Sub-pixel Reconstruction of a Variable Albedo Lambertian Surface

H. Shekarforoush, M. Berthod, J. Zerubia
INRIA - 2004 Route des Lucioles,
06902 Sophia Antipolis Cedex,
France
email: hshekar@sophia.inri.fr
Tel: (+33) 93 65 78 57
Fax: (+33) 93 65 76 43

Abstract

Using a probabilistic interpretation of an n dimensional extension of Papoulis's Generalized Sampling Theorem, an iterative algorithm has been devised for 3D reconstruction of a Lambertian surface at sub-pixel accuracy. The problem has been formulated as an optimization one in a Bayesian framework. The latter allows for introducing *a priori* information on the solution, using Markov Random Fields (MRF). The estimated 3D features of the surface are the albedo and the height which are obtained simultaneously using a set of low resolution images.

keywords: 3D Super resolution, Generalized Sampling Expansion, Low level image processing, Markov Random Fields (MRF).

1 Introduction

Most super-resolution algorithms proposed in the literature are confined to 2D applications and are concerned with the direct problem of merging low resolution 2D data on a finer grid [8][10][11]. In [4][17] a 3D version was proposed where the high resolution albedo of a Lambertian surface was estimated with the knowledge of high resolution height and vice versa. In this paper, the idea has been extended to the inverse problem of simultaneous reconstruction of albedo and height. The algorithm has been rigorously derived from the extension of Papoulis's Generalized Sampling Theorem [13] to nD cases.

The problem has been formulated using an optimization approach in a probabilistic framework, with Markov Random fields (MRF) [2] modeling the *a priori* knowledge. A generalized simplex algorithm [16] has been proposed for optimizing a cost function iteratively, giving sub-pixel 3D information on the observed surface. The simplex algorithm is initialized with a set of low resolution data.

Tests have been carried out on both synthetic and aerial or satellite images. The algorithm can be applied in the area of aerial and satellite image processing.

BMVC 1995 doi:10.5244/C.9.31

2 Preliminary Assumptions

We shall, henceforth, assume that our sensor is a pin-hole camera located far enough from the observed surface so that the projection of any surface point on the retina of the camera can be regarded as being orthographic. Moreover, camera parameters will be assumed to be known. As for the surface, we shall, hereafter, sample it into a two dimensional array and represent the surface albedo by $g(x, y)$ and the 3D graph of the surface by $z(x, y)$.

Image irradiance will be defined by the impulse response or the Point Spread Function (PSF) $H_w(x, y)$ of the imaging system. The remotely sensed view is usually affected by several impulse response functions: atmosphere h_a, the optics h_o and the sampling aperture h_s. Therefore the resultant PSF is given by $h_a * h_o * h_s$ (where $*$ denotes convolution). Each component introduces a type of smoothing operation and as a result, the net effect of these components in a complex remote sensing system can often be quite reasonably approximated by a 2D Gaussian pulse with a radius of gyration which depends on the resolution constraints of the imaging system [18]. In general the PSF can be approximated by:

$$H_{w_{xy}}(x_1, y_1) = Kexp\left(\frac{1}{2(1-\alpha^2)}\left(\frac{x_1^2}{w_{x1}^2} - \frac{2\alpha}{w_{x1}w_{y1}}x_1y_1 + \frac{y_1^2}{w_{y1}^2}\right)\right) \tag{1}$$

where K is a constant depending on the gain of the system, w_{x1} and w_{y1} are the radii of gyration about the two axes and α is the coupling factor between two directions. Using the following rotational transformation of our reference frame [18]:

$$\begin{cases} x = x_1 cos\varphi - y_1 sin\varphi \\ y = x_1 sin\varphi + y_1 cos\varphi \\ \varphi = \frac{1}{2}tan^{-1}\left(\frac{2\alpha w_{x1}w_{y1}}{w_{x1}^2 - w_{y1}^2}\right) \\ w_x^2 = w_{x1}^2 cos^2\varphi - w_{y1}^2 sin^2\varphi \\ w_y^2 = w_{x1}^2 sin^2\varphi + w_{y1}^2 cos^2\varphi \end{cases} \tag{2}$$

we will obtain:

$$H_{w_{xy}}(x, y) = Kexp\left(-\frac{1}{2}\left(\frac{x^2}{w_x^2} + \frac{y^2}{w_y^2}\right)\right) \tag{3}$$

