Title
Recovering Depth from Focus Using Iterative Image Estimation Techniques

Authors
Vitria, J.
Llacer, J.

Publication Date
1993-09-01

LBL-35158
UC-405

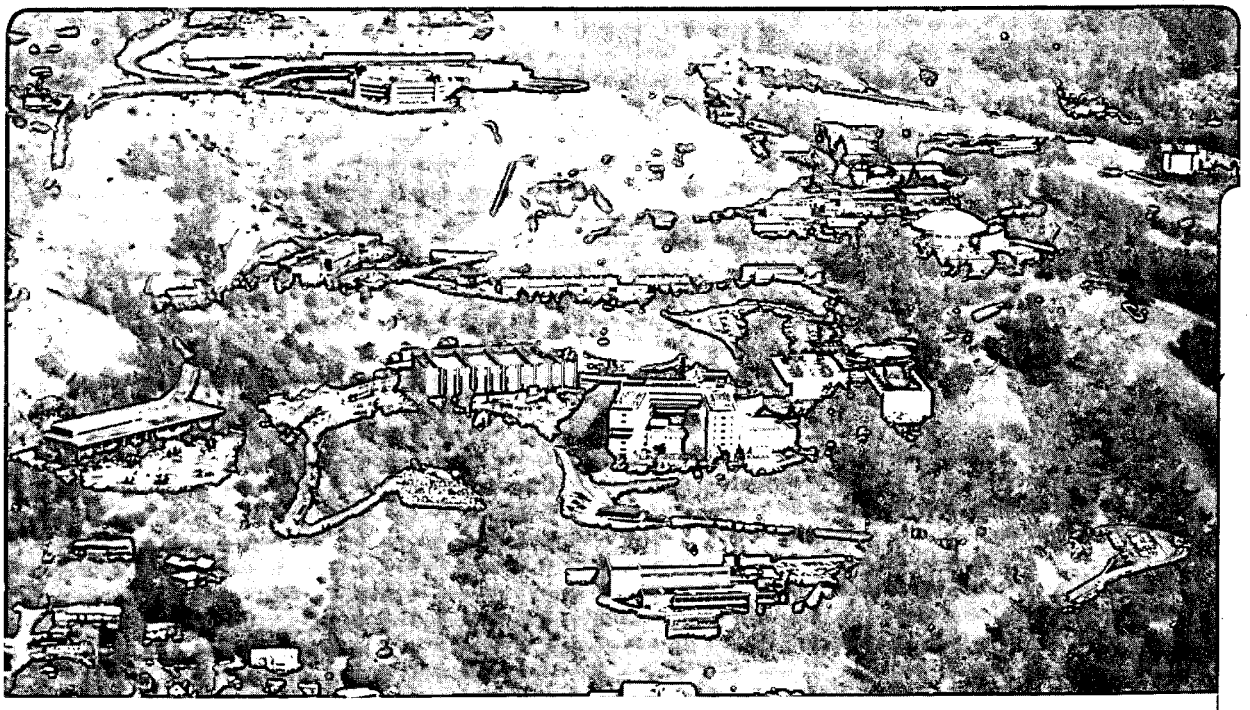
Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

Engineering Division

Recovering Depth from Focus Using Iterative Image Estimation Techniques

J. Vitrià and J. Llacer

September 1993



Prepared for the U.S. Department of Energy under Contract Number DE-AC03-76SF00098

DISCLAIMER

Technical Report
September 1993

LBL-35158
UC-405

Recovering Depth From Focus Using Iterative Image Estimation Techniques

Jordi Vitrià[†], Jorge Llacer
Lawrence Berkeley Laboratory
University of California, Berkeley, CA
[†] *on leave from Computer Science Dept.*
Universitat Autònoma de Barcelona
Bellaterra, Barcelona, Spain

Abstract

In this report we examine the possibility of using linear and nonlinear image estimation techniques to build a depth map of a three dimensional scene from a sequence of partially focused images. In particular, the techniques proposed to solve the problem of construction of a depth map are: 1) linear methods based on regularization procedures and 2) non linear methods based on statistical modelling. In the first case, we have implemented a matrix-oriented method to recover the *point spread function* (PSF) of a sequence of partially defocused images. In the second case, the chosen method has been a procedure based on image estimation by means of the EM algorithm, a well known technique in image reconstruction in medical applications. This method has been generalized to deal with optically defocused image sequences.

This report describes research done at the *Lawrence Berkeley Laboratory* of the University of California, Berkeley, and has been supported, in part, by a grant from the Generalitat de Catalunya *(Gaspar de Portola Catalonian Studies Program, University of California, Berkeley)* and in part by the Director, Office of Energy Research, Office of Health and Environmental Research, Physical and Technological Division, of the U.S. Department of Energy under contract no. DE-AC03-76SF00098.

1 Introduction

Computer Vision [27] is already a well defined scientific discipline, with a clear objective: the construction of systems that perform certain (intelligent) tasks based on visual sensing and feedback. The first processing stage in computer vision, also called *early vision*, consists in transforming two dimensional images to instances of a model representing 3D surfaces and properties. *Early vision* includes problems such as the recovery of motion or optical flow, shape from shading, etc. In the past two decades, several algorithms and sensing techniques have been proposed to recover properties of the physical world from images, but existing recovering methods suffer from several limitations. These limitations arise from the inability of these methods to deal with complex physical phenomena such as occlusions, shadows or interreflections. Physical based vision is an idea that has found widespread acceptance to solve this problem. One style of physics based processing that has recently emerged is the "inverse optics" approach. In this report we propose some methods to extract depth and shape information from a specific source: focal changes resulting from the limited depth of field inherent in most optical systems. There is a direct relationship between the amount of defocus or blurring in the image of an object point and the distance to the plane of exact focus from the object point . So, knowing this amount, it is straightforward to determine the distance between the object and the lens.

Using focus information in visual tasks is a fact in biological vision. There is some physiological evidence which shows that this depth cue could be used by the human visual system at least in two ways [24]:

- The depth of field for the red green retinal cells is different from that of the blue retinal cells, because of one diopter of chromatic aberration of the "lens". This provides two simultaneous views of the scene with dissimilar depth of field, albeit in different spectral bands.
- The focal length of the human eye is constantly varying in sinusoidal fashion at a frequency of about 2Hz.

From the computational point of view, depth from focus has two important features which makes it attractive as a depth cue with respect to classical techniques:

- It avoids the correspondence problem, which is the most difficult problem in stereo-based methods.
- It avoids the occlusion problem, which is the most difficult problem in motion-based methods.

Moreover, this method can be used to mutually reinforce other classical methods, as has been shown by Krotkov in [7], where a stereo system cooperates with a focus system. Our main objective when using this method can be defined as:

"Given a scene with objects placed at an unknown distance from the camera, derive the distance to visible points of the scene objects using information related to the focus."

Finally, and in order to build a practical system which uses this type of methods, some important points concerning this source of depth information must be stated:

1. Camera calibration and image formation model.
   (a) Point spread function (PSF) determination. Some PSF's are known in advance, but some must be determined experimentally.
   (b) Perspective equation. In some cases we can assume a parallel projection, but in "real world" scenes, we must consider perspective projection.
   (c) Relationship between the camera control parameters (focal length, magnification, spectral band, shutter time, sensor gain and camera position and orientation) and the parameters of the resulting images.
2. Focus processing.
   (a) How many images?. Actually, this number depends on the method to be used and on the required accuracy.
   (b) Which technique? Several techniques are available, based on different models: Fourier-based techniques, estimation techniques, etc.

(c) "A priori" information. Which kind of information can we provide "a priori"? How does this affect the choice of method?

3. Practical considerations.

   (a) Accuracy. There are applications where a high level of accuracy is not necessary (obstacle avoidance for robot navigation, for example) while in others is essential (3D VLSI circuit inspection, for example).

   (b) Robustness. How noise and non desired effects (intereflections, for example) affect the method?

   (c) Range of applicability. The applications for these methods range from microscopy to "real world" scenes.

In [48] a possible model for using computer-controlled cameras and lenses is described, and similar problems to the ones discussed above are presented.

2 Previous work.

In order to recover depth information from focus blur information, two approaches have been described in the literature: *depth from focus* and *depth from defocus*. The *depth from focus* technique uses a search for the sharpest focus position over a sequence of images taken at different lens focus settings or camera positions. The *depth from defocus* attempts to model the blurring process in a local image region as a function of depth. The goal is to determine surface depth by estimating the model parameters on the image. Pentland [25, 23, 24] and Grossman [20] both addressed the problem of recovering depth from blurred edges. Their work can be classified as *depth from defocus*. Also Subbarao has developed some methods of this type [18] [19]. Krotkov [7] and recent work of Subbarao [21] are examples of *depth from focus*.

Pentland proposes two methods [23, 24]. The first is based on measuring the blur of edges which are step discontinuities in the focused image. His second method is based on comparing two images, one formed with a very small (pin-hole) aperture and the other one formed with a normal aperture. In both cases, the methods are based on inverse filtering. In [26] an extension of Pentland's work is presented.

In [16, 17] a matrix method based on deconvolving the defocus operator from the images is presented. This is done by characterizing the problem as a system of linear equations. The main characteristics of the method are the independence of defocus operator models and least constraints on the scene. They argue against inverse filtering-based techniques, and give some interesting reasons.

The matrix-based approach to depth from defocus [17]

Two images $i_1(x,y))$ and $i_2(x,y)$ of the same scene $s(x,y)$ are acquired with two different defocus operators $h_1(x,y)$ and $h_2(x,y)$. A convolution ratio $h_3(x,y)$ is presented, where

$$h_1(x,y) \otimes h_3(x,y) = h_2(x,y)$$

This convolution ratio is a unique indicator of depth. Three methods of recovering $h_3(x,y)$ are presented in [17]. Assuming no noise, the first method solves for $h_3(x,y)$ by

$$\mathbf{h}_{3S} = [\mathbf{i}_{1BT}]^{-1}.\mathbf{i}_{2S}$$

where $[\mathbf{i}_{1BT}]$ is an $N^2 \times N^2$ block Toeplitz matrix constructed from $i_1(x,y)$ and $\mathbf{h}_{3S}$, $\mathbf{i}_{2S}$ are row-stacked vectors created from $h_3(x,y)$ and $i_2(x,y)$.
The second method uses regularization to deal with the noise, and the solution for $h_3(x,y)$ is

$$\mathbf{h}_{3S} = ([\mathbf{i}_{1BT}]^T[\mathbf{i}_{1BT}] + \lambda[C]^T[C])^{-1}[\mathbf{i}_{1BT}]^T\mathbf{i}_{2S}$$

where λ is a regularization parameter and C is a matrix minimizing the magnitude of the second term if $\mathbf{h}_{3S}$ belongs to a parametrc family of patterns. This method requires that $h_1(x,y)$ and $h_2(x,y)$ be represented by parametric families.
The third method is based in looking for a $\hat{h}_3(x,y)$ that minimizes

$$\sum_{x=0}^{N-k}\sum_{y=0}^{N-k}[i_1(x,y)[\otimes]\hat{h}_3(x,y) - i_2(x,y)]^2$$

where $[\otimes]$ is the restricted convolution.

Subbarao [19] proposes a method based on measuring the change in an image due to a very small change in one of the camera parameters, and in [18] a method for recovering depth from a measure on the blur of an edge.

In [76] the problem of depth from focus is formulated as an image restoration problem. Linear restoration techniques (constrained optimization) are used to *restore* a depth map of the object.

In [35] a depth from focus system is presented which is based on a *focus criterion function* which quantifies the notion of defocus. The position of sharpest focus is then given by the focus setting corresponding to the global maximum or minimum of the criterion function.

Bove [61] has recently proposed a method to solve the two-image depth from defocus problem (in the same sense as Pentland) in terms of entropy loss in the defocused image.

All these methods can be grouped in two broad classes: those based on inverse filtering (in Fourier domain), and those based on matrix techniques. While the first class presents a high performace in terms of speed and storage, the second class is able to give more accurate results. We do not know of any reference proposing statistical modelling to solve this problem.

