where the eigenvalues $\Lambda_h(i,i)$ are close to zero. This results in restored images that contain noise which is greatly amplified.

In the scientific literature such problems are called "ill-posed". Loosely speaking this means that small changes in the observed data can cause very large changes to the restored image, or at the worst case the original data cannot be recovered (restored in our case), even in the absence of noise (**H** is not invertible) [1,2]. In order to avoid this difficulties of solving an ill-posed problem the theory of regularization has been introduced which converts them to well-posed [1, 2]. This is achieved by constraining the set of admissible solutions using a priori knowledge about the image. Since regularization changes the problem to a well-posed one the solution also changes. However, if the a priori knowledge used is accurate and realistic, the results can be quite satisfactory.

A very commonly used estimate in image restoration is the linear minimum mean square error (LMMSE) estimate [3]. This estimate is based on second order statistics of the image and the noise. More specifically $\mathbf{R}_{\mathbf{ff}} = E\left[\mathbf{f}\mathbf{f}^{\mathrm{T}}\right]$ and $\mathbf{R}_{\mathbf{nn}} = E\left[\mathbf{n}\mathbf{n}^{\mathrm{T}}\right]$, where $E[\cdot]$ is the expectation operator, are the image and noise variances, respectively. If zero mean is assumed, the covariance is equal to the correlation matrix. The covariances can be computed from the degraded image for $\mathbf{R}_{\mathbf{ff}}$ and from a flat region of the image for $\mathbf{R}_{\mathbf{nn}}$. The LMMSE estimate which is also called Wiener filter [3] minimizes the expectation:

$$\min_{\hat{\mathbf{f}}}\left\{E\left[\left\|\mathbf{f}-\hat{\mathbf{f}}\right\|^2\right]\right\}.$$

Then, the estimated image is then given by:

$$\hat{\mathbf{f}} = \mathbf{R}_{\mathbf{ff}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}^{\mathrm{T}}\mathbf{R}_{\mathbf{ff}}\mathbf{H}+\mathbf{R}_{\mathbf{nn}}\right)^{-1}\mathbf{g}. \qquad (1.5)$$

In this case the inversion is better conditioned, and the ill-posed problem has been avoided. This can be seen in our case where the noise has simple covariance and it is circulant:

$$\mathbf{R}_{nn} = \sigma^2 \mathbf{I} = \mathbf{W}^{-1} \sigma^2 \mathbf{I} \mathbf{W}.$$

Thus the power spectrum $\mathbf{S}_{nn}$ is easily computed:

$$\mathbf{S}_{nn}(i) = \sigma^2, \quad i = 1,...,N.$$

If the process is assumed zero mean the covariance euals to correlation matrix. Also if the covariance matrix is assumed circulant, the equivalent equation of Eq.(1.5) to the DFT domain, following the same notation as in Eq.(1.4), is:

$$\hat{\mathbf{F}}(i) = \frac{\mathbf{S}_{ff}(i) \Lambda_h^*(i,i)}{\|\Lambda_h(i,i)\|^2 \mathbf{S}_{ff}(i) + \sigma^2} \mathbf{G}(i), \quad i = 1,...,N, \quad (1.6)$$

where $S_{ff}(i)$ is the power spectrum of the image [3]. The quantity that multiplies the observed image to obtain the restored is the inverse filter, since the multiplication is a convolution to the spatial domain. This formulation was possible because the image and the noise covariances were assumed circulant.

A stochastic process with statistical characteristics that do not change with time (or space) is called *stationary*. A random process $x(t)$ is stationary in a wide sense (wide sense stationary, WSS) if the mean is constant and independent of time, and the covariance (or correlation for zero mean) between $x(t)$ and $x(s)$ where t and s are two points in time of space depends only on the difference t-s:

$$E[x(t)x(s)] = r(t-s).$$

This for a discrete process yields as covariance a matrix which is Toeplitz. Circulant matrices are a special case of Toeplitz matrices and as in the case of the operator **H**

the Toeplitz covariances can be approximated by a circulant one. The main benefit of a stationary model is the ease of solution in the DFT domain. Unfortunately, these models usually cannot describe efficiently the real world phenomena.

At this point we will present how deterministic regularization is used to solve ill-posed problems. In deterministic approaches the criterion used to find the restored image is the minimization of the Euclidian norm:

$$\|\mathbf{g}-\mathbf{H}\mathbf{f}\|^2 .$$

However, this alone yields the pseudo-inverse solution

$$\hat{\mathbf{f}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{g},$$

which, is ill-posed [2]. To ameliorate this situation Tikhonov regularization can be used with an addition of a penalty (regularization) term [2] according to:

$$\min_{\mathbf{f}}\left\{\|\mathbf{g}-\mathbf{H}\mathbf{f}\|^2 + a\|\mathbf{Q}\mathbf{f}\|^2\right\}. \qquad (1.7)$$

The parameter a is a scalar and $\mathbf{Q}$ is a $N\times N$ matrix. The penalty term tries to bias the solution obtained by the first norm, toward to a different solution, which is constrained to have some properties common in images. In most cases this property is that the image must have small energy at high-frequencies and thus be rather smooth [7]. The Laplacian operator can form the regularization matrix $\mathbf{Q}$. For a one dimensional signal this is given by

$$\mathbf{Q} = \begin{bmatrix} 2 & -1 & 0 & \cdots & -1 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & \cdots & -1 & 2 \end{bmatrix}.$$

This form of Laplacian operator is obtained by writing each of the pixels after application to the image of this operator as:

$$\mathbf{Qf}[i]=\mathbf{f}'[i]=-\mathbf{f}[i-1]+2\mathbf{f}[i]-\mathbf{f}[i+1],\quad i=1,...,N. \qquad (1.8)$$

Circularity implies that: $\mathbf{f}[-1]=\mathbf{f}[N],\mathbf{f}[N+1]=\mathbf{f}[1]$ and it is incorporated for the same reasons as with $\mathbf{H}$ and the covariance matrices $\mathbf{R}_{ff}$ and $\mathbf{R}_{nn}$. If $\mathbf{Q}$ is viewed as a filter its frequency response is $Q(w)=2(1-\cos w)$, which is clearly a high-pass filter.

Eq.(1.8) implies that neighboring pixels of the image must have similar values. From a mathematical point of view the Laplacian operator is the discrete analogous of the second derivative operation. Thus, introducing this derivative as a penalty term automatically constrains the restored image to have bounded discontinuities. The parameter a, which is called regularization parameter, expresses the degree of belief to the constrained term and so to the prior knowledge. The solution of minimization of Eq. (1.7) is obtained by:

$$\hat{\mathbf{f}}=\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}+a\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{g}. \qquad (1.9)$$

Parameter a has to be estimated and various ways of doing this can be found in [7]. Due to the properties of circulant matrices, Eq. (1.9) can be written in the frequency domain as:

$$\hat{\mathbf{F}}(i)=\frac{\Lambda_{\mathbf{h}}^{*}(i,i)}{|\Lambda_{\mathbf{h}}(i,i)|^{2}+a|\Lambda_{\mathbf{Q}}(i,i)|^{2}}\mathbf{G}(i)=\frac{|\Lambda_{\mathbf{Q}}(i,i)|^{-2}\Lambda_{\mathbf{h}}^{*}(i,i)}{|\Lambda_{\mathbf{h}}(i,i)|^{2}|\Lambda_{\mathbf{Q}}(i,i)|^{-2}+a}\mathbf{G}(i),\quad i=1,...,N.$$

It is very important to notice that Eq.(1.6) is identical to the above equation if the inverse image covariance equals to $\mathbf{Q}^{\mathrm{T}}\mathbf{Q}$, or setting the inverse of the image power spectrum equal to Laplacian operator's eigenvalues:

$$\mathbf{S}_{\mathbf{ff}}(i) = \Lambda_c(i,i)^{-1}, \quad i = 1,...,N.$$

From this relation the close connection between stochastic and deterministic approaches becomes obvious. Precisely the deterministic method can be derived from a correspondent stochastic, following the maximum a posteriori (MAP) paradigm and incorporating a-priori a probability density function (pdf) for the image, based on the SAR prediction model [21]:

$$p(\mathbf{f}) = Z_1^{-1} \exp\left\{-\frac{1}{2} a \|\mathbf{Qf}\|^2\right\}.$$

Due to the additive Gaussian noise of the degradation model, the conditional pdf of the observed image is:

$$p(\mathbf{g}|\mathbf{f}) = Z_2^{-1} \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{g}-\mathbf{Hf}\|^2\right\}.$$

Z_1 and Z_2 are normalizing constants. According to Bayes' rule the posterior pdf is:

$$p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f})\,p(\mathbf{f})}{p(\mathbf{g})},$$

where $p(\mathbf{g})$ is the marginal distribution of $\mathbf{g}$ which does not depend on $\mathbf{f}$. It is very convenient to estimate the image by the mode of this density and obtain the maximum a posteriori (MAP) estimation:

$$\hat{\mathbf{f}}_{MAP} = arg\max_{\mathbf{f}} p(\mathbf{f}|\mathbf{g}) = arg\max_{\mathbf{f}} p(\mathbf{g}|\mathbf{f})\,p(\mathbf{f}).$$

This is equivalent to minimize the negative log-likelihood of the posterior:

$$\hat{\mathbf{f}}_{MAP} = arg\min_{\mathbf{f}}\{-\log p(\mathbf{g}|\mathbf{f})\,p(\mathbf{f})\} = arg\min_{\mathbf{f}}\left\{\frac{1}{\sigma^2}\|\mathbf{g}-\mathbf{Hf}\|^2 + a\|\mathbf{Qf}\|^2\right\}.$$

The variance parameters can be merged to one, since this has no effect to the solution:

$$\hat{\mathbf{f}}_{MAP} = arg\min_{\mathbf{f}}\left\{\|\mathbf{g}-\mathbf{Hf}\|^2 + a'\|\mathbf{Qf}\|^2\right\}, \text{ where } a' = a\sigma^2. \qquad (1.10)$$

The relation to Eq.(1.7) is obvious. There is a more extended discussion about this relationship is given in Chapter III.