For the case where the gray levels are preserved by the system [18], we have $K = \frac{1}{2\pi w_x w_y}$. Therefore assuming a circulary-symmetric (isotropic) Gaussian function the radius of gyration about both axes will be equal and given by $w = \sigma$, where σ is the standard deviation of the kernel. Thus:

$$H_w(x, y) = \frac{1}{2\pi\sigma^2}exp(-\frac{x^2 + y^2}{2\sigma^2}) \tag{4}$$

Therefore, by including a PSF [1][15] in our image formation model, we distinguish our image irradiance equation from the one defined in shape from shading algorithms [9]. Thus, assuming shift invariance the image irradiance is given by:

$$I(k, l) = \sum_{(x,y)\in w} H_w(x - x_c, y - y_c)\, g(x, y)R(x, y) \tag{5}$$

where $I(k,l)$ is the image intensity at coordinates (k,l), $H_w(x-x_c, y-y_c)$ is the PSF shifted by (x_c, y_c), and $R(x,y)$ is the reflection function of the Lambertian surface, whose product with the albedo $g(x,y)$ yields the luminance energy at coordinates (x,y) in the sampled array. The reflection function is, in fact, a function of the normal to the surface (see [9] and [3] for details).

Camera coordinates (k,l) are related to the world coordinate frame by:

$$k = a_{11}x + a_{12}y + a_{13}z + a_{14} \tag{6}$$

$$l = a_{21}x + a_{22}y + a_{23}z + a_{24} \tag{7}$$

Assuming that camera motions from one low resolution frame to another are only composed of translation and rotation around the origin of the world frame [7]:

$$a_{11} = a_{22} = S_x cos\theta \tag{8}$$

$$-a_{12} = a_{21} = S_y sin\theta \tag{9}$$

where S_x and S_y are the ratios of the sampling rates of the desired high resolution images to those of the low resolution frames along x and y axes, respectively, and θ is the rotation angle of the camera around the origin of the world frame. Working on square images of widths d and D for the low resolution and the high resolution frames, respectively, and assuming uniform and equal sampling rates along both axes, we have: $S_x = S_y = \frac{d}{D}$. As for other camera parameters, we have set $a_{13} = a_{23} = 0$ and a_{14} and a_{24} depend on the translation of the camera along the two axes. The latter should be in such a way that sub-pixel overlap is obtained between pixels of any two observed images in the sequence.

3 A Probabilistic view of Generalized Sampling Theorem

We can, now, formulate the problem as that of solving equation (5) which is a special case of the image irradiance equation [9][3]. To this end we need to recall some definitions as well as the Generalized sampling theorem in n dimensions:

Unless otherwise mentioned, all functions are n-dimensional vector functions and $f(\vec{t}) \leftrightarrow \mathcal{F}(\vec{\omega})$ denote a Fourier transform pair.

Definition: *A finite energy function $f(\vec{t})$ [14] is said to be σ-band limited if its Fourier transform $\mathcal{F}(\vec{\omega}) = 0$ outside the finite size hypercube $\mid \omega_i \mid \geq \sigma_i$, $i = 1 \ldots n$.*