The inverse filtering approach to depth from defocus [24]

The method proposed by Pentland is based on taking two images of a single view of a scene, with different lens aperture. We can consider a patch $f_1(r,\theta)$ [1] centered at (x_0, y_0) within the first image $i_1(x,y)$:

$$f_1(r,\theta) = i_1(x_0 + rcos\theta, y_0 + rsin\theta)$$

and calculate its two dimensional Fourier transform $\mathcal{F}_1(t,\theta)$. The same is done for a patch $f_2(r,\theta)$ at the corresponding point of the second image, yielding $\mathcal{F}_2(t,\theta)$. Now we consider

$$\frac{f_1(r,\theta)}{f_2(r,\theta)} = \frac{f_0(r,\theta) \otimes G(r,\sigma_1)}{f_0(r,\theta) \otimes G(r,\sigma_2)}$$

where $G(r,\sigma)$ is a gaussian approximation to the PSF. As is shown in [24], we can derive the following from the previous relation

$$\frac{\mathcal{F}_1(\lambda)}{\mathcal{F}_2(\lambda)} = \frac{G(\lambda, \frac{1}{\sqrt{2\pi}\sigma_1})}{G(\lambda, \frac{1}{\sqrt{2\pi}\sigma_2})} = e^{(\lambda^2 2\pi^2(\sigma_2^2 - \sigma_1^2))} \quad (1)$$

where $\mathcal{F}_i(\lambda) = \int_{-\pi}^{\pi} \mathcal{F}_i(\lambda,\theta)d\theta$. Thus, given $\mathcal{F}_1$ and $\mathcal{F}_2$ we can find σ_1 and σ_2 as follows. Taking the natural log of (1) we obtain

$$\lambda^2 2\pi^2(\sigma_2^2 - \sigma_1^2) = ln\mathcal{F}_1(\lambda) - ln\mathcal{F}_2(\lambda)$$

and we may formulate it as a linear regression equation in λ^2.
If $\sigma_1 = 0$ (a pinhole camera), then we have $2\pi^2(\sigma_2^2 - \sigma_1^2) = 2\pi^2\sigma_2^2$, and we can solve directly the depth from defocus problem.

3 General formulation

When an object is seen through an optical system, only its *visible surface* can be imaged. This surface is defined as all the points of the object surface that are visible from the viewer position. In the simplest form, we can model the optical system as a pinhole camera, and then every object is imaged without any loss of sharpness in the image plane. A real optical system can only image objects at a certain distance from the camera. To deal with this fact, it is necessary to model more accurately the optical system. We can assume also, that this later model doesn't change the imaging geometry. The range of distances within the objects that are imaged "sharply" is called *depth of focus.* The *depth of focus* is illustrated in figure 3.

The basic law governing image formation through a lens can be described by the lens formula used in geometrical optics:

$$1/d + 1/e = 1/f$$

where d is the object distance, e is the distance between the lens and the focused image plane of the object, and f is the focal length of the lens (all measured along the optical axis). When the distance between the object point and the lens is d, the point is imaged as a point. In other cases the image of a point smears to a disc with radius ϵ at the original image plane. The radius of this disc is directly related to the distance between the object point and the lens, and then it can be used to determine a "depth map" [2] of the imaged object.

Previously it has been seen that a point in 3D object space is not imaged onto a point in image space but onto a more or less extended area with varying intensities. Obiously, the function which describes the imaging of a point is an essential feature of the imaging system which is called the point spread function, abbreviated as PSF. We assume that the PSF is not position-dependent. Then the system is position invariant [5].

If we know the PSF we can calculate how any arbitrary 3D object would be imaged by this system. The intensity of a point (x,y) at the image plane is computed by integrating the contributions from all the other points previously transformed by their corresponding point spread functions:

[2]**Depth map**: *A depth map is a two dimensional function which gives the depth of an object point, d(x,y), -relative to the reference plane- as a function of the image coordinates.*

$$g(x,y,z) = \int h(x-x', y-y', z-z') f(x', y', z') dx' dy' dz'$$

We denote the point spread function $h(x,y,z)$ and consider the overall image of any object to be a superposition of appropiately shifted and weighted versions of $h(x,y,z)$.

During the process of image formation, the 3D coordinates of the visible surface are mapped onto the 2D coordinates of its image. The "depth from focus/defocus" problem can be stated as recovering the 3D coordinates of the visible surface.

3.1 Geometric Optics: a theory for defocus

Geometric optics is useful as a first approximation of light collection and image formation by almost any kind of optical system [13], and is a good aproximation to explain the theory of defocusing [17]. From fig. 3, we can deduce the formula for the radius R of a blured circle:

$$R = \frac{L\delta}{2e}$$

where L is the diameter of the lens or the aperture, e is the distance from the lens to a sharply focused image of a particular object, and δ is the displacent of the image plane from sharp focus.

To a first-order approximation, the brightness within the blur circle is uniform. The defocus operator, $h(x,y)$, where x and y are coordinates in the image plane, is defined as a pillbox of radius R

$$h(i,j) = \begin{cases} \frac{1}{\pi R^2} & if \quad x^2 + y^2 \leq R^2 \\ 0 & if \quad x^2 + y^2 > R^2 \end{cases}$$

Blurring due to defocus can be modelled as a convolution with this pillbox. Therefore, the amount of defocus or blurring depends solely on the distance to the surface of exact focus and the characteristics of the lens system [3]. The defocus operator $h(x,y)$ can be changed by varying one or more camera parameters, i.e position of image plane e_i, focal length f or aperture L. [4]

[3] We consider that we have a calibrated system and we know these characteristics

[4] In a microscope the image plane is constant and the object is displaced along the optical axis.

4 Linear Methods

As we have seen, in the simplest case, when we have a uniformly defocused image the basic equation is:

$$g(x,y) = \sum_{x'y'} h(x,x',y,y')f(x',y')$$

where f is the original image, h the point spread function and g the defocused image.

First, we considered some direct methods to restore the defocused image in the case of a known PSF. Let the 2D array $g(x,y)$ represent a discrete image with M rows and N columns. This data can be written in vector form through a 1D mapping that is known as lexicographic ordering. In essence, the 2D array is converted to a $M \times N$ vector by concatenating its rows:

$$f = [g_{(0,0)}, g_{(0,1)}, g_{(0,2)}, \cdots g_{(0,N-1)}, g_{(1,0)}, \cdots g_{(M-1,N-1)}]^t$$

If the image formation model is given by $g(i,j) = \sum h(i-m, j-n)f(m,n)$, by lexicographic ordering of f and g we have:

$$g = Df \tag{2}$$

where D is the blurring matrix of size $MN \times MN$. If the convolution is interpreted as a linear convolution (with zero boundaries) the matrix D has a block-Toeplitz structure. A (block-) Toeplitz matrix is often approximated by a (block-)circulant one because this two matrix types are structurally closely related, and operations involving (block-)circulant matrices can be efficiently performed in Fourier domain.

The simplest approach to solve the restoration process is to use a filter whose response is the inverse of D:

$$f' = D^{-1}g$$

Unfortunately , there are several problems with this solution of (2):

- The size of the system -typically each matrix dimension is in the order of thousands.

- Ill conditioning of the system. Typically D has a large condition number, so that a small change in the projection data g may cause a large change in the solution f.
- Ill-posedness. Typically, D may have a non-empty null space, so that some components of the image can not be recovered from the projection g without additional information.

One practical method to solve this problem is seeking to minimize the difference between the projection into data space of f' and the measurent g, i.e., minimizing the target function [6]:

$$W(f') = |g - Df|$$

Taking partial derivatives with respect f', setting equal to zero and solving for f' we obtain:

$$f' = (D^t D)^{-1} D^t g$$

which is called the **least square inverse filter.**

Another interesting technique that can be used in this problem is the *Singular Value Decomposition (SVD).* This technique permits a diagnosis of where the difficulty lies in trying to find a solution, and in some cases, overcomes those difficulties [2].

SVD methods are based on the following theorem of linear algebra:

Theorem 1 *Any $M \times N$ matrix A whose number of rows M is greater than or equal to its number of columns N, can be witten as a product of an $M \times N$ column-ortogonal matrix U, and $N \times N$ diagonal matrix W, with positive or zero elements (the singular values), and the transpose of a $N \times N$ othogonal matrix V.*

$$A = UWV^t$$

In our case, we can restrict its application to square matrices. Then, U, V, and W are all square matrices of the same size. Their inverses are also trivial to compute: U and V are ortogonal, so their inverses are equal to their transposes: W is diagonal, so its inverse is a diagonal whose elements are the reciprocals of the elements w_j:

$$A^{-1} = V[diag(1/w_j)]U^t$$

Given $A^{-1} = V[diag(1/w_j)]U^t$, if A is singular, we can select one particular member of the set of solutions. We can choose the one with minimum length by computing:

$$f' = V[diag(1/w_j)](U^t g)$$

If g is not in the range of the singular matrix A, then $g = Df$ has no solution, but the former equation can be used to "construct" a solution f', which is the minimum in the least squares sense.

Finally, we consider another way of solving our problem: **regularization** [28, 30, 32] Tikhonov and Arsenin were the first to study the concept of regularization. The idea is to define a criterion to select an approximate solution from a set of feasible solutions. Tikhonov defined the regularized solution as the one which maximizes a stabilizing functional O on the set of feasible solutions. Although a wide class of stabilizing functionals are available, usually a stabilizing functional of the following form is chosen to facilitate the mathematical analysis of the problem:

$$O(f') = ||Cf'||$$

Here C is a real valued matrix of size $NN \times NN$, known as the regularized operator. The computation of the regularized solution reduces to the minimization of this objective function:

$$\Phi(f') = ||g - Df'||^2 + \alpha||Cf'||^2$$

with respect to f'. Another related approach is to replace the minimization of $O(f')$ by the following bound:

$$O(f') = ||Cf'|| \leq E$$

Among the solutions satisfying this approach, a reasonable choice is the Tikhonov-Miller regularized solution, which leads to the following solution:

$$f' = (D^t D + \alpha C^t C)^{-1} D^t g$$

The main disavantage of these methods is their storage requirements. They need to manipulate and store squared matrices of $N \times N$ rows and $N \times N$ columns. Evidently, this is not possible using current workstations, and we must restrict ourselves to small images in order to test the effectiveness of these methods.

A common approach to overcome the storage problem is to use recursive algorithms to compute a feasible solution.

The first experiment we did was the estimation of the point spread function of a defocused image. The problem was defined as:

$$g = Af$$

where g is a defocused image in vector-form, A is the original (focused) image in block-circulant matrix form, and f is the point spread function (unknown) in vector form.

We tried to estimate this function from simple synthetic images and from real images (defocused by convolution).

From a computational point of view, the methods have been implemented in double precision on images of 32x32, and in some cases, of 64x64 or 128x128. The conclusions reached from these experiments are:

1. The errors induced in the image boundaries due to the block circulant approximation were quite visible. For this reason, we worked with block-Toeplitz matrices.

2. In the case of synthetic images, linear methods work quite well, but they require huge amounts of storage and processing.

3. In the case of synthetically defocused real images, the major problem arose from that fact that those methods do not obey a positivity constraint. Then, some negative values were present in the estimated PSF.

4.1 Recursive methods

Some linear methods of image restoration can be written as recursive algorithms [56, 3]. All these methods are derived from the classical equation:

$$g_m = \sum_k a_{mk} f_k + n_k$$

In most of these algorithms noise is ignored. The algorithms are expressed as recursive formulas using the following standard technique [3]. First the equations are written in the form

$$f_m = KF(\{f_m\})$$

for some function F where K is a constant chosen so that conservation of energy is satisfied. Then the equation is transformed into a recurrence relation:

$$f_m^{(n+1)} = KF(\{f_m^{(n)}\})$$

The convergence properties of this formula will depend on how F is constructed.