2. STATIONARY IMAGE RESTORATION

2.1 Motivation

The restoration methods discussed so far incorporate the stationary image assumption for the stochastic case and the spatially characteristics for the deterministic. Thus restoration is called stationary or spatially invariant, respectively. Despite the smaller complexity of these models, the main drawback is the inability to efficiently describe the nature of real images. More specifically, images contain edges, smooth areas, and areas with texture. Thus, a model that [illegible] must be spatially varying and cannot be stationary. The use of stationary models results in the appearance of restoration artifacts. The most common and known artifacts are the ringing artifact at the edges of the restored image and the texture-like shape in smooth areas due to noise amplification [3,6]. These artifacts are hard to suppress because if more regularization (lesser noise amplification) is used, more ringing artifact appear [7]. This motivated researchers to introduce spatial varying models.

Thus, the need for spatially adaptive methods has been motivated by heuristic arguments. However, an objective function is necessary to make restoration in this case rigorous. Images are destined to be observed by humans. Psychophysical experiments [8] showed that human visual system has the characteristic ability to be more sensitive to noise presented in smooth regions, in contrast to areas of high spatial activity where the noise is not so observable. Spatial invariance does not reflect this visual property of the human observer. As a result, the goal of spatial

CHAPTER II

PREVIOUS WORK ON NON STATIONARY IMAGE RESTORATION

2.1 Motivation

The restoration methods discussed so far incorporate the stationary image assumption for the stochastic case and unchanged spatially characteristics for the deterministic. Thus restoration is called stationary or spatially invariant, respectively. Despite the smaller complexity of these models, the main drawback is the inability to efficiently describe the nature of real images. More specifically, images contain edges, smooth areas, and areas with texture. Thus, a model that captures the image properties has to be spatially varying and cannot be stationary. The use of stationary models results in the appearance of restoration artifacts. The most common and known artifacts are the ringing artifact at the edges of the restored image and the texture-like shape in smooth areas due to noise amplification [5,6]. These artifacts are hard to suppress because if more regularization (lesser noise amplification) is used, more ringing artifact appear [7]. This motivated researchers to introduce spatial varying models.

Thus, the need for spatially adaptive methods has been motivated by heuristic arguments. However, an objective function is necessary to make restoration in this case rigorous. Images are destined to be observed by humans. Psychophysical experiments [8] showed that human visual system has the characteristic ability to be more sensitive to noise presented in smooth regions, in contrast to areas of high spatial activity where the noise in not so observable. Spatial invariance does not respect this visual property of the human observer. As a result, the goal of spatial

adaptive restoration is to make adaptive regularization dependant on the image local intensity transitions. Smooth regions need more regularization. In contrast, regions with edges are not affected so much from noise amplification. Also less regularization means more edge preservation. Thus for goal of spatially adaptive methods, the trade-off between regularization and noise amplification is not a severe.

2.2 Visibility Function Based Non-Stationary Restoration

In [8] a measure of spatial detail was defined by a noise masking function $M(i)$ at pixel i. The visibility function $f(i)$ was defined, which express the relationship between the noise visibility and the masking function, experimentally. Following in [9] the masking function was set to be the local variance of the pixel and accordingly to [8] the visibility function was defined to be:

$$f(i)=\frac{1}{\theta M(i)+1}, \quad i=1,...,N$$

where θ is a scale parameter depending on the image. This function goes to zero when local variance goes to infinity (pixel belongs to an edge) and to one when the variance close to zero (pixel belongs to a smooth region). Finally, as a result of an extended analysis (as in [9]), in contrast to Euclidean norm as in Eq. (1.7), a weighted norm is introduced for the penalty term:

$$a\|\mathbf{Qf}\|_{\Lambda}=a\mathbf{f}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\Lambda\mathbf{Qf},$$

where Λ is a $N\times N$ diagonal matrix with elements: $\Lambda(i,i)=f^2(i)$, $i=1,...,N$. The matrix $\mathbf{Q}$ is recommended to be a high-pass filter and the example of the p-th order (discrete) derivative operator is given. Notice that in Eq.(1.7) this is for $p=2$. The

resulting linear system $\left(\mathbf{H}^T\mathbf{H}+\alpha\mathbf{Q}^T\Lambda\mathbf{Q}\right)\hat{\mathbf{f}}=\mathbf{H}^T\mathbf{g}$ is solved iteratively with a constraint gradient descent algorithm.

The weights can be assumed known or can be approximated by a non-adaptive method. In [12] the weights are evaluated at every step from the partially restored image. Furthermore, in [11] the restored image is constrained to belong to a convex set (hard regularization). This achieved by application of a projection operator at every step.

2.3 Markov Random Field (MRF) Non-Stationary Restoration

The adaptive methods mentioned so far are deterministic. However, many non-stationary stochastic models have been proposed. Markov Random Fields (MRFs) are very known and used models in image processing. The main success point of MRFs is that can model local characteristics using local conditional probabilities. The physical analogous of these conditionals probabilities in images is the (realistic) assumption that a local pixel's intensity depends only from the neighboring pixels. An MRF describes a non-stationary stochastic process, in which elements are labeled with states following a procedure that obeys the conditional probabilities. (Note: there are probabilities and not probability density functions, due to the discrete nature of an MRF).

One of the most important works is that of Geman & Geman [12]. They made an analogy between statistical physics theory and image processing, by introducing MRFs to image restoration. At this work (as many later works) an image is regarded

as a pair $(\mathbf{F},\mathbf{L})$, where $\mathbf{F}$ is the $N\times N$ matrix of the observable pixel intensities and represents the *intensity process*, and $\mathbf{L}$ is a $N\times N$ matrix with unobservable edge binary elements (indicating when there is an edge or not between two pixels) and represents the *line process*. For brevity the later process will be temporally ignored, because can be easily introduced using a theorem that follows. For the present time the image is assumed to be a realization of an MRF process. Specifically, each pixel's $s\in S$ intensity is the realization random variable X_s, where $S=\{s_1,s_2,...,s_N\}$ is the set of image pixels. Each random variable is assigned with values (states) from a finite set $\Lambda=\{0,1,...,L-1\}$. Let $X=\{X_1,X_2,...,X_N\}$ be the set of the random variables. Thus a realization x_s of X_s belongs to this set $X_s=x_s\in\Lambda$ Then $\Omega=\{\omega=(x_{s_1},x_{s_2},...,x_{s_N})\}$ is the set of all possible configurations. Moreover the set $G=\{G_s,\forall s\in S\}$ is defined as a *neighborhood system*, where G_s is the set of neighbors of site s. For the intensity process $\mathbf{F}$, the four (or more) nearest pixels can be defined as neighbors. With this notation X is an MRF over the graph $\{S,G\}$ if:

$$P(X=\omega)>0,\qquad \forall\omega\in\Omega,$$

$$P(X_s=x_s\,|\,X_r=x_e,\forall r\in S)=P(X_s=x_s\,|\,X_r=x_r,\forall r=G_s).$$

In the previous chapter, the MAP estimator was presented. The same methodology can be adopted for MRFs; incorporate prior knowledge for the image with an MRF and find the MAP solution given the degradation model and the degraded image. Let $P(G=g\,|\,X=x)$ and $P(X=x)$ be the observed image density and prior respectively, where G is the random variable of the degradation process (it is a

stochastic degradation process due to random noise) and g a realization of it (the observed image). Then the posterior distribution, according to Bayes' Rule, is:

$$P(X=x|G=g)=\frac{P(G=g|X=x)P(X=x)}{P(G=g)},$$

and omitting the constant $P(G=g)$ the MAP estimation of the image is:

$$\hat{\mathbf{f}}_{MAP}=arg\max_{f}P(X=x|G=g)=arg\max_{f}P(G=g|X=x)P(X=x).$$

The disadvantage of this estimation is that the prior distribution of the image $P(X)$ must be determined by the conditional probabilities. This is a very difficult and usually impossible task. Fortunately, there is a connecting theorem between MRFs and Gibb's distributions. Preserving the previous notation, a probability distribution $\pi(\omega)$ relative to the graph $\{S,G\}$ is a Gibb's distribution if it is of the form:

$$\pi(\omega)=\frac{1}{Z}\exp\left\{-\frac{1}{\mathrm{T}}U(\omega)\right\}=\frac{1}{Z}\exp\left\{-\frac{1}{\mathrm{T}}\sum_{\forall C\in\mathbb{C}}V_C(\omega)\right\},\ \omega\in\Omega,$$

denoting by $\mathbb{C}$ the set of cliques of the graph. The term $V_C(\omega)$ is a function of the configuration ω only of the sites that belong to clique C and is called *potential function*. The normalizing constant Z is the *partition function*. The constant T is the temperature. The Hammersley-Clifford theorem states that a random variable X over a graph $\{S,G\}$ is an MRF if and only if it follows a Gibb's probability distribution, hence:

$$\pi(\omega)=P(X=\omega).$$

This is a powerful tool to define MRFs in a straightforward manner by the construction of the potential functions, avoiding the definition of the local conditional probabilities. For example, Geman & Geman used the following potential:

$$V_{r,s} = \begin{cases} 1, & if \quad x_r = x_s \\ 0, & if \quad x_r \neq x_s \end{cases}.$$

One other difficulty that can be solved is the introduction of the line process to the model, by introducing potentials as functions both of intensity and line process. Thus the MRF is expanded with the joint process $(\mathbf{F},\mathbf{L})$ and the neighborhood system is expanded to include the line process. Thus an element of $\mathbf{L}$, i.e. l_r can have five neighbors (the nearest edge points) and can take binary values (or more), indicating when there is an edge or no. Now one can define the potential with a manner that penalizes pixels placed between edges and have related values, and encourages pixels belonging to the same edge to take similar values. For example, the following potential can be defined:

$$V_{r,s}' = \begin{cases} 0, & if \quad l_d = 1 \\ V_{rs}, & if \quad l_d = 0 \end{cases},$$

where r,s are two neighboring pixels and l_d is the edge element between them. Because this function depends only on neighboring pixels (cliques), the proposed model is an MRF.