Theorem: *nD Generalised Sampling Theorem*
We apply a σ-band limited function $f(\vec{t})$ as a common input to m independent linear shift invariant systems with transfer functions $\mathcal{H}_1(\vec{\omega}) \ldots \mathcal{H}_m(\vec{\omega})$. The resulting outputs are:

$$\phi_r(\vec{t}) = \frac{1}{(2\pi)^n} \int_{-\sigma_1}^{\sigma_1} \ldots \int_{-\sigma_n}^{\sigma_n} \mathcal{F}(\vec{\omega}) \mathcal{H}_r(\vec{\omega}) exp(j\vec{\omega}^T \vec{t}) d\omega_1 \ldots d\omega_n \tag{10}$$

where $r = 1 \ldots m$, $\omega_1 \ldots \omega_n$ *are the components of* $\vec{\omega}$ *and* T *denotes the transposition. Next, we sample these outputs at* $\frac{1}{m}$*th of the Nyquist rate along each dimension, ie. with a sampling matrix* S *whose diagonal terms are* $\frac{m\pi}{\sigma_i}$*:*

$$S = [s_{ab}] \ , \ s_{ab} = \frac{m\pi}{\sigma_i} \ if \ a = b = i \tag{11}$$

It can be shown that [6] *:*

$$f(\vec{t}) = \sum_{r=1}^{m} \sum_{k_1=-\infty}^{\infty} \ldots \sum_{k_n=-\infty}^{\infty} \phi_r(S\vec{k}) y_r(\vec{t} - S\vec{k}) \tag{12}$$

where $\vec{k} = [k_1 \ldots k_n]$ *is an integer valued vector and:*

$$y_r(\vec{t}) = \frac{|\ S\ |}{(2\pi)^n} \int_{-\sigma_1}^{\sigma_1+c_1} \ldots \int_{-\sigma_n}^{\sigma_n+c_n} Y_r(\vec{\omega}, \vec{t}) exp(j\vec{\omega}^T \vec{t}) d\omega_1 \ldots d\omega_n \tag{13}$$

$Y_r(\vec{\omega}, \vec{t})$ *are given by the following set of simultaneous equations:*

$$\sum_{r=1}^{m} \mathcal{H}_r(\vec{\omega} + (\ell - 1)\vec{c}) Y_r(\vec{\omega}, \vec{t}) = exp(j(\ell - 1)\vec{c}^T \vec{t}) \ , \ \ell = 1 \ldots m \tag{14}$$

where $\vec{c} = [\frac{2\sigma_1}{m} \ldots \frac{2\sigma_n}{m}]$ *and* $-\sigma_i \leq \omega_i \leq -\sigma_i + \frac{2\sigma_i}{m}$.

Note that the n-dimensional expansion in (12) is valid iff the sampling matrix S is non-singular. Herein, we will assume that the matrix S is diagonal, ie. we will assume a rectangular sampling. This, merely, implies that the sampling is performed independently along each dimension, in which case, S will be automatically non-singular, since the sampling density $|\ S\ | = \prod_{i=1}^{n} \frac{m\pi}{\sigma_i} \neq 0$.

Now, let the number of available low resolution frames be $\ell < m$ then by simply applying the principle of superposition, one can attempt to reconstruct a sample of $f(\vec{t})$ at the resolution $\frac{\ell}{m}$th of the Nyquist rate by minimizing the following error:

$$\epsilon^2 = \mathcal{E}\{(f_\ell(\vec{t}) - \sum_{r=1}^{\ell} \sum_{k_1=-\infty}^{\infty} \ldots \sum_{k_n=-\infty}^{\infty} \phi_r(S\vec{k}) y_r(\vec{t} - S\vec{k}))^2\} \tag{15}$$

where $f_\ell(\vec{t})$ is a sample of $f(\vec{t})$ at $\frac{\ell}{m}$th of the Nyquist rate along each dimension and $\mathcal{E}$ represents the expected value. Alternatively, we can minimize the error after sampling. Therefore if $\hat{\phi}_r(S\vec{k})$ denotes an estimate of $\phi_r(S\vec{k})$:

$$\epsilon_\phi^2 = \mathcal{E}\{\sum_{r=1}^{\ell} (\phi_r(S\vec{k}) - \hat{\phi}_r(S\vec{k}))^2\} \tag{16}$$

An interesting situation arises when $f_\ell(\vec{t})$ is expressed in terms of two (or more) variables (eg. albedo and height): $f_\ell(\vec{t}) = f_\ell(\vec{s}_1(\vec{t}), s_2(\vec{t}))$. We can then seek high resolution information for $s_1(\vec{t})$ and $s_2(\vec{t})$. This is obviously an inverse problem

and can be tackled using regularization or equivalently MRF's.