Algorithms can be constructed from any of the following principles:

1. Let X_m be the noise-free image $X_m = \sum_k a_{mk} f_k$. Subtract X_m from both sides of the equation (ignoring the noise) and raise both sides to the power p, (any real number), giving

$$0 = (g_m - X_m)^P \tag{3}$$

2. Divide both sides of the equation by X_m and then raise both sides to the power of p, giving:

$$1 = (g_m/X_m)^p \tag{4}$$

3. Divide both sides by X_m and then subtract 1 from both sides and raise to the power p, giving,

$$0 = (g_m/X_m - 1)^p \tag{5}$$

It must be pointed out that X_m is function of f_m, and, it must be updated at each iteration.

There are several ways of designing new algorithms from these expressions. The first one is found by adding λ_m to both sides of equation (3) and scaling by K:

$$f_m^{(n+1)} = K[f_m^{(n)} + (g_m - X_m^{(n)})^p]$$

which for $p = 1$ is the *Van Citter* [1] recursive formula.

Another algorithm, from equation (4) can be found by multiplying both sides by f_m, giving the quasi inverse matrix formula:

$$f_m^{(n+1)} = K f_m^{(n)} (g_m / X_m^{(n)})^p$$

One important point about this formula is that it obeys positivity for optical objects. These techniques can be also applied to equations that consider the statistical model for the noise. In the case of Poisson noise:

$$p(\{g_m\}|\{f_m\}) = \prod_m (X_m g_m) \frac{exp(-X_m)}{g_m!}$$

and in the case of Gaussian noise:

$$p(\{g_m\}|\{f_m\}) = C \prod exp[\frac{-(g_m - X_m)^2}{2\sigma_m^2}]$$

The algorithms deduced from these formulas are, in the first case:

$$f_m^{(n+1)} = K f_m^{(n)} [\sum_k a_{mk} (\frac{g_k}{X_k^{(n)}})]^p \tag{6}$$

and in the second case:

$$f_m^{(n+1)} = K f_m^{(n)} [\sum_k a_{mk} (\frac{g_k}{X_k^{(n)}})^2] \tag{7}$$

The performance of these algorithms is not well characterized, and its applicability to a specific problem depends on testing of their behaviour on controlled experiments. In our case, they all give usable restorations, but non-linear methods yield better answers when noise is present in the image. The main advantage over preceding linear methods is that the formulation of the equations permits an easy implementation (little storage requirements), and also permits interactive display of the results as it converges.

We can find other iterative methods for solving this problem in the literature. Among those, we could highlight the *Landweber* and *Generalized*

Landweber iterations, which has been used in image restoration. The *Landweber* iteration method was proposed in 1951 [66]. This iteration is:

$$f^{(n+1)} = f^{(n)} + A^t(g - Af^{(n)})$$

This iteration will converge to the minimum norm least-squares solution if $||A^tA||_2 < 2$ and the iteration is initialized with $f^{(0)} = 0$.

The *Landweber* method was extended by Strand [67] to the *Generalized Landweber* iteration method, which uses:

$$f^{(n+1)} = f^{(n)} + DA^t(g - Af^{(n)})$$

for the iteration. The matrix D, called the shaping matrix, is a linear operator that can be designed as a polynomial function of A^tA to emphasize some frequency components of the image and to accelerate the convergence of those components. In [91] an investigation into the behaviour of these algorithms is presented [5].

5 Statistical Methods

The application of statistical models [8, 11, 39] to imaging problems has been a topic of much recent interest due to their generality for modelling and their capacity to incorporate various constraints. Among the imaging problems which has been solved using statistical models are the following: Single Photon Emission Computed Tomography (SPECT) [45], classification of satellite data [8], image restoration [40, 51, 78, 86], boundary detection [33], Positron Emission Tomography (PET) [38, 43, 44, 46, 50, 57, 88], surface reconstruction [53, 54, 87], microscopic image reconstruction [60, 62, 63, 64], stereo processing [84, 93], information fusion [10], image sequence analysis [92], etc.

Though some of the linear methods presented in the previous section acknowledge the presence of noise in the image, none of them is based on statistical models. *Statistical methods* are those that consider specific probability laws for the noise, and in some cases, for the observed data, to control

[5] Note that the *Generalized Landweber* iteration method is again similar to the Van Citter iteration. Differences between the algorithms are that the *Generalized Landweber* iteration involves matrices, and D is restricted to be a linear operator.

the propagation of noise to the estimation. There are two different ways to solve statistical problems: the direct and the Bayesian approach. In the first case, we have a model describing how the unknown data are related to the observational data (the model comes from some underlying theory that the data are supposed to satisfy), and a model for the noise (i.e. Gaussian or Poisson). We asume that data are subject to *measurement errors* called *noise* in the context of electronic imaging. Thus, typical data never exactly fit the model that is being used. In the case of Bayesian modelling we also have a statistical model for the observed data, called *prior* knowledge in the context of Bayesian estimation. This *knowledge* can be used to restrict the space of possible solutions to the estimation problem, as well as to accelerate some algorithms.

There are different ways of implementing these two appraches on imaging problems [31]. We have chosen two methods: the *Maximum Likelihood Estimation* (MLE), and the *Maximum a Posteriori* approach (MAP). In both cases we have used the *Estimation-Maximization* (EM) algorithm [59] to compute the image estimates. The relevance of the EM algorithm for imaging problems was first noticed by Shepp and Vardi [57], who applied it to a statistical model of PET. Since then, the EM algorithm has attracted a lot of research: its extension to Bayesian processing [38, 44, 89], its acceleration [68, 69, 70, 71], its smoothed versions [58], etc.

5.1 Experiment description

We have applied statistical methods to the problem of *depth from focus*. First of all we will describe the experiment and how the data are produced. This experiment simulates the acquisition of a sequence of images with an optical system with small depth of focus. We assume that the optical system is calibrated and that we know the exact form of the PSF's corresponding to each one of the possible states of the optical system.

Given a scene in an space Ω, usually R^3, located in front of the optical system and formed by opaque objects, we can acquire different images of the scene by changing the optical parameters. In the microscope case this means changing the distance between the scene and the lens (leaving the distance between the lens and the image plane constant). Also in this case, we assume that the projection of the scene on the image plane is represented by a paralell projection equation.

Ω, which is the *observable* space, emits photons according to a *emission density function* λ defined at all its points. As we are dealing with visual scenes, composed of 3D opaque objects, only some points of Ω will emit radiation: those corresponding to the surfaces of the objects. And among those points, the only observable radiation is that emited by the *visible surface* of the scene.

Definition 1 *The* **visible surface** *of a visual scene is the surface (not necessary continous) formed by all the emiting points of the objects composing the scene which are visible from the observer point of view.*

Then, it is more useful to assume that Ω holds a surface $s(x,y)$ - *the visible surface* from the observer position- which takes values $\lambda(x,y,s(x,y))$ corresponding to the radiometry of the physical object surface. [6]

Definition 2 *A* **scene** (λ, s) *is a pair composed by a* visible surface $s(x,y)$ *and an* emision density function $\lambda(x,y,s(x,y))$ *defined on the surface.*

When we reference points of the observable or the observed spaces, we can use two types of notation. The first one, written as λ_i or X_j, will be used when expressions denote the value of the function in those spaces and the x, y and z coordinates can be considered homogeneous (that is, they have no special relevance in the expression). On the other hand, if one coordinate, usually z, is processed in a different way than the others, we will use the classical expression $\lambda(x,y,z)$.

Another assumption we can make is that we are dealing with "real world" scenes, and then, we can consider a physically generic assumption about them: these surfaces are continuous almost everywhere, with the exception of the regions corresponding to the physical edges between objects. This fact will be important when discussing smoothness constraints on the solution of *ill posed* problems.

The objective of the method (in fact, the only one physically achievable) is the recovery of $s(x,y)$. Given $s(x,y)$, λ and a given state of the optical system (we will call this situation the *direct problem*), the generation of the image of the scene X is almost trivial [15]. Each value $\lambda(x,y,s(x,y))$ is spread

[6] Using a suitable coordinate frame $s(x,y)$ can be seen as a function representing the distance between the lens and the visible surface of imaged objects.

by a PSF depending on the value of $s(x,y)$. The superposition of all these blurred points generates the observed image $X_z(x,y)$, where z represents an image plane.

On the other hand, the *inverse problem* of recovering $s(x,y)$ (and eventually $\lambda(x,y,s(x,y))$) can be easily stated, but it is an ill posed problem. In the next sections we will propose a solution to this inverse problem.

We have generated a sequence of images, from different scenes, following this physical model. Images are of size 128×128. All images have 256 grey levels. Each of the experiments involve blur. Blur is generated by convolving the 3×3 mask:

$$\frac{1}{9}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

with itself. Two iterations yields the 7×7 mask:

$$\frac{1}{729}\begin{pmatrix} 1 & 3 & 6 & 7 & 6 & 3 & 1 \\ 3 & 9 & 18 & 21 & 18 & 9 & 3 \\ 6 & 18 & 36 & 42 & 36 & 18 & 6 \\ 7 & 21 & 42 & 49 & 42 & 21 & 7 \\ 6 & 18 & 36 & 42 & 36 & 18 & 6 \\ 3 & 9 & 18 & 21 & 18 & 9 & 3 \\ 1 & 3 & 6 & 7 & 6 & 3 & 1 \end{pmatrix}$$

and three iterations yields a 9×9 mask with an approximately Gaussian shape.

5.2 Image Estimation: the MLE approach

Assume an *emission density function* λ defined in Ω. This emission is sensed by a CCD camera, equiped with an optical system characterized by its small depth of focus. Physical theory allows us to express the observed emission Y as independent random variables with Poisson distribution [78, 65]:

$$Y_j \sim Poisson(X_j)$$

where the Poisson means X_j can be expressed in terms of the emision density function:

$$X_j = \sum_i h_{ij} \lambda_i$$

This transformation can be written in matrix notation as $X = H\lambda$, where $X = (X_1, \ldots, X_n)$ and $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ are vectors, and H is a matrix representing the 3D Point Spread Function (PSF), characteristic of the optical system. We wish to estimate the function λ from the observed $Y_j's$ and the known h_{ij}.

A good choice for the estimation of λ is the one that maximizes the likelihood function $p(Y|\lambda)$. The likelihood of the observations Y is maximized by any solution $\tilde{\lambda}$ of the equation $H\tilde{\lambda} = Y$. But there will not typically be any exact solution due to the noise. In this case we can propose a *Maximum Likelihood Estimation* (MLE) of λ.

EM Algorithm

The *Expectation-Maximization algorithm* ([59]) finds maximum likelihood parameter estimators in problems where some variables were unobserved. This algorithm was first presented by Dempster in [59]. It has been widely applied to image processing problems, mainly in image reconstruction and restoration.

The algorithm estimates parameters iteratively, starting from some initial guest. Each iteration consists of an Expectation step (E-step), which finds the distribution of the unobserved variables from the current estimate of the parameters, and a Maximization step (M-step), which reestimates the parameters to be those with maximum likelihood, under the asumption that the distribution found in the E-step is correct. It can be shown that each iteration improves the true likelihood, or leaves it unchanged, if a local maximum has already been reached. This is true for any GEM (Generalized EM) algorithm, in which only a partial maximization is performed in the M-step.

The key idea of the algorithm is to find a problem formulation involving B, the observed incomplete data, C, the unobserved complete data (in a such way that there is a known many-to-one mapping from C to B), and A, the parameters to be estimated.

Consider the estimation of A by maximizing the log-likelihood $log(B|A)$ based on B. These data can be regarded as an incomplete version of C, which we we would like to have been able to observe. The algorithm gives a two step iteration for increasing the likelihood of the current estimate A^k of A. The complete data have to be chosen in such a way that computing the ML estimate of A from the complete data is simpler than from the incomplete data.

$$\textbf{E-step} : E\{log(p(C|A))|B, A^k\}$$
$$\textbf{M-step} : A^{k+1} = arg\{max_A E\{log(p(C|A))|B, A^k\}\}$$

Sufficient convergence conditions are that $E\{log(p(C|A))|B, A^{(k)}\}$ is continuous in both A and A^k.

The likelihood is:

$$p(Y|\lambda) = \prod_j X_j^{Y_j} exp(-X_j)/Y_j! \tag{8}$$

and its logarithm:

$$log(p(Y|\lambda)) = \sum_j \Big[Y_j log(X_j) - X_j \Big] \tag{9}$$

This likelihood function has been shown to be concave in [57]: hence, a vector λ^{ML}, the maximum likelihood estimate at which $log(p(Y|\lambda))$ reaches its maximum, exists. However, a closed form formula for finding λ^{ML} as a maximizer of $log(p(Y|\lambda))$ does not exist.