Finally, Geman & Geman proposed the Gibb's sampler embedding a simulating annealing process (temperature T is for that reason) for finding the global mode of the posterior probability. First they showed that the posterior is also a Gibb's distribution assuming a more general degradation model (the present is a special case). Thus the degradation process is an MRF. Then they defined the local conditionals for the update equations for the optimization algorithm.

Despite the fact that MRFs are general non-stationary models, it is important to notice that when the potential function is the same for all pixels, they are homogeneous (stationary). The model can be easily converted to non-homogeneous when the potential function varies spatially. Nevertheless, the posterior probability is in general non-stationary thus the restoration is spatially adaptive.

The connection between stochastic and deterministic methods appears also in MRF models. The potential function can also be seen as a regularization function. These functions help to be quadratic because lead to simple linear inverse filters. However they suffer from the drawback of restoration artifacts, such that have mentioned in the previous chapter. Linearity cannot preserve high order discontinuities such as edges and texture. Some example works on non-quadratic restoration are [13, 14]. Non-quadratic functions have a disadvantage: the optimization is very difficult, because non-quadraticity means non-convexity, which leads to local minima. An important effort to solve this problem is that of D. Geman and C. Yang [15]. They proposed an expansion of the regularization function $\Phi(\tilde{X})$, introducing auxiliary variables following the property, (such as analog line process):

$$\Phi(X) = \arg\min_{b} \Phi^{*}(X,b)$$

The function $\Phi^{*}(X,b)$ has the same minima with $\Phi(X)$. Also it is half-quadratic with respect to X, which means fixing the auxiliary variables, the function becomes quadratic. In the probabilistic point of view the probability distribution of (X,b) (the random variable is the image $X = \mathbf{f}$) is half-gaussian. However $\Phi^{*}(X,b)$ is non-convex. The minimization of the new objective function is far more difficult due to the non-convexity:

$$\hat{\mathbf{f}}_{MAP} = arg\min_{\mathbf{f}} \left\{ \left\| \mathbf{g} - \mathbf{H}\hat{\mathbf{f}} \right\|^2 + \lambda \arg\min_{b} \Phi^{*}(X, b) \right\},$$

where λ is a regularization parameter. Geman and Yang proposed an optimization by implementation of the Gibb's sampler (with simulated annealing) and using the half-quadratic property.

An also interesting work on half-quadratic functions appears in [16]. First some conditions for these functions are imposed and then a general iterative optimization algorithm is given. To find the minimum of the objective function, introduce the auxiliary variables as in [15] and then iterate between two steps: (deterministic) minimization with respect to image $\mathbf{f}$ and then to b. The latter step is very simple. A proof of convergence is given.

In [17] theoretical aspects about the convexity of the objective function are discussed. More specifically Idier in [17] was able to find the conditions under which the convexity of the objective function guarantees the convexity of the regularized function, which was proposed by [15] and [14].

Generally MRF are very effective and model well images attributes. But the main drawback, as it appears in this discussion, is the complexity of estimating their parameters. This is due to the cyclic relationships between the variables. See Chapter III for the graphical models reference [31].

2.4 Wavelet Based Non-Stationary Restoration

Another used mathematical tool for non-stationary image restoration is the wavelet framework. For example, in [18] the matrix-vector multiplication as in Eq.(1.1) is reformulated based in a wavelet subband decomposition, which allows the computation of both the convolution operator and the image in the wavelet domain. Thus, the problem is altered to multichannel restoration. At each channel stationary restoration was applied, but the more channels used to model the image, the more spatially adaptive the restoration of the entire image (the composition of the channels) is. The restoration filter used there was the LMMSE.

In another work [19] a multiscale Kalman smoothing is used directly to the wavelet coefficients of a pre-filtered image. Pre-filtering by a constrained least square estimate (CLSE) filter produces an under-regularized solution. Post-filtering by the wavelet methodology, enables smoothing over the desired regions, and not on the edges where edge preservation is critical.

CHAPTER III

THE BAYESIAN METHODOLOGY

The Bayesian methodology is very popular and has been used successfully in many scientific fields. The Bayesian paradigm is applied to solve statistical inference problems. Since image restoration boils down to an inference problem, the Bayesian methodology is a very powerful tool for such problems. Inference is in essence the estimation of unknown random variables based on certain observations and a stochastic model [29]. Inverse problems can be viewed as inference problems because the unknown data is obtained from the observations of a stochastic model.

Stochastic models have parameters that need to be estimated. A very popular method is the use of the maximum likelihood (ML) estimator. This estimator is from the relationship:

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} p(\mathbf{y};\boldsymbol{\theta}), \qquad (3.1)$$

where $\mathbf{y}$ is the vector of the observations (data produced from the stochastic model), $\boldsymbol{\theta}$ is the vector consisting of the parameters and $p(\mathbf{y};\boldsymbol{\theta})$ the probability density function (PDF) of the data. Usually this likelihood is not directly known, because the generative model may have 'hidden' (or 'latent') random variables that connect the observations with the model parameters [35]. The hidden variables are unobserved and they are very important because they include all the information that we need to solve the inference problem. The introduction of the 'hidden' variables to the problem can be made by writing the likelihood of Eq.(3.1) using Bayes' rule and then marginalizing them according to the integral [31]:

$$p(\mathbf{y};\boldsymbol{\theta}) = \int p(\mathbf{y},\mathbf{x};\boldsymbol{\theta})d\mathbf{x} = \int p(\mathbf{y}\,|\,\mathbf{x};\boldsymbol{\theta})\,p(\mathbf{x};\boldsymbol{\theta})d\mathbf{x}, \qquad (3.2)$$

where $p(\mathbf{x};\boldsymbol{\theta})$ is the a priori PDF for the hidden variables $\mathbf{x}$. This is one of the basic principles of the Bayesian methodology which allows us to incorporate all the information that we have for $\mathbf{x}$ by means of the prior probability distribution. Estimation of the parameters can be based on the equations (3.1) and (3.2). Then, Bayes' rule is used to obtain the conditional posterior probability distribution of $\mathbf{x}$ under the observed data:

$$p(\mathbf{x}\,|\,\mathbf{y};\boldsymbol{\theta}) = \frac{p(\mathbf{y},\mathbf{x};\boldsymbol{\theta})}{p(\mathbf{y};\boldsymbol{\theta})} = \frac{p(\mathbf{y}\,|\,\mathbf{x};\boldsymbol{\theta})\,p(\mathbf{x};\boldsymbol{\theta})}{p(\mathbf{y};\boldsymbol{\theta})}. \qquad (3.3)$$

The estimate of hidden can be the mode or the mean of the posterior probability:

$$\hat{\mathbf{x}} = arg\max_{\mathbf{x}} p(\mathbf{x}\,|\,\mathbf{y};\boldsymbol{\theta}), \qquad (3.4)$$

$$\text{or} \qquad \hat{\mathbf{x}} = E\left[p(\mathbf{x}\,|\,\mathbf{y};\boldsymbol{\theta})\right]. \qquad (3.5)$$

The Expectation-Maximization (E-M) algorithm a very popular algorithm to find ML estimates in a sense alternates between the equations (3.2) and (3.4), for details see for example [36, 37].

In the Bayesian philosophy probability is treated as degree of belief or knowledge. It is obvious that the Bayesian methodology allows us to incorporate to the stochastic model used for estimation, the a priori knowledge about the hidden variables. This knowledge is represented by the prior PDF.

But this methodology has a serious drawback. In most models of interest the integral of Eq. (3.2) is intractable, since it is difficult or impossible to obtain it in a closed form. Thus, a number of approximations have been used by researchers of the years to

compute this integral, see for example [29, 31, 37]. The simplest approximation is to assume the joint probability (the integrand) is a delta function at the mode of this function [37]:

$$p(\mathbf{y}\,|\,\mathbf{x};\boldsymbol{\theta})\,p(\mathbf{x};\boldsymbol{\theta}) \approx p(\mathbf{y}\,|\,\mathbf{x};\boldsymbol{\theta})\,p(\mathbf{x};\boldsymbol{\theta})\,\delta(\mathbf{x}-\hat{\mathbf{x}}), \qquad (3.5)$$

where,

$$\hat{\mathbf{x}} = arg\max_{\mathbf{x}} p(\mathbf{y},\mathbf{x};\boldsymbol{\theta}). \qquad (3.6)$$

Then, the parameters are estimated by:

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} \int p(\mathbf{y}\,|\,\mathbf{x};\boldsymbol{\theta})\,p(\mathbf{x};\boldsymbol{\theta})\,\delta(\mathbf{x}-\hat{\mathbf{x}})\,d\mathbf{x} = \arg\max_{\boldsymbol{\theta}} p(\mathbf{y}\,|\,\hat{\mathbf{x}};\boldsymbol{\theta})\,p(\hat{\mathbf{x}};\boldsymbol{\theta}). \qquad (3.7)$$

Of course this method does not to follow exactly the Bayesian methodology; and it is called maximum a posteriori (MAP) estimation.