The n-dimensional extension of Papoulis' sampling expansion is an ideal tool for our purpose. In this context, in a sequence of low resolution images, each frame can be assumed to contain the recuring samples of a nonuniform sampling sequence obtained by applying a common input function to a set of linear shift invariant systems. We, obviously, need to register the recuring samples using registration algorithms such as [7][19].

Let g and z denote the vectors of unknown variables, ie. the vectors of the albedo $g_{(x,y)}$ and the height $z_{(x,y)}$ on the super-resolution grid at $\frac{\ell}{m}$th of the Nyquist rate. Let also I denote the vector of all observed pixel values in our low resolution frames. Then using Bayes law on probability distributions and assuming that g and z are independent:

$$p(g,z \mid I) = K\ p(I \mid g,z)\ p(g)p(z)\ \ ,\ \text{where } K \text{ is a constant} \tag{17}$$

Applying the Hammersley-Clifford theorem [12] to $p(g)$ and $p(z)$ and assuming a Gaussian distribution for $p(I \mid g,z)$ we can calculate the Maximum a Posteriori (MAP) estimator as follows:

$$E(g,z) = -ln\ p((g,z) \mid I) = e^T C_e^{-1} e + u_g^T C_g^{-1} u_g + u_z^T C_z^{-1} u_z + constant \tag{18}$$

Using a membrane model for $p(g)$ and $p(z)$ [5], and assuming that $e^T C_e^{-1} e = \epsilon_\phi^2$, given by (16), we can write the following cost function at any point (x,y) of our sampled array:

$$E_{(x,y)} = \sum_{\ell} \sum_{(k',l') \in \nu_{(k,l)}} \frac{\left(I^{\ell}_{(k',l')} - \hat{I}^{\ell}_{(k',l')}\right)^2}{2\sigma_e^2} + \sum_{(x',y') \in \nu_{(x,y)}} \frac{\left(\hat{g}_{(x,y)} - \hat{g}_{(x',y')}\right)^2}{2\sigma_g^2} + \sum_{(x',y') \in \nu_{(x,y)}} \frac{\left(\hat{z}_{(x,y)} - \hat{z}_{(x',y')}\right)^2}{2\sigma_z^2} \tag{19}$$

where $\nu_{(k,l)}$ is the set of all pixels in the neighbourhood of (k,l) whose intensities are affected by the irradiance at (x,y), $\nu_{(x,y)}$ is the neighbourhood structure depending on the order of the MRF associated with the estimates $\hat{g}$ and $\hat{z}$, σ_e^2, σ_g^2 and σ_z^2 are the variances of the error vector, g and z, respectively. Therefore, the MAP estimator amounts to minimizing (19) for which we have employed a generalized simplex algorithm [16]. Equation (19) has been obtained by approximating the covariance matrices as being diagonal with constant terms on the diagonal.

4 Proposed Algorithm

The algorithm is iterative and is shown schematically in Figure 1. It proceeds as follows: we first initialize it using a set of low resolution images. The imaging process is then simulated to create a set of low resolution estimates of sensor observations. The objective function in (19) characterizing the estimation error and the Gibbs energy of our MRF is minimized simultaneously with respect to the albedo and the surface height. The algorithm is repeated iteratively until no more

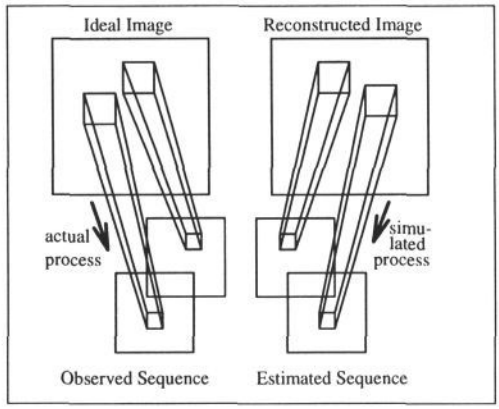

Figure 1: Schematic diagram of the algorithm

reduction in the objective function is achieved for a preset number of successive iterations.