Consider the estimation of $\tilde{\lambda}$ by maximizing the log-likelihood $log(p(Y|\lambda))$ based on Y. These (observed) data can be regarded as an incomplete version of ξ, a *complete* data that we would like to have been able to observe, and such that there is a known many-to-one mapping from ξ to Y. The EM algorithm gives a two step process for increasing the current estimate $\lambda^{(n)}$ of λ. In the first step, the E-step, the distribution of ξ given Y and $\lambda^{(n)}$ is estimated. In the second step, the M-step, the parameters $\lambda^{(n+1)}$ must be reestimated in such way to be those with maximum likelihood under the assumption that the distribution found in the E-step is correct. It can be shown that each iteration improves the true likelihood. That is, starting from a strictly positive initial estimate [7] this algorithm computes a sequence of estimates which converge to the MLE.

In our case, the E-step and the M-step can be expressed in a single iterative step [92]. Under the assumption that $\{\xi_i\}$ are independent Poisson variables, it has been shown [57] that:

$$E\{log(p(\xi|\lambda))|Y, \lambda^{(n)}\} = \sum_i \sum_j [-h_{ij}\lambda_j + \frac{Y_i h_{ij} \lambda_j^{(n)}}{\sum_k h_{ik} \lambda_n^{(n)}} log(h_{ij}\lambda_j)]$$

This is the E-step of the algorithm. The M-step is performed by taking the first derivative of this equation with respect to λ_j and equating to zero. The solution is:

[7] Theoretically, the results at convergence are independent of the initial guess, but if the process is stopped before reaching the convergence, which is the common case, the resulting image depends on the starting point. See [43, 47] for a discussion on this point.

$$\lambda_i^{(n+1)} = \frac{\lambda_i^{(n)}}{\sum_j h_{ij}} \sum_j \frac{Y_j h_{ij}}{\sum_k \lambda_k^{(n)} h_{kj}} \tag{10}$$

which constitutes the single iterative step of the plain EM algorithm.

Until now we have not discussed the point of how many *image planes* must be acquired in order to have a good behaviour of the algorithm. Each image plane will be composed by all the observed values $Y(x, y, z)$ corresponding to the specific z of that plane. If we take only one plane, the problem will be underdetermined, and the ML estimation will converge to a solution which can be wrong (and it will converge very slowly). If we look at the iterative step of the EM algorithm, we can see that it is based on a quotient between the *observed* data Y_j and the projection of the *estimated* data: $\sum_k \lambda_k^{(n)} h_{kj}$. In order to have a good estimation it is necessary to have as many *observed* data as *estimated* data That is, every point of the visible surface must be perfectly focused in almost one image plane. That is, we need as many images as depth levels we want estimate. If we have more than this, the problem is overdetermined, but it appears that this does not result in a perceptible improvement of the estimation. Nevertheless, it must be pointed out that the form of the algorithm does not change whatever the number of image planes is.

The main avantages of a MLE approach are:

- It is a sound optimization method that is based on the physics of the imaging system, and which acknowledges the presence of noise.
- It results in a tractable and implementable method.

In general, this additive Poisson model is treated as an artificial statistical model, i.e. it is merely used as a gimmick to generate an algorithm rather than because it is assumed to model any statistical variability in the data. An exception is in Emision Tomography, where the Poisson model is regarded as a realistic aproximation for modelling the data.

Vardi and Lee [92] have recently proposed an alternative interpretation for the EM algorithm, which does not presupose Poisson statistics for the data. The formulation of the problem is based on the following problem: given a vector $Y = \{y_j\}$ and a matrix $H = \{h_{ij}\}$ with non negative entries, the objective is to solve the set of equations:

$$y_j = \sum_i \lambda_i h_{ij} \qquad f_i \geq 0, i = 1, \ldots, M, j = 1, \ldots, N. \tag{11}$$

for a vector $\lambda = \{\lambda_i\}$. This problem, known as *linear inverse problem with positivity restriction (LININPOS)* can be interpreted as an statistical estimation problem from incomplete data based on infinitely large "samples", and that ML estimation and the EM algorithm provide a straighforward method of solution for such problems.

Consider the Kullbach-Leibler information divergence between two probability mass functions $p = (p_1, \ldots, p_N) > 0$ and $q = (q_1, \ldots, q_N) \geq 0$:

$$D(\{y_i\} || \{\sum_i \lambda_i h_{ij}\}), f \geq 0, \sum_i f_i = 1$$

If equation (11) has solution, it can be obtained as the argument that minimizes:

$$\frac{argmin}{f \geq 0}[D(\{y_i\} || \{\sum_i \lambda_i h_{ij}\})] = \frac{argmin}{f \geq 0}[\sum_j y_j log(\sum_i \lambda_i h_{ij})] \tag{12}$$

It follows that the right-hand side of (12) is a solution of equation (11) whenever it exists.

Also, the right-hand side of (12) can be interpreted as the MLE of the following incomplete data problem: the complete (unobservable) data are independent and identically distributed pair of random variables with joint distribution:

$$p\{A = i, B = j\} = \lambda_i h_{ij}, 0, i = 1, \ldots, M, j = 1, \ldots, N.$$

The incomplete (observed) data are the B's alone, and y_j is the observed proportion of B's with value j. The MLE based on the observed data is the right-hand side of (12), and since this is a standard estimation problem for incomplete data we can apply the EM algorithm, giving the iteration (10).

If the problem does not have a non-negative solution, then the algorithm finds the closest aproximation to it, in the Kullbach-Leibler sense.

The algorithm (10) can be derived in at least one more way. As it has been shown previously, there are some techniques to derive iterative algorithms to solve non-linear equations. If we consider equation (6) with $p = 1$, the resulting formula is identical to (10). This fact shows that the EM algorithm can be generated in different ways to solve the same problem.

5.3 Image Estimation: the Bayesian Approach

The Bayesian approach provides an elegant way of formulating imaging problems in terms of probability theory, using the Bayes rule:

$$p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)} \tag{13}$$

Here X could be a specific visual scene and Y the image of a scene. The probabilities can be interpreted as : (i) $p(X|Y)$ the probability of a given scene given the image (also known as *Posterior Probability*), (ii) $p(Y|X)$ the probability of an image given the scene, (iii) $p(X)$ the probability of looking at a given scene, and (iv) $p(Y)$ the probability of an image. Its clear that $p(Y|X)$ represents the model of how an image is formed from the scene (a physical model), and $p(X)$ the *a priori* probability of a given scene ($p(Y)$, the probability of an image, is a normalization factor which can be determined from $p(Y|X)$ and $p(X)$). In general, $p(X)$ will represent the constraints on the solution.

The problem is usually formulated as finding the most probable scene X given some *a priori* knowledge about it $p(X)$ (i.e. some general condition like smoothness), the image Y and a model of image formation $p(Y|X)$, that is, finding a scene which maximizes the posterior probability, which is called the MAP estimate.

In this context, Bayesian inference can be seen as a natural consequence of the idea of using energy functionals to impose smoothness constraints in early vision problems. This idea has been very influential, and has been formalized by Poggio and Torre [30]. They noted the similarity of these methods to a branch of mathematics called *regularization theory*, and proposed a unified framework for vision algorithms. Regularization required that the solution to a problem depends smoothly on the data. The inability to deal with discontinuities had important practical advantages: the energy functions tended to be convex and not have local minima. Thus, they could be minimized by simple methods such as gradient descent.

For example, in the one dimension case, if data d is specified on a regular lattice, the energy functional is discretized, and we choose the smoothness operator to be $\partial f/\partial x$ [10], the energy functional becomes:

$$E(f_i) = \sum_i \{f_i - d_i\}^2 + \lambda \sum_i \{f_{i+1} - f_i\}^2$$

This minimum can be found by steepest descent techniques by iterating:

$$f_i(t+\delta t) = f_i(t) - K\frac{\partial E}{\partial f_i}$$

where K is a constant.

This energy-regularization approach can be incorporated directly in the Bayesian framework using the Gibbs distribution (see below):

$$p(f_i) = \frac{e^{-\beta E(f_i)}}{Z}$$

In this way we can express constraints in a probabilistic framework.

It is possible in some cases to find an extension of the EM algorithm to MAP (Maximum a Posteriori Probability) problems. Maximum likelihood (ML) and MAP approaches to parameter estimation are related by Bayes rule:

$$p(\lambda|Y) = \frac{p(Y|\lambda)p(\lambda)}{p(Y)} \tag{14}$$

where $p(\lambda)$ and $p(Y)$ are the "a priori" probability distributions of the parameters and the observed data. Of course, this approach requires the effort of specifying a prior model and computing and summarizing the joint distribution.

In this case, assuming a normal distribution of the prior (which corresponds to a penalty function which is a weighted sum of the squared deviations of image components from their a priori mean values), the function to be maximized is [38]:

$$log(p(\lambda|Y)) = log(p(Y|\lambda)) - \frac{\alpha}{2}(\lambda - m)^t R(\lambda - m)$$

where R is a diagonal matrix with elements r_j. The expected value of $log(p(\lambda|Y))$ given $\lambda^{(n)}$ is:

$$\sum_i \sum_j [-h_{ij}\lambda_j + \frac{Y_i h_{ij}\lambda_j^{(n)}}{\sum_k h_{ik}\lambda_k^{(n)}} log(h_{ij}\lambda_j)] - \frac{\alpha}{2}(\lambda - m)^t R(\lambda - m)$$

This is the E-step of the algorithm. The M-step is performed by setting to zero the first derivative of this equation with respect to λ_j:

$$\frac{\partial log(p(\lambda|Y))}{\partial \lambda_j} = -1 + \frac{\lambda_j^{(n)}}{\lambda_j} \sum_i \frac{h_{ij} Y_i}{\sum_k h_{ik} \lambda_k^{(n)}} - \alpha r_j (\lambda_j - m_j) = 0$$

This system can be solved easily, yielding the new estimate λ_i^{n+1}:

$$\lambda_i^{(n+1)} = \frac{\alpha r_i m_i - 1 + \sqrt{(\alpha r_i m_i - 1)^2 + 4\alpha r_i \hat{\lambda}_i^{(n+1)}}}{2\alpha r_i} \tag{15}$$

$$\hat{\lambda}_i^{(n+1)} = \frac{\lambda_i^{(n)}}{\sum_j h_{ij}} \sum_j \frac{Y_j h_{ij}}{\sum_k \lambda_k^{(n)} h_{kj}}$$

that only involves an additional evaluation to the standard EM algorithm.

This algorithm constitutes an extension of the EM algorithm to deal with normally distributed priors in the framework of Bayesian estimation. In the next section we will see other techniques to express the *prior* knowledge about the estimated data.

5.3.1 Markov Random Fields and Bayesian processing

A MRF *(Markov Random Field)* is a probability distribution defined over a discrete field where the probability of a particular variable u_i depends only on a small number of its neighbors,

$$p(u_i|\mathbf{u}) = p(u_i|\{u_j\}), u_j \in N_i \tag{16}$$

We can use MRF's to model the correlated structure of dense fields or the smoothness inherent in visible surfaces and natural patterns.

We can specify conditional probabilities for particular configurations. However, calculating $p(u)$ such that all the marginal distributions are correct is a difficult problem, in general. There is a single way of specifying a probability distribution whose conditional probabilities are Markovian [9]. We can use a Gibbs (or Boltzman) distribution of the form:

$$p(u) = 1/Z_p exp(-E_p(u)/T_p) \tag{17}$$

where T_p is the temperature of the model and Z_p the partition function:

$$Z_p = \sum_u exp(-E_p(u)/T_p)$$

The energy function $E_p(u)$ can be written as:

$$E_p(u) = \sum_{c \in C} E_c(u)$$

where each clique energy $E_c(u)$ depends only on a very few neighboring points. In this way one can give a probabilistic interpretation to problems involving minimization of energy functionals, which are a special case of the Bayesian formalism corresponding to MRF's.