Another approximation is often used when only a part of the hidden variables can be integrated out explicitly. Let us assume the two sets of hidden variables $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2\}$. Then, we can first integrate out $\mathbf{x}_1$ and then approximate the remaining integral using its mode as in the MAP methodology according to the equation:

$$p(\mathbf{y};\boldsymbol{\theta}) \approx \int p(\mathbf{y}\,|\,\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta})\,p(\mathbf{x}_1,\mathbf{x}_2;\boldsymbol{\theta})\,\delta(\mathbf{x}_2-\hat{\mathbf{x}}_2)\,d\mathbf{x}_1 d\mathbf{x}_2 = p(\mathbf{y}\,|\,\hat{\mathbf{x}}_2;\boldsymbol{\theta})\,p(\hat{\mathbf{x}}_2;\boldsymbol{\theta}), \qquad (3.8)$$

where,

$$\hat{\mathbf{x}}_2 = arg\max_{\mathbf{x}_2} p(\mathbf{y},\mathbf{x}_2;\boldsymbol{\theta}). \qquad (3.9)$$

This approach of course is not full Bayesian, but may be called partially-Bayesian.

It is interesting to notice the solution of equations (3.6) and (3.7) is obtained by the Bayes' rule of Eq. (3.3) and it is the mode of the posterior probability distribution. For example equation (3.6) is:

$$\hat{\mathbf{x}} = arg\max_{\mathbf{x}} p(\mathbf{y},\mathbf{x};\boldsymbol{\theta}) = arg\max_{\mathbf{x}} \frac{p(\mathbf{y},\mathbf{x};\boldsymbol{\theta})}{p(\mathbf{y};\boldsymbol{\theta})} = arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y};\boldsymbol{\theta}), \quad (3.10)$$

because the likelihood $p(\mathbf{y};\boldsymbol{\theta})$ is constant with respect to $\mathbf{x}$. Then the optimization algorithm for both MAP and partial-Bayesian approaches, alternates between (approximate) likelihood maximization and hidden variables estimation. This is exactly is done when the previous mentioned E-M algorithm is used.

In Chapter I an example of a stationary stochastic model and a method of finding the solution was given with Eq. (1.10), when the all parameters are known. In this image restoration example, using the Bayesian framework $\mathbf{g}$ is defined as the observed data, since only the degraded image is known, and $\mathbf{f}$ is the 'hidden' variables (the unknown original image). These are random variables with known joint probability distribution:

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

where the conditional is:

$$p(\mathbf{g}|\mathbf{f}) = N(\mathbf{H}\mathbf{f},\sigma^2)$$

and the prior on the hidden variables is:

$$p(\mathbf{f}) \propto N\left(0,\left(a\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\right)^{-1}\right).$$

The parameters of the model, denoted by a vector $\boldsymbol{\theta}$, are the variances:

$$\boldsymbol{\theta} = \left[a,\sigma^2\right].$$

Following the Bayesian methodology, if we manage to estimate the parameters (for example via the E-M algorithm) then the estimation of the hidden data is given as in Eq.(1.10) from Bayes' rule:

$$\hat{\mathbf{f}} = arg\max_{\mathbf{f}} p(\mathbf{f} \mid \mathbf{g};\boldsymbol{\theta}) = arg\min_{\mathbf{f}} \left\{ \frac{1}{\hat{\sigma}^2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \hat{a} \|\mathbf{Q}\mathbf{f}\|^2 \right\}.$$

where $\hat{\sigma}^2, \hat{a}$ are the estimated parameters.

A stochastic generative model can be described by a graph, which is called a *graphical model* [31]. A graphical model describes the data generation process, defining random variables and relations between them. For the previous example the graphical model is shown in Figure 3.1. The cycles denote random variables and the square the parameters that governs the probability distributions. This graphical model is a directed acyclic graph (DAG).

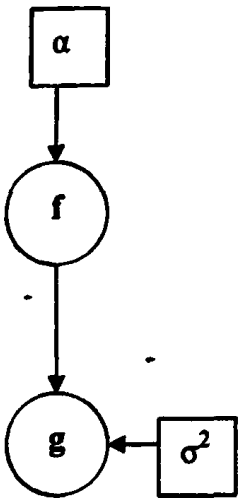

Figure 3.1: Graphical model of the degradation process.

Of course more variables can be introduced to the model and make it more complex. Following from the Bayesian methodology, hierarchical priors can be introduced to the model. At this case one parameter of the model can be assumed stochastic (hence a hidden random variable) with a prior distribution. The Bayesian formulation is the same, with only difference that the set of hidden variables is altered. For example if

we assume θ to be random variable, with prior (named hyperprior): $p(\theta;\theta')$, where θ' are the parameters (hyperparameters), and $(\mathbf{f},\theta)$ are the hidden variables. The hierarchical model can also be described by a graphical model (Figure 3.2). The parameters are omitted for brevity. The directed graph denotes a hierarchy between the random variable set. Not all the hidden are the same, but they belong to different levels of hierarchy. Then to obtain an estimation of the hidden data **f**, we can use Bayes' rule, but with the difference that the remaining hidden data (belonging to a different level of hierarchy) must be integrated out:

$$\hat{\mathbf{f}} = arg\max_{\mathbf{f}} p(\mathbf{f} \mid \mathbf{g};\theta') = arg\max_{\mathbf{f}} \int p(\mathbf{f},\theta \mid \mathbf{g};\theta')\, d\theta .$$

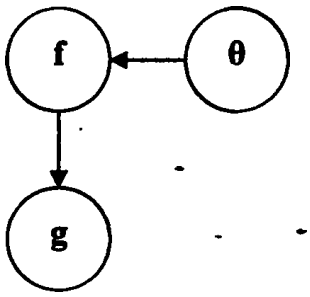

Figure 3.2: Graphical model after the hyperprior introduction.

It is worth to mention that MRF are described by undirected graphical models [31]. An interesting discussion about Bayesian methods and regularization in image processing can be found in [7, 20].

CHAPTER IV

NON-STATIONARY IMAGE MODEL BASED ON HIERARCHICAL PRIORS

As discussed in the previous Chapters, Bayesian methods have been applied extensively for many signal processing problems including image restoration. The Bayesian formulation offers many advantages for the image restoration problem since it allows the incorporation of a priori knowledge in the form of priors about the image and the unknown parameters. Strictly speaking the maximum a posteriori (MAP) methodology is *not Bayesian* since the mode of the posterior which is used in MAP may not be representative of the posterior distribution which is integrated in Bayesian formulation [31].

In many Bayesian and MAP formulations for the image restoration problem Gaussian stationary models have been used for the image prior; see for example [21, 23]. A very popular model is the simultaneously autoregressive (SAR) in which the statistics of the image are assumed invariant for the different spatial locations (has already been introduced in Chapter I); see for example [21-23].

This model greatly facilitates the parameter estimation process since only one parameter is used and thus can be easily estimated. However, it is seriously handicapped because it does not provide the flexibility to model the spatially varying correlations of the image. In other words, such prior enforces smoothness uniformly across the entire image ad corresponds to uniform "regularization". Furthermore, the SAR model assumes Gaussian statistics for the autoregressive predictor residuals.

This assumption is well known to be inaccurate for certain images where these residuals are heavy tailed due to large prediction errors in the regions of the image edges and texture.

There have been numerous efforts to ameliorate either the problems of uniform regularization or the Gaussian statistics assumption for the residuals in image restoration. However, there has not been an attempt to ameliorate both of them simultaneously. One of the most successful such efforts to ameliorate the first problem has used spatially adaptive regularization [8,9,26]. The motivation and the justification for this approach is based on psycho visual arguments about the visibility of the noise in images. Furthermore, for its application the parameters used to define the noise visibility weights are selected in an ad hoc manner. There have been numerous efforts to ameliorate the second problem, see for example [32] and [33]. However, in most these methods the parameters of the statistical models that are used are not estimated but are assigned empirically.

In this thesis we first propose new non stationary image prior models which incorporate both spatially varying variances for the SAR predictor residuals and generalized Gaussian statistics. Thus this prior provides the flexibility to model both the spatially varying correlations of images and the long tailed behavior of the SAR prediction residuals in edge and texture areas. Based on this model for the prior we present a MAP based methodology in which all the restored image and all parameters can be estimated. However, due to the non stationary nature of the image prior model it contains too many parameters. Thus, the MAP estimate of all unknown parameters turns out to be unreliable. For this purpose in order to ameliorate the estimation of the

spatially varying variances a hierarchical model is also proposed with hyperpriors within and a Bayesian setting. More specifically, based on the graphical model for the observations the likelihood with respect to the "hidden" variables is marginalized [29].