The major obstacle is the non-convexity of the objective function. Different optimization algorithms have been proposed in the litterature for minimising a non-convex function which have been briefly compared in [3]. In this article, we have considered the use of a generalized simplex algorithm which is based on the idea of constructing a non-local linear approximation of a function by updating an n-dimensional simplex in $\mathbb{R}^n$ (see [16] for details). We have initialized the simplex algorithm using a set of low resolution images and a noisy/sparse low resolution height map.

5 Discussion

By interpreting the Generalized Sampling Theorem in probabilistic terms we have shown that it is possible to reconstruct, simultaneously, sub-pixel information on the albedo and the height of a Lambertian surface, provided that a sequence of independent images have been given at a lower resolution.

The method exploits the non-redundancy of information in a sequence of image frames with interframe sub-pixel overlap, to enhance the spacial resolution by data fusion. It is, indeed quite reasonable to assume that a set of images taken from a surface, almost always contain non-redundant information, since the chance of exact correspondence is extremly low.

Two main problems were encountered when testing the algorithm: one is the trade-off between the regularization term and the error term in the cost function, which proved to be not so easy when dealing with real images. For the synthetic

image the problem was almost non-existent due to the fact that the test images did not exhibit discontinuities. This problem could, therefore, be handled by taking into acount a discontinuity field and hence by including extra terms in the cost function. The second problem was the convergence of the simplex algorithm which slows down as we get closer to the optimum point. A possible solution would be to switch to a gradient method in the neighbourhood of the solution where the cost function exhibits convexity. The simplex algorithm, however, is robust to noise and is very efficient far from the solution.

Since the problem has been formulated as a constrained optimization one, further extensions of the proposed method can be considered along the same direction. Therefore, we might consider optimizing with respect to parameters of the imaging system which could yield for example the interframe registration, optimal light source direction, etc.

Experimental results follow.

6 Experimental Results

In this section we present some tests on both synthetic and real data. Signal to Noise Ratio (SNR) is used for measuring the quality of results. Original images were 50×50 and 100×100 for low resolution and high resolution, respectively.

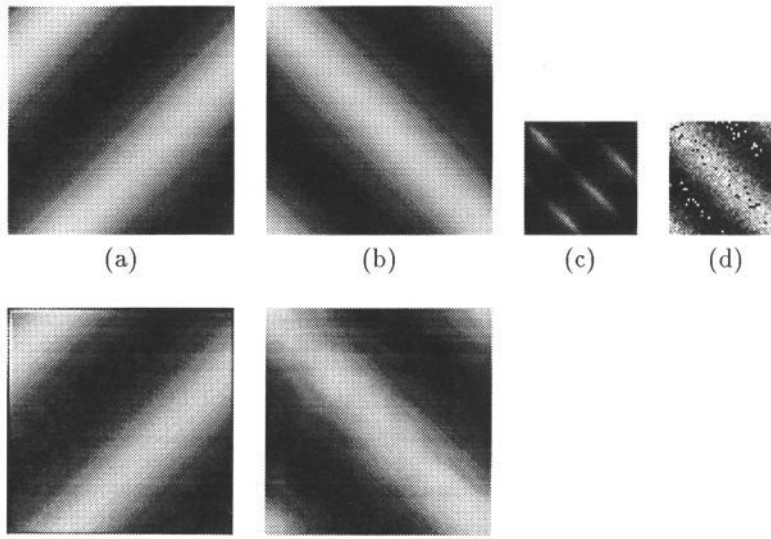

Figure 2: (a), (b) albedo & height, (c) one of the simulated low resolution camera images, (d) a noise corrupted low resolution height SNR=5 dB, (e) reconstructed albedo SNR=34 dB, (f) reconstructed height SNR=27 dB