If we wish to generate a random sample from the distribution (16), we can use an algorithm called *Gibbs sampler*. This iterative algorithm successively updates each state variable u_i by randomly picking a value from

$$p(u_i|u) = \frac{1}{z_i} exp(-E_p(u_i|u)/T_p)$$

where $Zi = \sum_{u_i} exp(-E_p(u_i|u)/T_p)$. This random updating is guaranteed to converge to a representative sample from the Gibbs distribution. To speed its convergence, simmulated annealing can be used.

The MRF techniques have attracted much attention due to their capacity to incorporate various constraints [40, 41, 42, 44, 46]. Furthermore, MRF's fit naturally in a Bayesian formulation. Finally, it was pointed out in [14] that standard regularization techniques can be considered as special cases of the MRF approach.

A coupled Markov random field model was introduced in [33] in which image restoration and edge detection are performed simultaneously. The basic idea was to include a *line process* that is indexed by the dual lattice and suspends the continuity constraint associated with the pixel pair (f, l) when the line is "on", but preserves the constraint when the line is "off". In one dimension, this can be done by defining an energy functional:

$$E(F_i, l_j) = \sum_i \{f_i - d_i\}^2 + \lambda \sum_i \{f_{i+1} - f_i\}^2 (1 - l_i) + \mu \sum_i l_i$$

Using the Gibbs distribution on $E(F_i, l_j)$ we obtain two coupled MRF's, f_i and l_j. The goal is to find the most probable configuration of f_i and l_i

given this distribution, or alternatively, to minimize $E(F_i, l_j)$ with respect f_i and l_i simultaneosly. This minimization is not straightforward as usually $E(F_i, l_j)$ will have many local minima, and involves using a Monte Carlo algorithm [33].

Gibbs Sampler

This method was proposed in [33]. The optimization of a non convex energy function can be carried out via simulated annealing procedure that uses the *Metropolis* algorithm. The intensity of a pixel in the current image estimate is altered, and the change of energy $\triangle E$ due to the alteration is calculated. A move that reduces energy is accepted always, and a move that increases energy is accepted with probability $exp(-\triangle E/T)$, where T is the temperature of the system:

1. Choose a new state u_i randomly from the set of allowable states.
2. Compute $\triangle E$.
3. if $\triangle E \leq 0$, make the move.
   if $\triangle E > 0$, generate a random number $r \in [0, 1]$,
   (a) if $r \leq exp(-\triangle E/T)$, make the move.
   (b) if $r > exp(-\triangle E/T)$, leave u_i unchanged.

We can formulate an energy minimization problem to describe the shape from focus paradigm, in the same way as other computer vision problems have been stated [10]. In this case the expression to be minimized could be:

$$E(\lambda, o) = \sum_{x,y,z} \left| \sum_{x',y',z'} h(x,y,z,x',y',z')[\lambda(x',y',z')o(x',y',z')] - Y(x,y,z) \right| + \mu \sum_{x,y} [(\sum_{z'} o(x,y,z')) - 1] \quad (18)$$

where μ is a constant and $o(x, y, z)$ is a binary function which represents the support of the visible surface. Obviously, this function has a minimum when $o(x, y, z)$ represents the visible surface and λ the intensity pattern on the surface. We have not implemented this version of the *shape from focus* problem because of the high nonlinearity of the expression. This optimization problem could be solved using nonlinear optimization techniques like the Gibbs sampler, but in this case the cost for each iteration of the algorithm (which includes a projection in the *observable* space of the estimated scene) is too high to be implemented. Nevertheless, it is a good example of MRF modelling for the shape from focus problem.

6 Surface recovery

As we are dealing with opaque objects, the recovery of λ will give us the observed surface in a implicit way. λ is defined in 3D space, and because of the image formation process, for every possible (x, y) there exists only one point (x, y, z) with a positive value. This value corresponds to the radiance value of the surface on that point. If we iterate the algorithm until its convergence, and the noise due to the imaging system is relatively low, the solution λ will represent a surface that can be recovered in a straightforward way by looking for all z where (x, y, z) is positive (in the case of black points -with null emission- the surface cannot be recovered). But as we have seen previously, some practical problems prevent it. To overcome these problems, we have adopted the Bayesian way -with some new strategies- which allows a slight improvement of the convergence rate and an iterative recovery of the surface.

6.1 1st Case: Avalability of a prior image.

Sometimes, in spite of the fact that we are using an optical system with limited depth of focus, we can acquire a perfectly focused image, free of out-of-focus effects.This image can be obtained using a small aperture (in the limit, a pinhole camera would produce always perfectly focused images). By acquiring images with a greater aperture and a neutral density filter to equalize the exposure, we can obtain a series of partially defocused images of the same scene, which are rich in depth information. These last images can

be regarded in the framework of image estimation as the "observed" degraded (by noise and out-of-focus) images, and the focused image as the prior grey level distribution of the "restored" image. In this case, the problem of depth perception is reduced to the estimation of the prior deformation during the restoration process.

First of all, it may be useful to remember that the algorithm may be seen as a special type of gradient-ascent algorithm. In each step, the algorithm reaches an estimated solution which is nearer from the ML solution than the previous ones. When using the Bayesian version of the algorithm, the estimated solutions will increase the posterior probability distribution.

As we have seen, at the MAP solution

$$\frac{\partial(ln(p(\lambda|Y)))}{\partial\lambda_i} = \frac{\partial(ln(p(Y|\lambda)))}{\partial\lambda_i} - \frac{\partial(\frac{\alpha}{2}(\lambda-m)^t R(\lambda-m))}{\partial\lambda_i} = 0$$

To build geometric flexibility into the prior, we can consider m to be a function of several parameters, $m(d)$, which will represent the shape of the prior. Subsequently, the Bayesian solution that maximizes the *a posteriori* probability must satify

$$\frac{\partial(ln(p(\lambda|Y)))}{\partial\lambda_i} = 0 \quad \text{and} \quad \frac{\partial(ln(p(\lambda|Y)))}{\partial d_i} = 0$$

The use of parametric constraints results in a special case of the Bayesian formalism which is an alternative to the standard smoothness constraint. It assumes a parametrized form for the solution and determines the parameters to best match the data. The choice of the parameterized form corresponds to the prior knowledge. These ideas can be found in [77, 94]. It extends the idea of smoothness constraints to a more general idea of *constraint* which includes other types of constraints like rigidity, Lambertian reflexion assumption, etc. Thus, through the choice of the parameters and the transformations governed by these parameters we can impose constraints to the problem which could be hardly stated in other ways. If visible surfaces could be easily represented using parametric forms (like quadric surfaces), the gradient of $ln(p(\lambda|Y))$ with respect the parameters could be computed from the dependence of m on d_j:

$$\frac{\partial(ln(p(\lambda|Y)))}{\partial d_j} = \sum_i \frac{\partial(\frac{\alpha}{2}(\lambda-m)^t R(\lambda-m))}{\partial m_i}\frac{\partial m_i}{\partial d_j}$$

where the sum is over all the points. But this approach is very restrictive when dealing with the representation of visible surfaces [9]. In this case, the most useful techniques to model visibles surfaces have been spline-based representations and MRF. Both of these approaches are examples of *locally* parameterized models of shape.

In a first aproximation we have not considered any *prior* model for visible surfaces. Surfaces are represented by a function $s(x, y)$ with no shape constraints. We allow a warped prior restricted by only a constraint: the prior must be a surface and must "cover" all the observable space. In other words, for every (x, y), the prior must be defined in one and only one $(x, y, z) \in \Omega$. This fact significantly simplifies the estimation of the prior deformation, since it is reduced to seeking the z which makes the difference $|\lambda^{(k)}(x, y, z) - m(x, y)|$ smaller for every (x, y). The set of z coordinates determined in this way define the shape of the prior.

We leave the use of more advanced methods to model the warping as an open problem to be solved in the future . It seems that linear models developed for surface reconstruction (based on generalized splines), and statistical models can be incorpored in this framework as a way of expressing the *prior* knowledge about the surface. In general, this knowledge is reduced to considering smooth surfaces with discontinuities.

Using this technique, it has been possible to recover visible surfaces from sequences of computer generated images representing different scenes acquired with an optical system. Besides the estimation of the visible surface, this approach results in the acceleration of the convergence of the plain EM algorithm.

6.2 2ond Case: Non-avalability of a prior image.

In some situations, where depth from focus estimation could be of great interest (i.e. when images are acquired with a microscope and depth differences in the scene are greater than the depth of focus of the optical system) it is not possible to have a focused image of the scene to be used as prior image. In these cases, we can introduce the grey level values of the prior as new parameters to be determined during the estimation step. Then, this step involves the determination of the following parameters: the "restored" scene, the prior values, and the deformation of the prior.

While the estimation of the "restored" scene -given an estimation of the

prior values and a deformation- can be achieved with the same method that is used in the first case, the question of how to estimate the intensity values of the prior must be carefully discused. The deformation can be determined in the same way as in the previous case, but now the prior must also be determined at each step. It must be noted that the prior image is not an objective of its own, and that we are mainly interested in the resulting deformation.

Our method for estimating the prior has its origin in the observation of the behaviour of the EM algorithm. When we iterate the algorithm in order to estimate the scene which has produced the observation, we observed that while *high frequencies* are well estimated in the (x, y) planes for every z, the algorithm has some problems in eliminating some low frequencies which have been produced in the first steps. Thus, the scene $\lambda(x, y, z)$ appears as a dense space, that is, almost all the points have positive values. As we have discussed, the algorithm must converge to a solution where there is only one positive value for every (x, y), but in practice this convervenge is too slow. The reason for this slowness is that scenes presenting these problems are very close, from the statistical point of view, to the ML solution. On the other hand, these scenes are not close enough, given a classical norm between images, to the ML scene and they cannot be considered as good estimation [8]. The idea is to force the algorithm to a solution which embeds a surface using the Bayesian approach.

From the fact that the estimated scene is, from the statistical point of view, close to the ML solution we can conclude that it produces an *observation* which is very close to the *observed* image. At the same time, it is very different from a correct scene. In other words, the region of the posterior density function near the maximum is "flat" -which means that large variations in the "solution" produces "small" variations in the value of this function- and the scenes that correspond to this zone have a great variance. The only way we can overcome this problem is by using the Bayesian term, the prior, as a tool to navigate in the right direction.

The prior values are determined in the following way: given an estimate $\lambda^{(n)}$, the prior values may be computed as $I^{(n+1)}(x, y) = \sum_z \lambda^{(n)}(x, y, z)$. At each iteration step, the prior is computed by integrating the values of

[8]This fact is characteristic of the problem we are dealing with, and has no correspondence with any similar problem when using the algorithm to restore or reconstruct two dimensional images or dense three dimensional volumes

the estimated solution for each z tube. In this way, we build a prior which supports the fact that the solution must be a surface in Ω. We have tested this algorithm with synthetic data, and we have observed no bad effects in the reconstruction algorithm. Moreover, the prior estimated in this way converges in a few steps to the right one.

The whole algorithm can be written as follows:

1. Initialize $\lambda^{(0)}$ and $m^{(0)}$ with a constant value for all (x,y,z);
2. $k = 0$;
3. Until we reach convergence {
   (a) Compute $\lambda^{(n+1)}$ with (15) with $m_i = m_i^{(n)}$;
   (b) $I^{(n+1)}(x,y) = \sum_z \lambda^{(n+1)}(x,y,z)$
   (c) For every (x,y,z), determine the best z such that if $m^{(n+1)}(x,y,z) = I(x,y)$, then $|m(x,y,z)^{(n+1)} - \lambda^{(n+1)}(x,y,z)|$ is minimized. }
4. The depth map is computed from $m^{(n+1)}$ by looking at every (x,y) which is the z such that $m^{(n+1)}(x,y,z) \neq 0$.

The surfaces estimated after 500 and 20000 iterations of the algorithm are shown in figure (6). The original data was generated from a scene with a pyramid shape. As we can see, the algorithm converges to the exact solution.