4.1 Imaging and Image Models

In the introduction the model of the degradation process was given. The imaging model is linear. Let $\mathbf{g}$ be a $N\times 1$ vector, representing the observed degraded image. We assume that this image is formed as

$$\mathbf{g}=\mathbf{H}\mathbf{f}+\mathbf{n}, \qquad (4.1)$$

where, $\mathbf{f}$ the unknown $N\times 1$ original image to be estimated, $\mathbf{H}$ a $N\times N$ known degradation matrix, and $\mathbf{n}$ additive white noise. We assume Gaussian statistics for the noise given by:

$$\mathbf{n}\sim N\left(\mathbf{0},\beta^{-1}\mathbf{I}\right),$$

where $\mathbf{0}$ and $\mathbf{I}$ are a $N\times 1$ vector with zeros and the $N\times N$ identity matrix, respectively, and β the inverse of the noise variance is assumed unknown. The image $\mathbf{f}$ is assumed to be generated by a zero mean SAR prediction model [21], given by:

$$\mathbf{f}(k,j)=\frac{1}{4}\sum_{l=-1}^{1}\sum_{m=-1}^{1}\mathbf{f}(k+l,j+m)+\varepsilon(k,j),\ m\neq l$$

with $\varepsilon(k,j)$ the prediction residual for the image location (k, j). With out loss of generality, in what follows we use for convenience one dimensional notation, as we have done in this thesis so far. The above equation can be also written in matrix vector form for the entire image as:

$$\mathbf{Q}\mathbf{f}=\boldsymbol{\varepsilon},$$

where $\mathbf{Q}$ is a $N \times N$ matrix operator (the already introduced Laplacian operator), and $\varepsilon = [\varepsilon(1), \varepsilon(2), ..., \varepsilon(N)]^T$ the $N \times 1$ vector of the residuals. We assume that the residuals have Generalized Gaussian statistics which induces prior for the image given by:

$$\varepsilon(i) \sim p(\varepsilon_i) = \frac{c a_i^{\frac{1}{c}}}{2\Gamma(1/c)} \exp\left\{-a_i |\varepsilon(i)|^c\right\}$$

where:

$$a_i = (k(c))^c \sigma_i^{-c}, \; k(c) = \sqrt{\frac{\Gamma(3/c)}{\Gamma(1/c)}} \quad (4.2).$$

$\Gamma(\cdot)$ is the Gamma function. The standard deviation of the GGD prediction error at location i is denoted by σ_i. The parameter c determines the shape of the pdf. For the special case where $c = 1$ we have the Laplacian distribution, and for $c = 2$ the Gaussian.

This model is *non-stationary*, because the standard deviation changes spatially. It introduces N parameters a_i's that have to be estimated from N data points, which is clearly not a desirable situation from an estimation point of view. For this purpose apart from the MAP approach we also propose a Bayesian methodology to bypass this difficulty and we introduce a Gamma hyperprior for all the a_i's. The rational for using this Gamma prior in the non stationary case is threefold. First, it is "conjugate" for the variance of a Gaussian and ameliorates the over parameterization problem of this model. Second, similar hierarchical models have been used successfully in Bayesian formulations of other statistical learning problems; see for example [8]. Finally, as we shall see in what follows it produces update equations for the a_i's

previously derived using different principles. We parameterized the Gamma hyperprior as:

$$p(a_i) = \frac{1}{\Gamma(l/2)} \alpha^{\left(\frac{l-2}{2}\right)} \exp\{-m(l-2)a_i\}. \qquad (4.3)$$

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

$$E[a_i] = \frac{l}{2m(l-2)}, \text{ and } Var[a_i] = \frac{l}{2m^2(l-2)^2},$$

respectively [11]. This representation is used because the value of the parameter l can be interpreted as the level of confidence to the prior knowledge provided by the Gamma hyperprior [2, 9]. More specifically, as:

$$l \to \infty \text{ then } E[a_i] \to \frac{1}{2m} \text{ and } Var[a_i] \to 0.$$

Thus the prior becomes very restrictive. In contrast, as:

$$l \to 2 \text{ both } E[a_i] \to \infty \text{ and } Var[a_i] \to \infty,$$

thus the prior becomes uninformative.

The hierarchical model, from the Bayesian perspective has been build. The graphical model is shown at Figure 4.1.

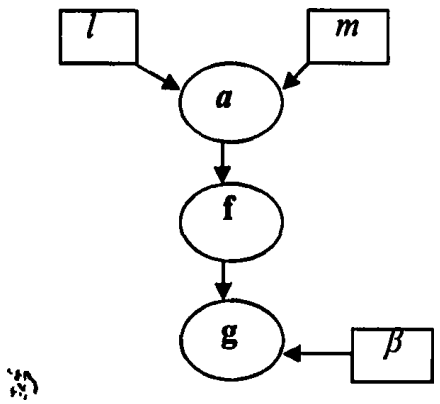

Figure 4.1: Graphical model of the observations

In the next section the restoration algorithms are described, based on the proposed model and the Bayesian methodology discussed in Chapter III.

4.2 The Restoration Algorithms

4.2.1 Maximum a posteriori (MAP) approach

At first we propose a MAP approach to estimate all the parameters of our model and the restored image. This is based on maximization of the posterior probability using Bayes rule. This is given by:

$$p(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l) = p(\mathbf{g}|\mathbf{f};\beta)\,p(\mathbf{f}|\mathbf{a})\,p(\mathbf{a};m,l).$$

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

$$\left[\mathbf{f}^*,\mathbf{a}^*,\beta^*,m^*,l^*\right] = \arg\max_{[\mathbf{f},\mathbf{a},\beta,m,l]} \log p(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l) = \arg\min_{[\mathbf{f},\mathbf{a},\beta,m,l]} \log J(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l)$$

where:

$$J(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l) =$$

$$-\frac{N}{2}\log\beta + \frac{1}{2}\beta\|\mathbf{Hf}-\mathbf{g}\|^2 - \frac{1}{c}\sum_{i=1}^{N}\log a_i + \sum_{i=1}^{N} a_i \left|[\mathbf{Qf}](i)\right|^c - \frac{l-2}{2}\sum_{i=1}^{N}\log a_i + m(l-2)\sum_{i=1}^{N} a_i. \quad (4.4)$$

Setting $\nabla_a J(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l) = 0$ yields

$$a_i^* = \left(\frac{1}{c}+\frac{1}{2}(l-2)\right)\left(\left|[\mathbf{Qf}](i)\right|^c + m(l-2)\right)^{-1} \quad (4.5)$$

Taking derivative and setting to zero, we have the maximum for β is

$$\beta = \frac{N}{\|\mathbf{g}-\mathbf{Hf}\|^2} \qquad (4.6)$$

To find the parameters m,l that maximize the approximation, it is enough to estimate the parameters of the gamma distribution, treating a_i's as samples from that gamma distribution. Formally this means:

$$\left[m^*,l^*\right]=\arg\max_{[m,l]}\log p\left(\mathbf{g}\,|\,\mathbf{f};\beta\right)p\left(\mathbf{f}\,|\,\mathbf{a}\right)=\arg\max_{[m,l]}p\left(\mathbf{a};m,l\right)$$

It is known that the ML estimated parameters of the Gamma distribution $a\sim Gam\left(a\,|\,b,d\right)$ from its samples are given from the relationships [33]:

$$b=\frac{\bar{a}}{s^2} \qquad (4.7)$$

$$d=\left(\frac{\bar{a}}{s^2}\right)^2 \qquad (4.8)$$

where:

$$\bar{a}=\frac{1}{N}\sum_{i=1}^{N}a_i \qquad (4.9)$$

$$s^2=\frac{1}{N}\sum_{i=1}^{N}\left(a_i-\bar{a}\right)^2 \qquad (4.10)$$

Easily derived from the above, each parameter is:

$$l^*=\frac{2\bar{a}}{s^2} \qquad (4.11)$$

$$m^*=\left(\frac{\bar{a}}{s^2}\right)^2\frac{1}{l^*-2} \qquad (4.12)$$

For the Gaussian case (c=2) $\nabla_{\mathbf{f}}J\left(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l\right)=0$ yields

$$\mathbf{f}^*=\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}+\boldsymbol{\beta}^{-1}\mathbf{Q}^{\mathrm{T}}\mathbf{A}\mathbf{Q}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{g}. \qquad (4.13)$$

For all other values of c, $\mathbf{f}^*$ cannot be found in closed form.

4.2.2 Bayesian (partial) Algorithm

For the Bayesian formulation we select first what will be considered as hidden variables and what as parameters. A graphical model is used that describes the observed data generation process and is shown in *Figure 5*. In this figure ellipses represent the random variables and rectangles the parameters. Thus, $\mathbf{f}$ and $\mathbf{a}$ are considered "hidden" (latent) variables, while m, l and β are unknown parameters. In the Bayesian inference paradigm hidden variables are marginalized while parameters are estimated [10]. Given the observations $\mathbf{g}$, the parameters are estimated by maximizing the likelihood $p(\mathbf{g};\beta,m,l)$.