Below are further results on real data. The aerial image has been assumed to represent the albedo (Figure 3(a)). Using the high resolution height map a sequence of shifted and degraded low resolution images have been produced according to the Lambertian model, described above (note that only one low resolution frame has been shown below). The reconstructed images are given in Figure 3(e) and 3(f)

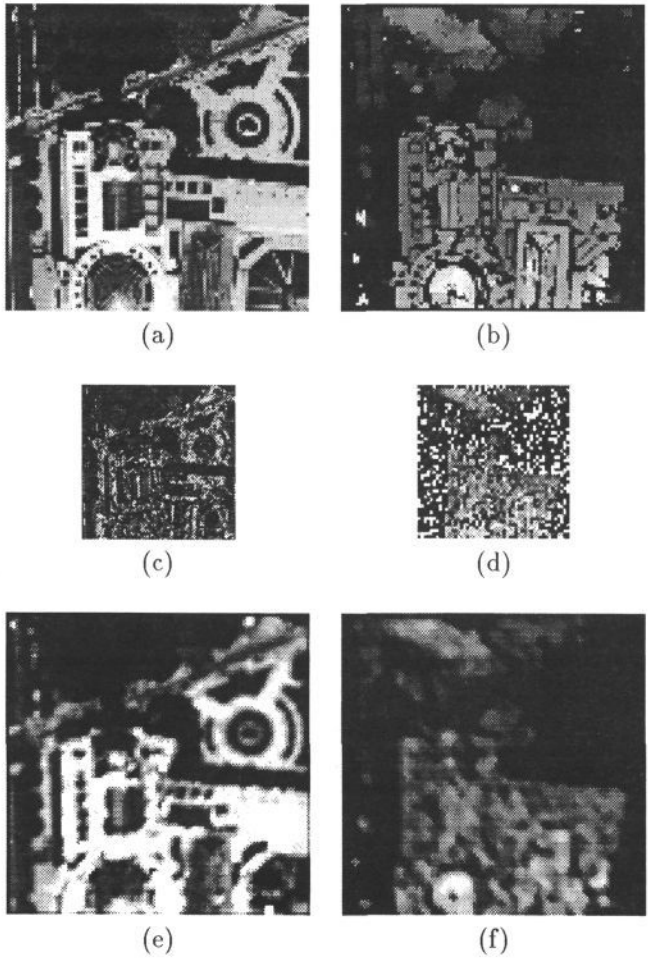

Figure 3: (a) albedo, (b) height, (c) one of the low resolution camera images, (d) a noise corrupted low resolution height SNR=5 db, (e) reconstructed albedo SNR=25 dB, (f) reconstructed height SNR=22 dB

Here are further results on an aerial image. Same comments apply:

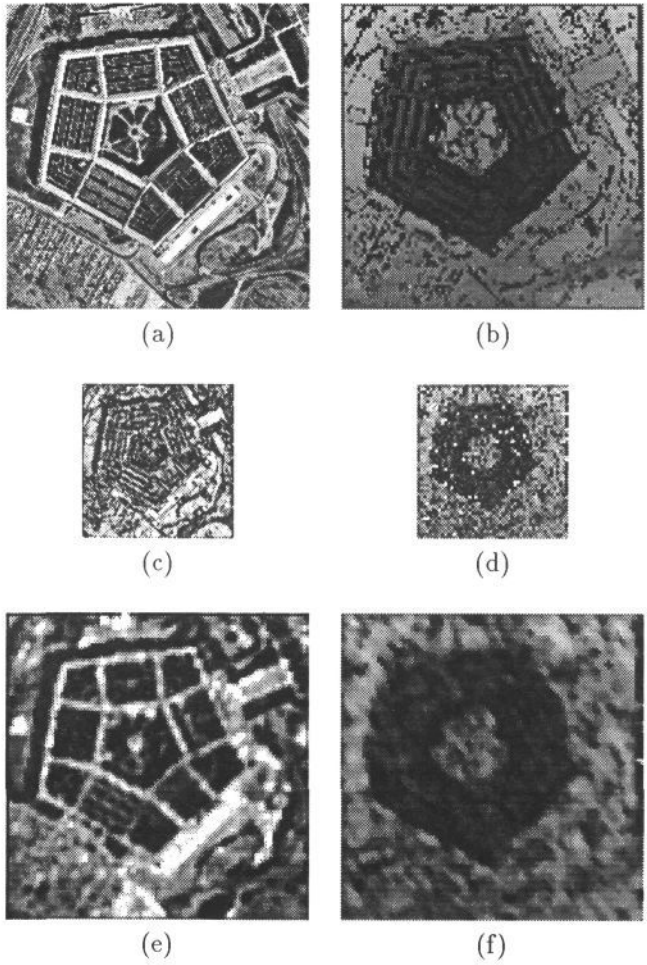

Figure 4: (a) albedo, (b) height, (c) one of the simulated low resolution camera images, (d) a noise corrupted low resolution height SNR=5 db, (e) reconstructed albedo SNR=26 dB, (f) reconstructed height SNR=24 dB

References

[1] H. C. Andrews and B. R. Hunt. *Digital Image Restoration.* Prentice-Hall Inc., 1st edition, 1977.

[2] R. Azencott. Image analysis and markov fields. In *Int. Conf. Ind. & Appl. Math*, Paris, 1987. SIAM.

[3] M. Berthod, H. Shekarforoush, M. Werman, and J. Zerubia. Reconstruction of high resolution 3d visual information, 1993. INRIA research report, France.

[4] M. Berthod, H. Shekarforoush, M. Werman, and J. Zerubia. Reconstruction of high resolution 3d visual information. In *Proc. CVPR*, pages 654–657, Seattle, Washington, 1994.

[5] A. Blake and A. Zisserman. *Visual Reconstruction.* MIT Press, 1987.

[6] K. F. Cheung and R. J. Marks. Papoulis' generalization of the sampling theorem in higher dimensions and its applications to sample density reduction. In *Proc. Int. Conf. on circuits and systems*, Nanjing, China, 1989.

[7] E. De Castro and C. Morandi. Registration of translated and rotated images using finite fourier transforms. In *IEEE PAMI*, volume 9, No 5, pages 700–703, 1987.

[8] D. Gross. Super-resolution from sub-pixel shifted pictures. Master's thesis, Tel-Aviv University, Oct. 1986.

[9] B. K. P. Horn. Understanding image intensities. *Artificial Intelligence*, 8(2):201–231, 1977.

[10] M. Irani and S. Peleg. Super-resolution from image sequences. Technical Report 89-7, The Hebrew University of Jerusalem, June 1989.

[11] D. Keren, S. Peleg, and R. Brada. Image sequence enhancement using sub-pixel displacement. In *Proc. CVPR*, pages 742–746, Ann Arbor, Michigan, June 1988.

[12] J. Moussouris. Gibbs and markov random systems with constraints. *Journal of Statistical Physics*, 10(1), 1974.

[13] A. Papoulis. Generalized sampling expansion. In *IEEE Trans. on Circuits and Systems*, volume 24, No 11, Nov. 1977.

[14] A. Papoulis. *Signal Analysis.* McGraw-Hill Book Company, 1977.

[15] A. Rosenfeld and A. C. Kak. *Digital Picture Processing*, volume 1 and 2. Academic Press Inc., 1982.

[16] H. Shekarforoush, M. Berthod, and J. Zerubia. Direct search generalized simplex algorithm for optimizing non-linear functions. research report N^o 2535, INRIA - France, 1995.

[17] H. Shekarforoush, M. Berthod, J. Zerubia, and M. Werman. Sub-pixel bayesian estimation of albedo and height. *International Journal of Computer Vision*, to appear.

[18] D. S. Simonett and F. T. Ulaby, editors. *Manual of Remote Sensing*, volume 1. American Society of Photogrammetry, second edition, 1983.

[19] R. Tsai and T. Huang. Multiframe image restoration and registration. *Adv. Comp. Vis. Im. Proc.*, 1, 1984.