7 Conclusions and future research

The most important conclusions are the following:

- Iterative linear methods are better options for image restoration than matrix-based methods because of their implementability. Non-linear iterative methods, however, perform better in the presence of noise in the image than linear methods.
- Theoretically, it would appear that the MLE approach to the problem of depth from focus permits the recovery of visible surfaces, but

when using the EM algorithm (or other iterative algorithms to find the ML estimate), practical considerations prevent the application of this method.

- The Bayesian approach allows one to express visible surfaces (or their parametric forms) as a set of parameters to be estimated by the EM algorithm, resulting in an intensity-based method to recover depth from focus.

- The use of parametric constraints results in a special case of Bayesian processing which in some cases provides better control over the estimated parameters than standard smoothness constraints. This approach is useful when visible surfaces are easily represented by parametric forms like quadric surfaces.

- Estimation of visible surfaces based on physical modelling requires simultaneous estimation of the object surface and the intensity pattern on it. The characteristics of this pattern (texture, gradient, etc.) have realty influence on the convergence speed of the EM algorithm. Textured objects, obviously, are better estimated than smooth-patterned objects.

- EM algorithm reaches convergence in a few steps (200-300), but estimated parameters can not be taken as good estimations of the shape from focus problem. This apparent contradiction is caused by the special form of our problem : ML and MAP estimates, which are far from representing physical scenes (composed of opaque objects), are "projected" on the "observable" space very close to real scenes. In spite of this fact, the intensity pattern on the surface can be "reconstructed" very accurately from these data, and can be used within the bayesian framework to obtain correct surface estimates.

- In the general case we need to acquire as many image planes as depth-levels we want to estimate to get a "good" estimate. In the optimal case, we need a focused observation of each point of the visible surface. In the case of having a prior image perfectly focused in all its points, the algorithm can work with only one observation.

One possible way of extending this method is by combining it with an edge-based method to recover depth. These methods are faster than intensity-based methods, and give better estimates of depth for object edges. This combination has been implemented succesfully in *depth from stereo* [93] where the integration was performed by coupling the disparity estimates from an independent edge-based stereo algorithm to the energy functional of the Markov Random Field associated with an intensity-based stereo algorithm. In figure 7 a result from [95], using an edge-based focus method, is shown. The fusion of this information with the estimation of our method could result in a better depth estimation.

Another possible improvement is the use of different forms of *a priori* knowledge about the estimated surface. Some of these methods allow a very real modelling of visible surfaces, although in many cases they are computationally expensive. The use of *generalized splines* models [10, 9, 37] or MRF models [33] could be a way of better regularizing the estimated surface as well as accelerating the convergence of the algorithm (in the sense of getting better surface estimates in fewer steps).

Finally, some effort should be made in order to accelerate the algorithm, like the study of multigrid techniques (in order to accelerate the iterative reconstruction by first using a coarse-grid iteration to provide an initial condition for the more computationally demanding fine-grid iteration) [90, 73, 91] or fast convergence algorithms [71].

8 Aknowledgements

J. Vitrià would like to acknowledge the support of the *Gaspar de Portola Catalonian Studies Program* (University of California, Berkeley and Generalitat de Catalunya, Barcelona) and the Engineering Division of the *Lawrence Berkeley Laboratory* for the use of their facilities during the course of the research reported here.

References

[1] R.L. LAGENDIJK, J.BIEMOND, **Iterative Identification and Restoration of Images.** Kluwer Academic Publishers, 1991.

[2] W.H.PRESS, S.A.TEUKOSLSKY, W.T.VETTERING, B.P.FLANNERY, **Numerical Recipes in C. 2ond Edition.** Cambridge university Press, 1992.

[3] F.B HILDEBRANT, **Introduction to Numerical Analysis. 2ond Edition.** Dover Pub. 1984.

[4] B.K.P.HORN, **Robot Vision,** McGrawHill, 1986.

[5] H.C.ANDREWS, B.R.HUNT, **Digital Image Restoration.** Prentice Hall, 1977.

[6] W.MENKE, **Geophysical Data Analysis: Discrete Inverse Theory.** Academic Press, 1989.

[7] E.P. KROTKOV, **Active Computer Vision by Cooperative Focus and Stereo,** Springer, New York, 1989.

[8] B.D.RIPLEY, **Statistical Inference for Spatial Processes.** Cambridge University Press, 1988.

[9] R.SZELISKY, **Bayesian Modeling of Uncertainty in Low-Level Vision.** Kluwer Academic Pub. 1989.

[10] J.C.CLARK, A.L.YUILLE, **Data Fusion for Sensory Information Processing Systems.** Kluwer Academic Pub. 1990.

[11] NATIONAL RESEARCH COUNCIL, **Spatial Statistics and Digital Image Analysis.** National Acady Press, 1991.

[12] A. BLAKE AND A. ZISSERMAN, **Visual Reconstruction,** MIT Press, Cambridge, MA, 1987.

[13] W.T.WELFORD, **Useful Optics.** Chicago Lectures in Physics, The University of Chicago Press, 1991.

[14] M. BERTERO, T.A. POGGIO, AND V. TORRE *Ill-posed problems in early vision.* Proc. of the IEEE 76, 1988, 869-889.

[15] M.POTMESIL, I.CHAKRAVARTY *Synthetic Image Generation with a Lens and Aperture Camera Model.* ACM Trans. on Graphics, vol.1, no.2, 1982, 85-108.

[16] J. ENS AND P. LAWRENCE. *A matrix method for determining depth from focus.* Int. Conf. on Comp.Vision and Pattern Recognition , 600-606, 1990.

[17] J. ENS AND P. LAWRENCE. *An investigation of methods for determining depth from focus.* Trans. on Pattern Anal. and Machine Intelligence , vol 15, no.2, 1993, pp.97-108.

[18] M. SUBBARAO AND N. GURUMOORTHY. *Depth recovery from blurred edges.* Int. Conf. on Comp.Vision and Pattern Recognition , 1988, 498-503.

[19] M. SUBBARAO. *Parallel depth recovery by changing camera parameters.* Int. Conf. on Computer Vision, 1988, 149-155.

[20] P.GROSSMAN. *Depth form focus.* Pattern Recognition Letters 5, 1987, pp. 63-69.

[21] M. SUBBARAO AND T.C. WEI. *Depth from defocus and rapid autofocusing: A practical approach.* Int. Conf. on Comp.Vision and Pattern Recognition , 1992, 773-776.

[22] K. VENKATESH PRASAD AND R.J. MAMMONE. *Depth restoration from defocused images using simulated annealing.* Int. Conf. on Pattern Recognition -A, 1990, 227-229.

[23] A.P. PENTLAND, *A new sense for depth of field.* Int. Joint Conf. on Art. , 1985, 988-994.

[24] A.P. PENTLAND, *A new sense for depth of field.* Trans. on Pattern Analysis and Machine Intelligence 9, 1987, 522-531.

[25] A. PENTLAND, T. DARRELL, M. TURK, AND W. HUANG. *A simple, real-time range camera.* Int. Conf. on Comp.Vision and Pattern Recognition , 1989, 256-261.

[26] S.LAI, C.FU, S.CHANG, *A Generalized depth estimation algorithm with a single image.* Trans. on Pattern Analysis and Machine Intelligence 14, no.4, April 1992, pp.405-411.

[27] Y. ALOIMONOS AND A. ROSENFELD. *Computer vision.* Science 253, 1991, 1249-1254.

[28] T. POGGIO AND V. TORRE. *Ill-posed problems and regularization analysis in early vision.* Image Understanding Workshop, 1984, 257- 263.

[29] T. POGGIO. *Early vision: from computational structure to algorithms and parallel hardware.* Comp.Vision,Graphics and Image Processing 31, 1985, 139-155.

[30] T. POGGIO, V. TORRE, AND C. KOCH. *Computational vision and regularization theory.* Nature 317, 1985, 314-319.

[31] J. MARROQUIN, S. MITTER, AND T. POGGIO. *Probabilistic solution of ill-posed problems in computational vision.* Image Understanding Workshop, 293-309, 1985.

[32] V. TORRE AND T.A. POGGIO. *On edge detection.* Trans. on Pattern Anal. and Machine Intelligence 8, 1986, 147-163.

[33] S. GEMAN AND D. GEMAN. *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,* Trans. on Pattern Anal. and Machine Intelligence 6, 1984, 721-741.

[34] S.K. NAYAR, *Shape from focus system,* Int. Conf. on Comp.Vision and Pattern Recognition , 1992, 302-308.

[35] H.N. NAIR AND C.V. STEWART, *Robust focus ranging,* Int. Conf. on Comp.Vision and Pattern Recognition , 309-314.

[36] A. BLAKE, *Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction,* Trans. on Pattern Anal. and Machine Intelligence 11, 1989, 2-12.

[37] A. BLAKE AND A. ZISSERMAN, *Localizing discontinuities using weak continuity constraints,* Pattern Recognition Letters 6, 1987, 51-59.

[38] E. LEVITAN AND G.T.HERMAN, *A Maximum a Posteriori probability Expectation Maximization algorithm for image reconstruction in emision Tomography.* Trans. on Medical Imaging 6, 1987, 185-192.

[39] J.BESAG, *On the statistical analysis of dirty pictures,* J.R.Statist.Soc.B, 48, no.3, 1986, 259-302.

[40] V.E.JOHNSON, W.H.WONG, X.HU, C.CHEN, *Image restoration using Gibbs priors: boundary modelling, treatment of blurring, and selection of hyperparameter.* Trans. on Pattern Anal. and Machine Intelligence 13, 1991, 413-425.

[41] V.E.JOHNSON, *A model for segmentation and analysis of noisy images,* Tech.Rep. DP 91-A15, Institute of Statitics and Decision Sciences, Duke University, 1991.

[42] M.GINDI, M.LEE, A.RANGARADAN, I.G.ZUBAL, *Bayesian Reconstruction of functional images using registered anatomical images as priors,* in **"Information Processing in Medical Imaging"** A.Colchester and D.Hawkes (Ed) Lecture Notes in Computer Science 511, Springer Verlag, 1991.

[43] J.LLACER,E.VEKLEROV,J.NUNEZ, *Statistically based image Reconstruction for emision tomography,* Int.J. of Imaging Systems and Technology, Vol.1, 1989, 132-148.

[44] T.HEBERT, R.LEAHY, *A generalizad EM algorithm for 3D Bayesian reconstruction from Poisson data using Gibbs priors,* Trans. on Medical Imaging 8, 1989, 194-202.

[45] S.GEMAN,D.MCCLURE, *Bayesian image analysis: an application to single photon emision tomography,* in Proc.Statis.Comp.Sect., Amer.Statis.Assoc., Washington DC, 1985, 12-18.

[46] R.LEAHY, X.YAN, *Incorporation of anatomical MR data for improved functional imaging with PET,* in "Information Processing in Medical Imaging" A.Colchester and D.Hawkes (Ed) Lecture Notes in Computer Science 511, Springer Verlag, 1991, 105-119.

[47] E.VEKLEROV, J.LLACER, *Stopping rule for the MLE algorithm based on statistical hypothesis testing.* Trans. on Medical Imaging 6, No.4, 1987, pp. 313-319.

[48] R.WILSON, S.SHAFER, *Precision imaging and control for machine vision research at Carnegie Mellon University,* Tech.Rep. CMU-CS-92-118, 1992.

[49] J.LLACER, B.M. TER HAAR ROMENY, M.VIERGEVER, *The use of visual response functions in Bayesian reconstruction,* Tech.Rep. LBL-33054, Lawrence Berkeley Lab., University of California, 1993.

[50] J.NUNEZ, J.LLACER, *A fast Bayesian reconstruction algorithm for emision tomography with entropy prior converging to feasible images,* Trans. on Medical Imaging 9, 1990, 159-171.

[51] W.RICHARSON, *Bayesian-based iterative method of image restoration,* J.Opt.Soc.Ame. 62, no 1, 1972, 55-59.

[52] H.HART,Z.LIANG, *Bayesian image processing in two dimensions,* Trans. on Medical Imaging 6, 1987, 201-208.

[53] J.L.MARROQUIN, *Surface reconstruction preserving discontinuities,* A.I.Mo 792, Massachusetts Institute of Technology, Aug.1984.