Based on the graphical model in *Figure 4.1* the likelihood is obtained by marginalizing the joint probability density function (pdf) according to Bayes rule:

$$p(\mathbf{g};\beta,m,l)=\int\int p(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l)d\mathbf{f}d\mathbf{a}=\int\int p(\mathbf{g}\,|\,\mathbf{f};\beta)p(\mathbf{f}\,|\,\mathbf{a})p(\mathbf{a};m,l)d\mathbf{f}d\mathbf{a}, \quad (4.14)$$

where, $d\mathbf{a}=da_1da_2...da_N$. The exact evaluation of the complete Bayesian integral is not possible thus we resort to an approximation. In this approach we evaluate only the integral for the variable $\mathbf{a}$.

$$\begin{aligned} p(\mathbf{g};\beta,m,l)&=\int\int p(\mathbf{g}\,|\,\mathbf{f};\beta)\,p(\mathbf{f}\,|\,\mathbf{a})\,p(\mathbf{a};m,l)\,d\mathbf{f}d\mathbf{a}= \\ &=\int p(\mathbf{g}\,|\,\mathbf{f};\beta)\Big(\int p(\mathbf{f}\,|\,\mathbf{a})\,p(\mathbf{a};m,l)\,d\mathbf{a}\Big)d\mathbf{f} \end{aligned} \quad (4.15)$$

The evaluation of the inner integral becomes:

$$\int p(\mathbf{f}\,|\,\mathbf{a})\,p(\mathbf{a};m,l)\,d\mathbf{a}=$$

$$=\Gamma\left(\frac{l}{2}\right)^{-N}\left(m\left(l-2\right)\right)^{N\frac{l}{2}}\Gamma\left(\frac{l}{2}+\frac{1}{c}\right)^{N}\prod_{i=1}^{N}\left(m(l-2)+\left\|\left[\mathbf{Qf}(i)\right]\right\|^{c}\right)^{-\frac{l}{2}-\frac{1}{c}} \quad (4.16)$$

where the term outside the integral is the constant of the original gamma pdf. To explain better this result we must say that all of the N terms were the integration of a gamma function. Thus, integration of them produces the constant term of a gamma density function. Combining equations (4.15) and (4.16), it is:

$$p(\mathbf{g};\beta,m,l) =$$

$$\Gamma\left(\frac{l}{2}\right)^{-N}\Gamma\left(\frac{l}{2}+\frac{1}{c}\right)^{N}\left(m(l-2)\right)^{\frac{Nl}{2}}\int\exp\left\{-\frac{1}{2}\beta\|\mathbf{Hf}-\mathbf{g}\|^2\right\}\prod_{i=1}^{N}\left(m(l-2)+\left|[\mathbf{Qf}(i)]\right|^{c}\right)^{-\frac{l}{2}-\frac{1}{2}}d\mathbf{f}$$

The remaining integral is impossible to be evaluated. Thus we resort to the previous introduced MAP methodology. At the same manner the approximation of the integral is the value of the integrand at the maximum. We want to minimize a new function:

$$\left[\mathbf{f}^*,\beta^*,m^*,l^*\right] = \arg\max_{[\mathbf{f},\beta,m,l]} p(\mathbf{g}\,|\,\mathbf{f};\beta)\,p(\mathbf{f};m,l) = \arg\min_{[\mathbf{f},\beta,m,l]} J'(\mathbf{g},\mathbf{f};\beta,m,l),$$

where:

$$J'(\mathbf{g},\mathbf{f};\beta,m,l) = -N\log\Gamma\left(\frac{l}{2}\right)+\frac{Nl}{2}\log\left(m(l-2)\right)+N\log\Gamma\left(\frac{l}{2}+\frac{1}{c}\right)-\frac{N}{2}\log\beta+$$

$$+\frac{1}{2}\beta\|\mathbf{Hf}-\mathbf{g}\|^2-\left(\frac{l}{2}+\frac{1}{c}\right)\sum_{i=1}^{N}\log\left(m(l-2)+\left|[\mathbf{Qf}(i)]\right|^{c}\right)$$

The derivative with respect to $\mathbf{f}$ is:

$$\nabla_{\mathbf{f}}J'(\mathbf{g},\mathbf{f};\beta,m,l) = \frac{\partial J(\mathbf{g},\mathbf{f},\mathbf{a};\beta,m,l)}{\partial\mathbf{f}} = -\beta\mathbf{H}^{\mathrm{T}}(\mathbf{Hf}-\mathbf{g})-\left(\frac{l}{2}+\frac{1}{c}\right)\mathbf{Qv}$$

where $\mathbf{v}$ is an $N\times 1$ vector, defined as:

$$\mathbf{v}(i) = \frac{c\left|[\mathbf{Qf}](i)\right|^{c-1}sign\left([\mathbf{Qf}](i)\right)}{m(l-2)+\left|[\mathbf{Qf}](i)\right|^{c}}.$$

See Appendix B for a detailed derivation of this result.

Using the closed form of the gradient, the maximum is obtained by application of the iterative scheme of a nonlinear Conjugate-Gradients algorithm, using the Secant method for the line search. This algorithm was preferred because it is designed to quadratic-like functions.

Of course the iterations take place in the DFT domain to reduce the computations, since all the matrices are circulant.

Now since we have an approximation (at the point $\mathbf{f}^*$), remains to find the values of the model parameters that maximize the likelihood. Tthe derivative with respect to l of the logarithm of the approximation is:

$$\frac{\partial J'(\mathbf{g},\mathbf{f}^*;\beta,m,l)}{\partial l} = G(l) + \frac{N}{2}\frac{\partial l \log(m(l-2))}{\partial l} - \frac{\partial\left(\left(\frac{l}{2}+\frac{1}{c}\right)\sum_{i=1}^{N} log\left(m(l-2)+\left|[\mathbf{Qf}^*](i)\right|^c\right)\right)}{\partial l}$$

$$= G(l) + \frac{N\left(\log(m(l-2)) + \frac{1}{(l-2)}\right)}{2} - \sum_{i=1}^{N}\left(\frac{log\left(m(l-2)+\left|[\mathbf{Qf}^*](i)\right|^c\right)}{2} - \frac{m\left(\frac{l}{2}+\frac{1}{c}\right)}{m(l-2)+\left|[\mathbf{Qf}^*](i)\right|^c}\right).$$

where: $$G(l) \triangleq \frac{d\left(\log\Gamma\left(\frac{l}{2}\right)^{-N}\Gamma\left(\frac{l}{2}+\frac{1}{c}\right)^{N}\right)}{dl}$$

The computation of the function $G(l)$ that represents the derivative of the logarithm of the two gamma functions is described in Appendix A. The derivative for the second parameter is simpler:

$$\frac{\partial J'(\mathbf{g},\mathbf{f}^*;\beta,m,l)}{\partial m} = \frac{N}{2m} - \left(\frac{l}{2}+\frac{1}{c}\right)\sum_{i=1}^{N}\frac{(l-2)}{\left(m(l-2)+\left|[\mathbf{Qf}^*](i)\right|^c\right)}.$$

The two gradients are used to the application of a semi-Newton optimization algorithm, based on the BFGS update formula for the Hessian matrix.

The parameter for the noise is estimated again from Eq. (4.6).

4.3 Experiments

We present numerical experiments to demonstrate our algorithms using 256×256 images. The images were blurred by Gaussian shaped blur and Gaussian noise of variance 10^{-2} was added to it and is shown in *Figure 4.2, 4.7* and *4.12*. The restored image using a stationary SAR prior and a MAP approach in [22] is shown in *Figure 4.3, 4.9, 4.14*. In *Figure 4.4, 4.9, 4.14,* we show the restored image using the MAP approach with a non stationary prior with c=2. In our numerical implementation we observed that the MAP approach could not estimate reliably all unknowns. In other words we could estimate simultaneously $\mathbf{f}, \boldsymbol{\alpha}, \beta, l,$ and m iterating between (4.5)-(4.13). This is a serious drawback of the MAP formulation for this model. Thus, we estimated β and m from a stationary SAR model and $l = 2.1$ was used which was empirically found to give good restorations. In contrast for the Bayesian approach all parameters can be computed automatically. To demonstrate the Bayesian approach and the effect of the selection of the exponent of the generalized Gaussian model we show restorations with c=2.0 and 0.8 in *Figures* 4.*(5, 10, 15)* and *4.(6, 11, 16)* respectively.

The error metric used to evaluate our results is the weighted MSE (WMSE) that takes into account the visibility of the errors [8, 26]. This metric is defined as $\text{WMSE} = (\mathbf{f} - \hat{\mathbf{f}})^T \mathbf{A} (\mathbf{f} - \hat{\mathbf{f}})$. The matrix $\mathbf{A} = diag(\lambda_1, \lambda_2 \ldots \lambda_N)$, where λ_i^{-1} the local

variance of the original image at location *i*. From the restored images shown in *Figures 4.3-4.16* it is clear that the non stationary model yields visually more pleasing results. Furthermore, the WMSE metric that incorporates the visibility of the error in the image is about 50% smaller for the non stationary model in all experiments. We can also observe the difference between the non stationary restored with the Bayesian algorithm shown in *Figures 4.(5, 6), 4.(10, 11)* and *4.(15,16)*. The larger is the generalized shape parameter *c*, the smoother the restored image seems to become.

The Conjugate-Gradients algorithm used for optimization, for the line search the Secant method was used. The line search terminated after a limited number of repetitions for two reasons. For c=2 the algorithm converged very fast, so the sooner it terminated the less time was needed for the restoration. For c=0.8, the convergence was much slower. But the main reason for terminating after a limited number of repetitions, and not using a criterion that takes into account the change in the likelihood, is that for the Gaussian case after a certain number of steps, the increasie in the likelihood did not imply a decrease in the WMSE. This behavior was not observed when c=0.8. In the experiments presented here we adopt the limited number repetitions for the Gaussian case, so the error metric appears better than the non-Gaussian case. For the non-Gaussian case the line search was terminated when convergence in the likelihood was detected. The "early stopping" technique was adopted also for the MAP case. Another very important result is the number of iterations. Two iterations of the optimization step were enough to restore satisfactorily the images. So the time required for the restoration was about 60' in a Pentium 4 2.4 GHz computer. Also the parameters l and m never exceeded their bounds (both positive and l greater from 2).

Figure 4.2: 'Lena' degraded image.

Figure 4.3: Stationary restoration (WMSE = 4.07e+010)

Figure 4.4: MAP non stationary restora c=2 (WMSE = 1.85e+010).

Figure 4.5: Bayesian non stationary restoration c=0.8 (WMSE= 2.67e+010)

Figure 4.6: Bayesian non static restoration c=2 (WMSE = 2.44e+010)

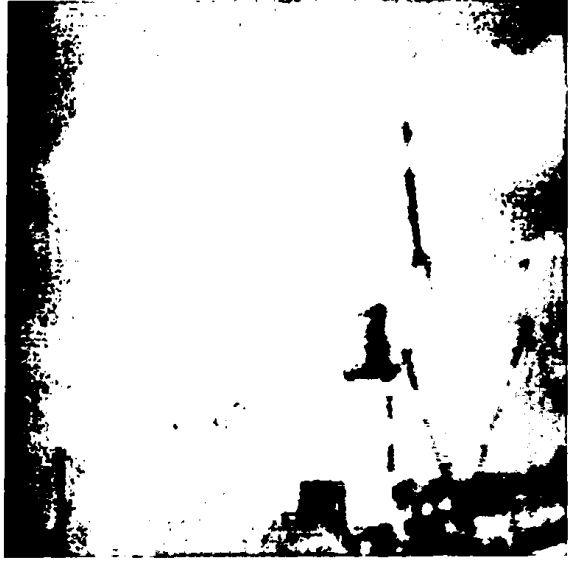

Figure 4.7: 'Boat' degraded image.

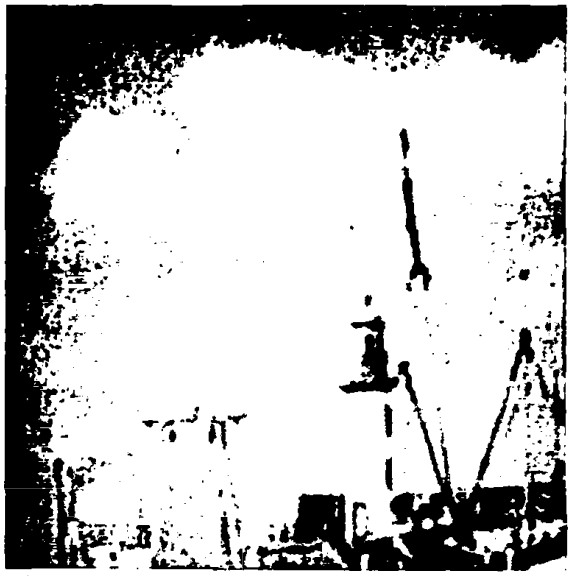

Figure 4.8: Stationary restoration (WMSE = 1.79e+010)

Figure 4.9: MAP non stationa restoration c=2 (WMSE = 1.06e+010).

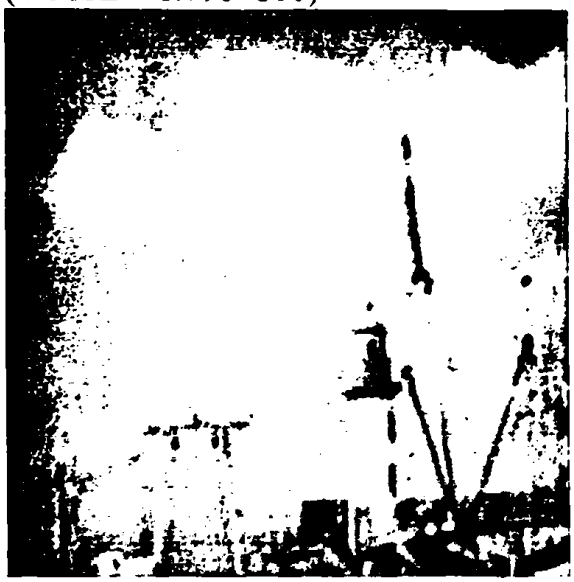

Figure 4.10: Bayesian non stationary restoration c=0.8 (WMSE= 1.27e+010)

Figure 4.11: Bayesian non stationa restoration c=2 (WMSE=1.63e+010)

Figure 4.12: Degraded image

Figure 4.13: Stationary restoration (WMSE =3.77e+010)

Figure 4.14: MAP non stationary restoration c=2 (WMSE =1.54e+010).

Figure 4.15: Bayesian non stationary restoration c=2 (WMSE=2.12e+010)

Figure 4.16: Bayesian non stationary restoration c=0.8 (WMSE=2.34e+010)

CHAPTER V

CONCLUSIONS AND FUTURE WORK

The non stationary prior image model that we propose in this thesis provides better restorations since it has the ability to smooth the uniform areas of the images while at the same time it maintains its edges. An interesting observation about the MAP approach proposed in this paper is that it yields update equations for variances of the residuals of the non stationary SAR model for c=2 which are *identical* in form to the equations proposed for obtaining the visibility weights of the noise in images [6, 7]. In other words, using the MAP formulation we were able to obtain the same form of equations as in [8, 26] which were derived using heuristic arguments.

Also, the MAP and the Gaussian non-stationary algorithms give smaller WMSE, but this is due to the early stopping tactic that we adopted for the experiments. The fast convergence in these cases can be explained if we notice that the linear Conjugate-Gradients algorithm is designed to optimize quadratic functions. Thus when the function is not quadratic, more time is needed for the algorithm to converge and thus to give a small WMSE. We further noticed that the increasing likelihood did not decrease the WMSE when c=2. This leads us to the conclusion that the Gaussian model is not appropriate for this images

In the future we plan to explore methodologies to estimate the best shape parameter c for the generalized Gaussian and priors that capture the direction of the edges also. A better approximation of the marginal distribution will be always a challenge. We also plan in the future, to use the variational methodology [31] for non stationary restoration.

APPENDIX A

Here we show the calculation of the logarithmic derivative of the two gamma functions (denoted by the term $G(l)$). The logarithmic derivative of the general function $\Gamma(x)$ is:

$$\frac{d\log\Gamma(x)}{dx} = -C - \frac{1}{x} + \sum_{n=1}^{+\infty}\left(\frac{1}{n} - \frac{1}{x+n}\right),$$

where C is a constant. Using this result the derivative we want to evaluate is:

$$G(l) \triangleq \frac{d\left(\log\Gamma\left(\frac{l}{2}\right)^{-N}\Gamma\left(\frac{l}{2}+\frac{1}{c}\right)^{N}\right)}{dl} = N\frac{d\log\Gamma\left(\frac{l}{2}+\frac{1}{c}\right)}{dx} - N\frac{d\log\Gamma\left(\frac{l}{2}\right)}{dx} =$$

$$= -N\frac{2}{l+\frac{2}{c}} - N\sum_{n=1}^{+\infty}\left(\frac{2}{l+\frac{2}{c}+2n}\right) + N\frac{2}{l} + N\sum_{n=1}^{+\infty}\left(\frac{2}{l+2n}\right)$$

For not large l, the approximation is quite satisfactory for summation to small n.

APPENDIX B

The maximum of the integrand is obtained taking the derivative of the logarithm:

$$\frac{\partial\left(\log p(\mathbf{g}\,|\,\mathbf{f};\beta)-\left(\frac{l}{2}+\frac{1}{c}\right)\log\sum_{i=1}^{N}\left(m(l-2)+\left\|\left[\mathbf{Qf}(i)\right]\right\|^{c}\right)\right)}{\partial\mathbf{f}}=$$

$$=\frac{\partial\log p(\mathbf{g}\,|\,\mathbf{f};\beta)}{\partial\mathbf{f}}-\left(\frac{l}{2}+\frac{1}{c}\right)\frac{\partial\log\sum_{i=1}^{N}\left(m(l-2)+\left\|\left[\mathbf{Qf}(i)\right]\right\|^{c}\right)}{\partial\mathbf{f}}$$

Evaluation of the first derivative is straightforward:

$$\frac{\partial\log p(\mathbf{g}\,|\,\mathbf{f};\beta)}{\partial\mathbf{f}}=-\frac{1}{2}\frac{\partial\left(\beta\left\|\mathbf{Hf}-\mathbf{g}\right\|^{2}\right)}{\partial\mathbf{f}}=-\beta\mathbf{H}^{\mathrm{T}}\left(\mathbf{Hf}-\mathbf{g}\right)\qquad\text{B.1}$$

The second needs more analysis. Derivation with respect to an $\mathbf{f}(i)$, which is the $i-th$ point of the vector $\mathbf{f}$, yields:

$$\frac{\partial\sum_{i=1}^{N}\log\left(m(l-2)+\left\|\left[\mathbf{Qf}(i)\right]\right\|^{c}\right)}{\partial\mathbf{f}(i)}=\qquad\text{B.2}$$

$$=\frac{-2c\left\|\left[\mathbf{Qf}(i-1)\right]\right\|^{c-1}sign\left(\mathbf{Qf}(i-1)\right)}{\left(m(l-2)+\left\|\left[\mathbf{Qf}(i-1)\right]\right\|^{c}\right)}+\frac{c\left\|\left[\mathbf{Qf}(i)\right]\right\|^{c-1}sign\left(\mathbf{Qf}(i)\right)}{\left(m(l-2)+\left\|\left[\mathbf{Qf}(i)\right]\right\|^{c}\right)}+\frac{c\left\|\left[\mathbf{Qf}(i+1)\right]\right\|^{c-1}sign\left(\mathbf{Qf}(i+1)\right)}{\left(m(l-2)+\left\|\left[\mathbf{Qf}(i+1)\right]\right\|^{c}\right)}$$

because:

$$\frac{\partial\left\|\left[\mathbf{Qf}(k)\right]\right\|^{c}}{\partial\mathbf{f}(i)}=sign\left(\mathbf{Qf}(k)\right)\times\begin{cases}0, & if\quad i\neq k,k-1,k+1\\ -2c\left\|\left[\mathbf{Qf}(k)\right]\right\|^{c-1}, & if\quad i=k\\ c\left\|\left[\mathbf{Qf}(k)\right]\right\|^{c-1}, & if\quad i=k\pm1\end{cases}$$

Of course, the function $sign(x)$ returns the sign of the real number x.