[54] D. GEIGER AND F. GIROSI, *Parallel and deterministic algorithms from MRF's: Surface reconstruction,* Trans. on Pattern Anal. and Machine Intelligence 13, 1991, 401-412.

[55] Z.LIANG, *Statistical models of a priori information for image processing: neighboring correlation constraints,* J.Opt.Soc.Ame. A, vol.5, 1988, 2026-2031.

[56] E.MEINEL, *Origins of linear and nonlinear recursive restoration algorithms,* J.Opt.Soc.Ame. A, vol.3, 1986, 787-799.

[57] L.SHEPP, Y.VARDI, *Maximum likelihood reconstruction for emision tomography,* Trans. on Medical Imaging 1, no.2, 1982, 113-122.

[58] B.SILVERMAN, M.JONES, J.WILSON, D.NYCHKA, *A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emision tomography,* J.R.Statis.Soc. B, 1990, 277-324.

[59] A.P.DEMPSTER,N.LAIRD,D.RUBIN, *Maximum likelihood from incomplete data via the EM algorithm,* J.R.Statis.Soc. B, 39, 1977, 1-38.

[60] C.PREZA, M.MILLER, L.THOMAS, J.MCNALLY, *Regularized linear method for reconstruction of three dimensional microscopic objects from optical sections.* J.Opt.Soc.Am.A., Vol.9, No.2, Feb 1992, 219-228.

[61] V.M.BOVE, *Entropy-based depth from focus,* J.Opt.Soc.Am.A., Vol.10, No.4, April 1993, 561-566.

[62] J.CONCHELLO, E.HANSEN, *Enhanced 3-D reconstruction from confocal scanning microscope images. 1: Deterministic and maximum likelihood solutions.* Applied Optics, Vol.29, No.26, Sept. 1990, 3795-3804.

[63] T.HOLMES, Y.LIU, *Acceleration of maximum likelihood image restoration for fluorescence microscopy and other noncoherent imagery.* J.Opt.Soc.Am.A., Vol.8, No.6, Feb 1991, 893-907.

[64] T.HOLMES, Y.LIU, *Richarson-Lucy/maximum likelihood image restoration algorithm for fluorescence microscopy: further testing.* Applied Optics, Vol.28, No.22, Sept. 1989, 4930-4935.

[65] T.HOLMES, *Blind deconvolution of quantum-limited incoherent imagery: maximum-likelihood approach.* J.Opt.Soc.Am.A., Vol.9, No.7, Jul 1992, 1052-1061.

[66] L.LANDWEBER, *An iterative formula for Fredholm integral equations of the first kind,* Amer.J.Math, vol.73, 1951, pp.615-624.

[67] O.N.STRAND, *Theory and methods related to the singular function expansion and Landweber's iteration for integral equations of the first kind.* SIAM J.Numer. Anal., Vol.11, Num.4, 1974, pp.798-825.

[68] R.LEWITT, G.MUEHLLEHNER, *Accelerated iterative reconstrucction for positron emision tomography based on the EM algorithm for maximum likelihood estimation.* Trans. Medical Imaging -5, No.1, March 1986, pp. 16-22.

[69] L.KAUFMAN, *Implanting and accelerating the EM algorithm for positron emision tomography.* Trans. Medical Imaging -6, No.1, March 1987, pp. 37-51.

[70] K.LANGE, M.BAHN, R.LITTLE, *A theoretical study of some maximum likelihood algorithms for emision and transmision tomography.* Trans. Medical Imaging -6, No.2, June 1987, pp.106-114.

[71] N.RAJEEVAN, K.RAJGOPAL, G.KRISHNA, *Vector extrapolated fast maximum likelihood estimation algorithms for emision tomography.* Trans. Medical Imaging -11, No.1, March 1992, pp. 9-20.

[72] K.LANGE, *Convergence of image reconstruction algorithms with Gibbs smoothing.* Trans. Medical Imaging -9, No.4, Dec.1990, pp. 439-446.

[73] M.RANGANATH,A.DAWHAN,M.MULLANI,
A multigrid expectation-maximization reconstruction algorithm for positron tomogrphy. Trans. Medical Imaging -7, 1988, pp.273-278.

[74] R.DANTU, N.DINOPOULOS, R.PATEL, A.AL-KHALILI, *Depth perception using blurring and its application in VLSI wafer probing.* Machine Vision and Applications, 1992, no.5, pp. 35-45.

[75] D.GEMAN, *Random Fields and inverse problems in imaging,* in **Ecole d'Ete de Probabilites de Saint-Flour XVIII-1988**, edited by A.Ancona, D.Geman and N.Ikeda, Lecture Notes in Mathatics, Springer Verlag, 1988.

[76] K.PRASAD, R.MAMMONE, J.YOGESHWAR, *Three-dimensional image restoration using constrained optimization techniques.* Optical Engineering, Apr. 1990, Vol.29, No.4, pp. 279-288.

[77] K.M.HANSON, *Bayesian Reconstruction based on flexible prior models.* J.Opt.Soc.Am.A, Vol.10, No.5, May 1993, pp. 997-1004.

[78] D.SNYDER, A.HAMMOUD, R.WHITE, *Image Recovery from data acquiered with a charge-coupled-device camera.* J.Opt.Soc.Am.A, Vol.10, No.5, May 1993, pp. 1014-1023.

[79] D.AGARD, *Optical Sectioning Microscopy: Cellular Architecture in Three Dimensions.* Ann.Rev.Biophys.Bioeng., 13, 1984, pp.191-219.

[80] N.STREIBL, *Three-dimensional imaging by a microscope.* J.Opt.Soc.Am.A, Vol.2,No.2, Feb 1985, pp. 121-127.

[81] A.ERHARDT, G.ZINSER, D.KOMITOWSKI, J.BILLE, *Reconstucting 3-D light microscopic images by digital image processing.* Applied Optics, Vol.24, No.2, 1985, pp.194-200.

[82] S.LAVALLE, S.HUTCHINSON, *Considering multiple surface hypotheses in a Bayesian hierarchy.* SPIE Vol. 1569, Stochastic and Neural Methods in Signal Processing and Computer Vision, 1991, pp.2-15.

[83] K.HANSON, *Reconstruction based on flexible prior models.* SPIE Vol. 1652, Medical Imaging IV: Image Processing, 1992, pp.183-191.

[84] B.CERNUSCHI-FRIAS, D.COOPER, Y.HUNG, P.BELHUMEUR, *Toward a Model-Based Bayesian Theory for Estimating and Recognizing Parametrized 3-D Objects Using Two or More Images Taken from Different Positions.* Trans. on Pattern Analysis and Machine Intelligence 11, no.10, Oct.1989, pp. 1028-1052.

[85] K.M.HANSON, *Making binary decisions based on the posterior probability distribution associated with tomographic recostructions.* Proc. of the 11th Workshop Maximum Entropy and Bayesian Methods, G.J.Erickson (ed.), 1991, Seattle, Washington, (Kluwer, Dordrecht).

[86] Y.AMIT, U.GRENANDER. M.PICCIONI, *Structural image restoration through deformable templates.* Journal of the American Statistical Association, June 1991, V.86, No.414, Theory and Methods, pp. 376-387.

[87] D.TERZOPOULOS, *The computation of Visible-Surface Representations.* Trans. on Pattern Anal. and Machine Intelligence 10, No.4, July 1988, pp. 417-438.

[88] E.TANAKA, *A fast reconstruction algorithm for stationary Positron Emission Tomography based on a modified EM algorithm.* Trans. Medical Imaging 6, No.2, June 1987, pp.98-105.

[89] P.GREEN, *On the use of the EM algorithm for penalized likelihood estimation.* J.R.Statist.Soc. B, 52, No.3, 1990, pp.443-452.

[90] K.ZHOU,C.RUSHFORTH, *Image restoration using multigrid methods.* Applied Optics, Vol.30,No.20,July 1991, pp. 2906-1912.

[91] T.PAN, A.YAGLE, *Numerical study of multigrid implementations of some iterative image reconstruction algorithms.* Trans. Medical Imaging 10, num 4, 1991, pp.572-588.

[92] Y.VARDI, D.LEE, *From image deblurring to optimal investments: maximum likelihood solutions for positive linear inverse problems.* J.R.Statist.Soc. B, 55, No.3, 1993, pp.569-612.

[93] K.LIM, R.PRAGUER, *Using Markov Random Fields to integrate Stereo Modules,* Tech. Report CUED/F-INFENG/TR109, University of Cambridge, 1992.

[94] L.MALONEY, B.WANDELL, *A computational model of color constancy,* J.Opt.Soc.Am.A, 1986, pp.29-33.

[95] X.BINEFA, J.VITRIA, J.J.VILLANUEVA, *Three dimensional inspection of integrated circuits: a depth from focus approach.* SPIE/IS&T Symposium on Electronic Imaging, Conference on Machine Vision in Microelectronics Manufacturing, San Jose (CA), 1992.

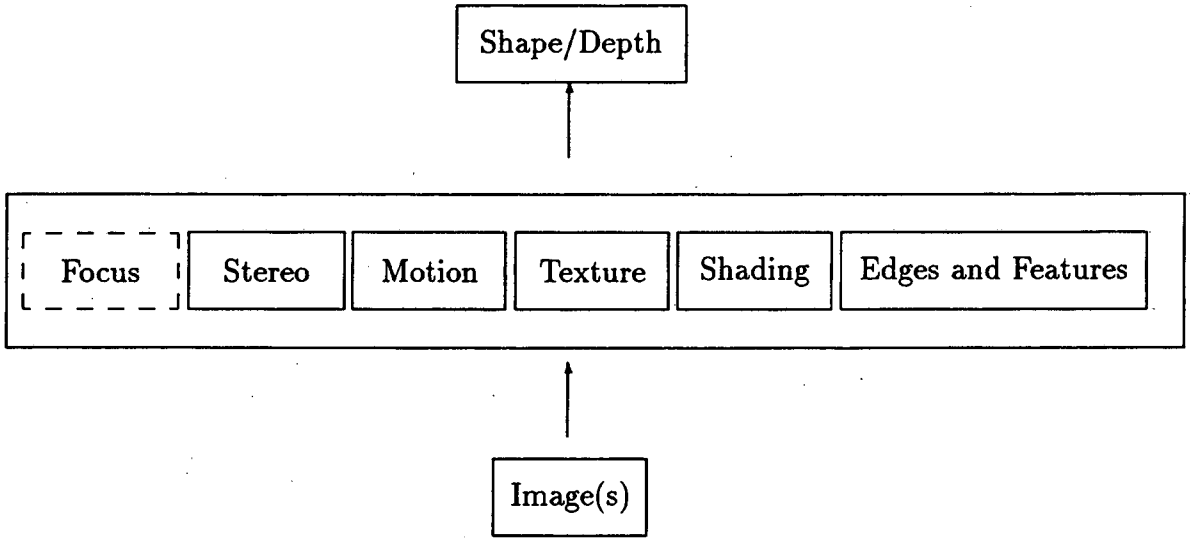


Figure 1: Early vision modules

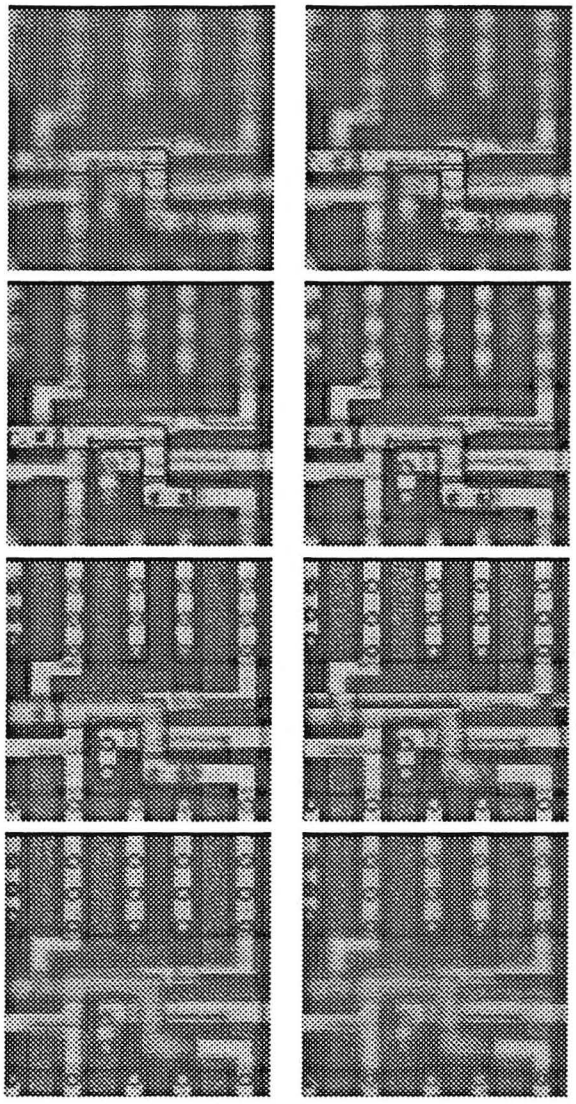


Figure 2: Light microscope focus series corresponding to different layers of an integrated circuit. Three-dimensional inspection of integrated circuits is an important point in the manufacturing process because of the relationship between the *topography* and the electric properties of the circuit. Depth from focus methods provide a non-invasive technique for the estimatation of the three-dimensional profile of the circuit using low-cost equipment. In this series we can clearly perceive how focus gives depth information to the observer.