Equation B.2 shows that the derivative with respect to $\mathbf{f}$ is a vector with elements the quantity in B.2 at each position i. If we consider the $N \times 1$ vector $\mathbf{v}$ where each element i is equal to:

$$\mathbf{v}(i) = \frac{c\left|\left[\mathbf{Qf}(i)\right]\right|^{c-1} sign\left(\mathbf{Qf}(i-1)\right)}{\left(m(l-2)+\left|\left[\mathbf{Qf}(i)\right]\right|^{c}\right)}, \qquad i = 1,2,...,N$$

then the above derivative is obtained by application to vector $\mathbf{v}$ the Laplacian operator, as B.2 shows:

$$\frac{\partial \sum_{i=1}^{N} \log\left(m(l-2)+\left|\left[\mathbf{Qf}(i)\right]\right|^{c}\right)}{\partial \mathbf{f}} = \mathbf{Qv}.$$

Each element of the $N \times 1$ vector is:

$$\mathbf{Qv}(i) = -2\mathbf{v}(i) + \mathbf{v}(i-1) + \mathbf{v}(i+1),$$

and it is equal to the derivative in Equation B.2.

REFERENCES

[1] M.Z. Nashed, "Operator Theoretic and Computational Approaches to Ill-Posed Problems with Applications to Antenna Theory," IEEE Trans. Antennas Prop., vol. 29, pp. 220-231, March 1981.

[2] A.N. Tikhonov and V.Y Arsenin, *Solutions of Ill- Posed Problems*, New York: Wiley, 1977.

[3] H.C. Andrews and B.R. Hunt, *Digital Image Restoration*, Englewood Cliffs, NJ: Prentice Hall, 1997.

[4] B.R. Hunt, "The Application of Constrained Least Squares to Image Restoration by Digital Computer", IEEE Trans. Computers, vol. 22, pp. 805-812, September 1973.

[5] A.K. Katsaggelos, *Constrained Iterative Image Restoration Algorithms*, Ph.D. Thesis, Georgia Institute of Technology, August 1985.

[6] A.M. Tekalp, H. Kaufman, and J.W. Woods, "Edge Adaptive Image Filtering for Image Restoration with Ringing Suppression", IEEE Trans. Acoust., Speech, and Signal Proc., vol. 37, no. 6, pp. 892-899, June 1989.

[7] N.P. Galatsanos and A.K. Katsaggelos, "Methods for Choosing the Regularization Parameter and Estimating the Noise Variance in Image Restoration", IEEE Trans. Image Proc., vol 1., pp. 322-336, July 1992.

[8] G. L. Anderson and A. N. Netravali, "Image restoration based on a subjective criterion", *IEEE Trans. Syst., Man, Cybern*, vol. SMC-6, pp. 845-853, Dec. 1976.

[9] A.K. Katsaggelos, J. Biemond, R.W. Schafer, and A.R. Mersereau, "A regularized Iterative Image Restoration Algorithm", IEEE Trans. Acoust., Speech, Signal Proc., vol. 39, pp. 914-929, April 1991.

[10] D. M. Titterington, "General structure of regularization procedures in image reconstruction," *Astron. Astrophys.*, vol. 144, pp. 381-387, 1987.

[11] A.K. Katsaggelos and M.G. Kang, "A Spatially Adaptive Iterative Algorithm for Restoration of Astronomical Images", invited paper, *International Journal of Imaging Systems and Technology*, special issue on "Image Restoration and Reconstruction in Astronomy", vol. 6, pp. 305-313, Winter 1995.

[12] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs distribution, and the Bayesian Restoration of Images", *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 6, pp. 228-238, November1984.

[13] A. Blake and A. Zisserman, *Visual Reconstruction*, Cambridge: MIT Press, 1987.

[14] S. Geman and D. McClure, "Bayesian Image Analysis: An application to single photon emission tomography", in 1985 *Proc. Stat. Comput. Sect., Amer. Stat. Assoc.*, 1985.

[15] D. Geman and C. Yang, "Non-linear Image Recovery with Half-Quadratic Regularization and FFT's", IEEE Trans. Image Processing, vol.4, pp. 932-946, July 1995.

[16] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and Michel Barlaud, "Deterministic Edge-Preserving Regularization in Computed Imaging", IEEE Trans. Image Proc., vol. 6, pp. 298-311, February 1997.

[17] J. Idier, "Convex Half-Quadratic Criteria and Interacting Auxiliary Variables for Image Restoration", IEEE Trans. Image Proc., vol. 10, pp. 1001-1009, no. 7, July 2001.

[18] M.R. Banham, N.P. Galatsanos, H.L. Gonzales, and A.K. Katsaggelos, "Multichannel Restoration of Single Channel Images using a Wavelet-Based Subband Decomposition", IEEE Trans. Image Proc., vol. 5, pp. 821-833, November 1994.

[19] M.R. Banham and A.K. Katsaggelos, "Spatially-Adaptive Wavelet-Based Multiscale Image Restoration", IEEE Trans. Image Proc., vol. 5, pp. 619-634, April 1996.

[20] G. Archer and D. M. Titterington, "On Some Bayesian/Regularization Methods for Image Restoration", IEEE Tran. Image Proc., vol.4 , no. 7, July 1995.

[21] R. Molina, and B. D. Ripley, "Using spatial models as priors in astronomical images analysis", J. Appl. Stat., vol.16, pp.193-206, 1989.

[22] Molina, R., Katsaggelos, A. K., Mateos, J., "Bayesian and regularization methods for hyper-parameter estimation in image restoration," IEEE Trans. on Image Processing, Vol: 8 No: 2 , pp. 231 -246, Feb. 1999.

[23] Galatsanos N. P., Mesarovic V. N., Molina R. M. and Katsaggelos A. K., "Hierarchical Bayesian Image Restoration from Partially-Known Blurs," IEEE Trans. on Image Processing, Vol. 9, No. 10, pp. 1784-1797, October 2000.

[24] Ruanaidh J., and Fitzegerald W., Numerical Bayesian Methods Applied to Signal Processing, Springer Verlag, 1996.

[25] Galatsanos N. P., Mesarovic V. N., Molina R. M. and Katsaggelos A. K., "Hierarchical Bayesian Image Restoration from Partially-Known Blurs," IEEE Trans. on Image Processing, Vol. 9, No. 10, pp. 1784-1797, October 2000.

[26] Katsaggelos A. K. "Iterative Image restoration", in Handbook on Image and Video Processing, Editor Al. Bovik, pp. 191-206, Academic Press 2000.

[27] Tipping M. E. "Sparse Bayesian Learning And The Relevance Vector Machine" Journal Of Machine Learning Research 1, 211-244, 2001.

[28] Galatsanos N., V. N. Mesarovic, R. M. Molina, J. Mateos, and A. K. Katsaggelos, "Hyper-parameter Estimation Using Gamma Hyper-priors in Image

Restoration from Partially-Known Blurs," Optical Engineering, 41(8), pp. 1845-1854, August 2002.

[29] D. J. MacKay, Information Theory, Inference and Learning Algorithms, Cambridge University Press, 2003.

[30] D. J. MacKay, "Bayesian Methods for Backpropagation Networks", In E. Domany, J. L. Van Hammen and K. Schulten, editors, Models of Neural Networks III, chapter 6, pp. 211-254, Springer Verlag 1994.

[31] Beal, M.J., "Variational Algorithms for Approximate Bayesian Inference", PhD. Thesis, Gatsby Computational Neuroscience Unit, University College London, 2003.

[32] Schultz, R.R.; Stevenson, R.L., "A Bayesian approach to image expansion for improved definition", IEEE Transactions on Image Processing, Vol. 3 , No. 3 , pp. 233 – 242, May 1994.

[33] Bouman, C.; Sauer, K., "A generalized Gaussian image model for edge-preserving MAP estimation", IEEE Transactions on Image Processing, Vol. 2 , No. 3 , pp. 296 – 310, July 1993.

[34] http://www.xycoon.com, Statistics - Econometrics – Forecasting

[35] C. Bishop, "Latent Variable Models", In M. I. Jordan (ed.) Learning in Graphical Models, pp. 371-430. Cambridge, MA: MIT Press 1998.

[36] A. D. Dempster, N. M. Laird, and D. B.Rubin, "Maximum Likelihood from Incomplete Data via the EM algorithm", J. Roy. Stat. Soc., vol. B39, pp. 1-37, 1997.

[37] R. M. Neal, and G. E. Hinton, "A View of the E-M Algorithm that Justifies Incremental, Sparse and Other Variants", In M. I. Jordan (ed.) Learning in Graphical Models, pp. 355-368. Cambridge, MA: MIT Press 1998.

[38] B. Charlin, and T. Louis, Bayes and Empirical Bayes Methods for Data Analysis, CRC Press; 2nd edition, 2000.