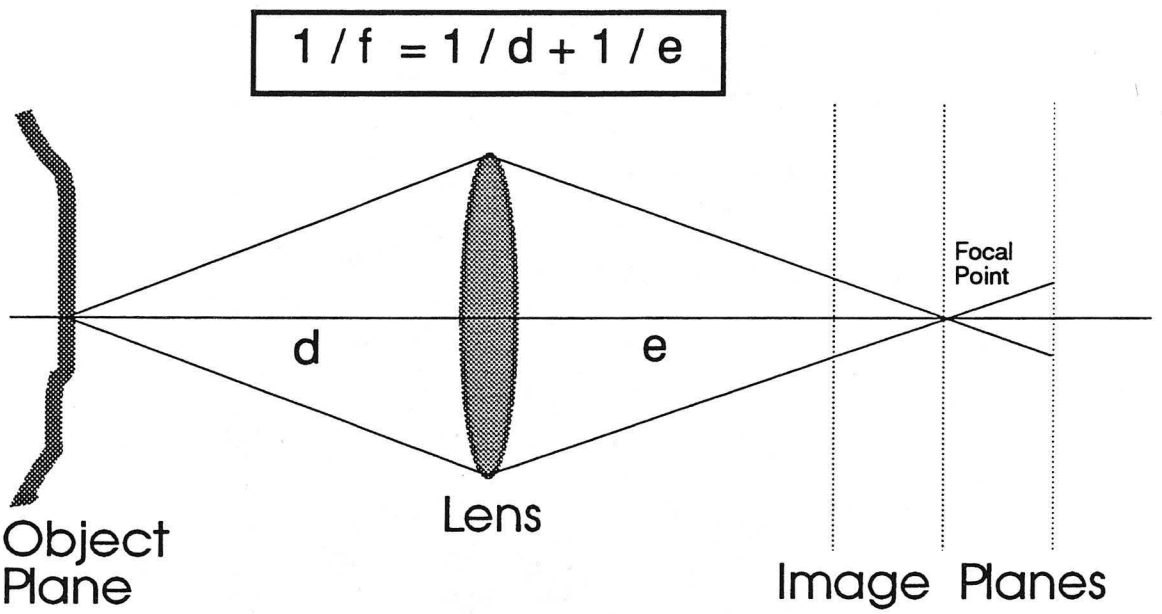


Figure 3: Illustration of the depth of focus.

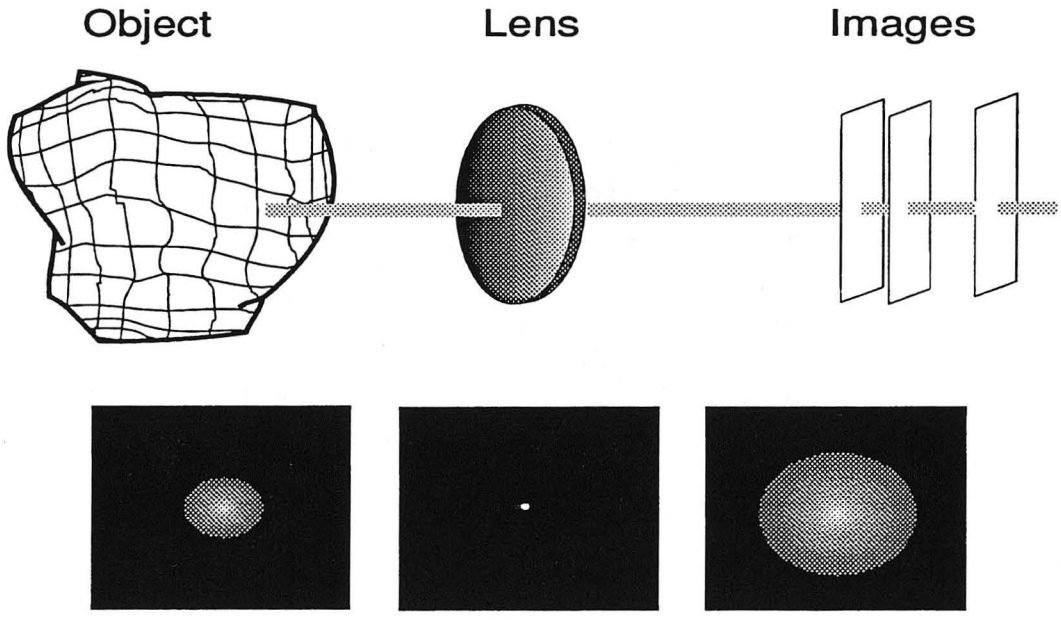


Figure 4: Description of the imaging experiment. Given an object located in front of the the optical system, we can acquire different images of it by changing the distance between the lens and the image plane. Tha image of a visible point of the object will correspond to different versions of the *point spread function* of the optical system

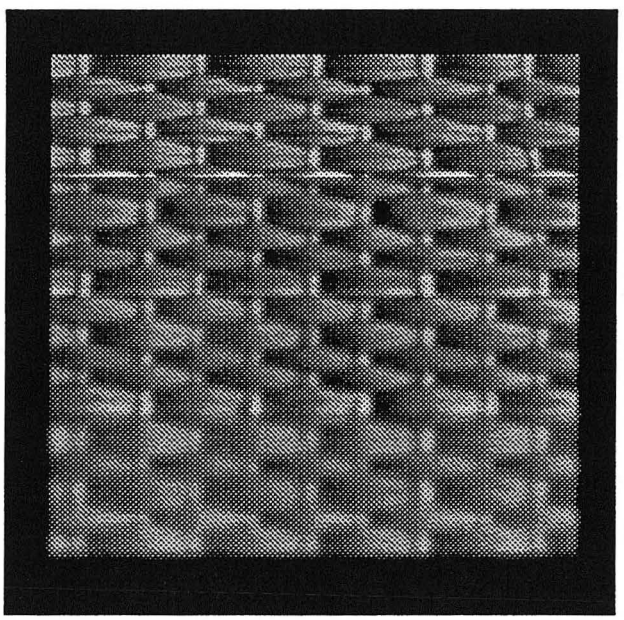


Figure 5: Synthetic image used in the experiment (stair).

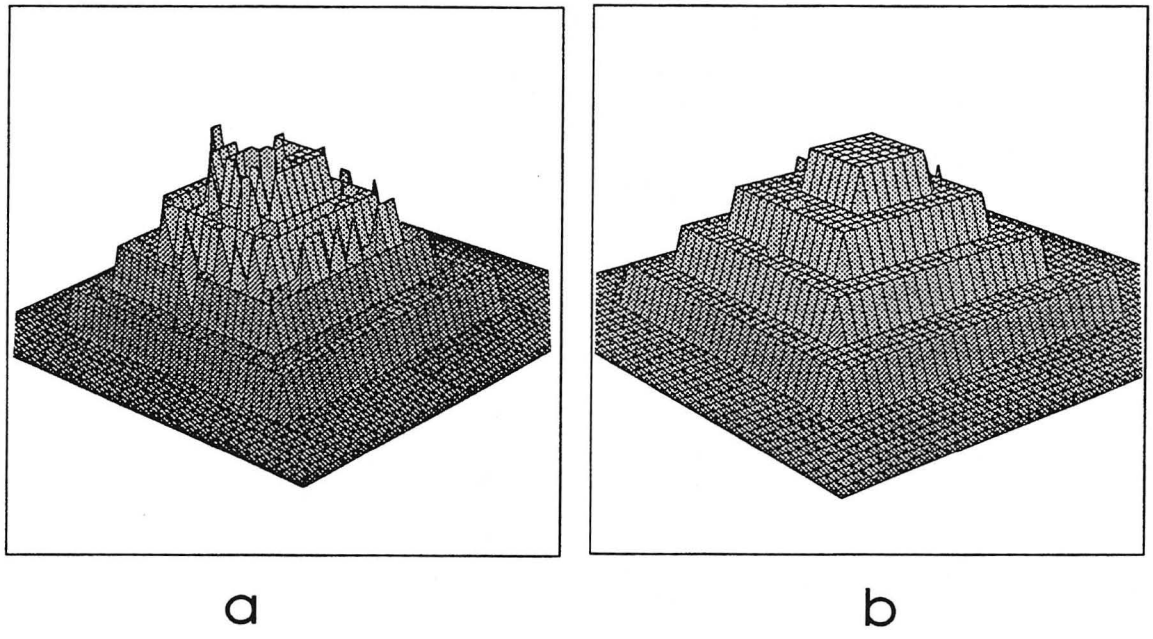


Figure 6: Recovered surfaces after 500 and 20000 iterations of the algorithm.

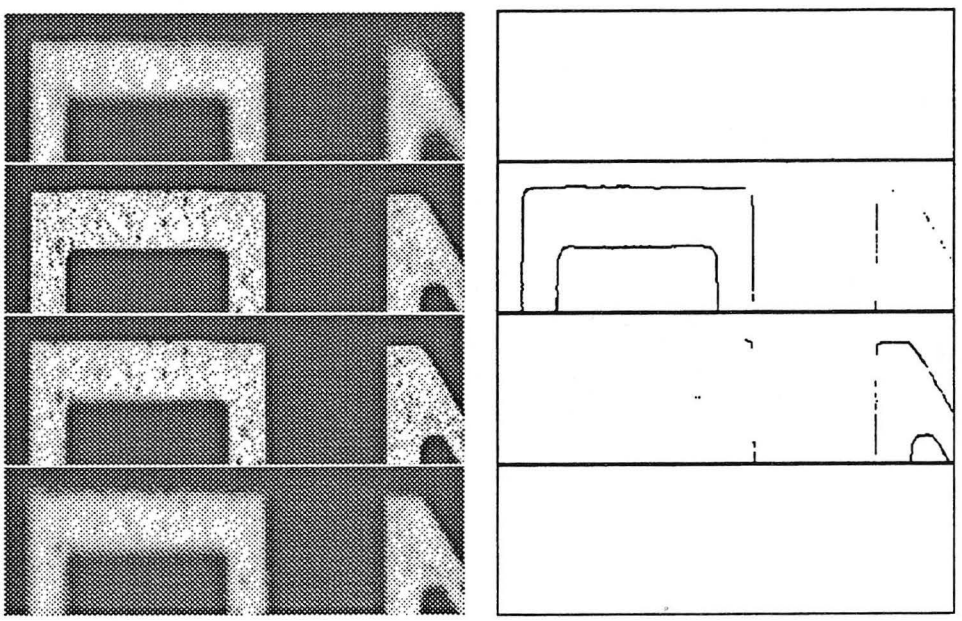


Figure 7: Depth estimation using an edge-based method.

LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
TECHNICAL INFORMATION DEPARTMENT
BERKELEY, CALIFORNIA 94720