Reconstruction Of High Resolution 3D Visual Information

*M. Berthod *H. Shekarforoush **M. Werman *J. Zerubia

*INRIA - 2004 Route des Lucioles
06902 Sophia Antipolis Cedex
France
email: berthod@sophia.inria.fr

**Institute Of Computer Science
Hebrew University of Jerusalem
91904 Jerusalem - Israel
email: werman@cs.huji.ac.il

Abstract

Given a set of low resolution camera images, it is possible to reconstruct high resolution luminance and depth information, specially if the relative displacements of the image frames are known. We have proposed iterative algorithms for recovering high resolution albedo and depth maps that require no a priori knowledge of the scene, and therefore do not depend on other methods, as regards boundary and initial conditions. The problem of surface reconstruction has been formulated as one of Expectation Maximization (EM) and has been tackled in a probabilistic framework using Markov Random Fields (MRF) [1][3]. *As for the depth map, our method is directly recovering surface heights without refering to surface orientations, while increasing the resolution by camera jittering* [2]. *Conventional statistical models have been coupled with geometrical techniques to construct a general model of the world and the imaging process.*

1 Introduction

Image resolution depends on the physical characteristics of the sensor: the optics, the density and the spatial response of the sensing elements. Increasing the resolution by sensor modification may not always be an available option. An increase in the sampling rate can, however, be achieved by obtaining more samples of the same scene from a sequence of displaced images. In this paper we have proposed an algorithm for reconstructing high resolution images given a set of low resolution observations. Earlier works on super-resolution such as [6],[7],[8] and [10] provide a good background on the subject. In our approach, however, Markov Random Field's (MRF) [1] [3] have been used for modeling the properties of the processed images. We specifically have used the Expectation Maximization (EM) algorithm for reconstructing the optimal solution. Tests have been carried out on both synthetic and real (aerial) images.

2 Image Formation

We have, primarily, assumed that our low resolution images are formed by projecting a Lambertian surface of varying albedo onto the retina of a distant camera (orthographic projection) while the surface is being illuminated by a distant light source. It is well known that various possible combinations of the albedo and the surface shape can lead to the same light intensity. Therefore, recovering scene characteristics and structure is an ill-posed problem and needs to be constrained in order to converge to a unique solution. Given the above assumptions the reflectance map of the surface is given by:

$$R = \frac{\vec{N}.\vec{L} \times \vec{N}.\vec{S}}{\sqrt{p^2+q^2+1}} \qquad (1)$$

where $\vec{N} = (p, q, -1)$ is the normal vector to the surface with $p = \frac{\partial z}{\partial x}$ and $q = \frac{\partial z}{\partial y}$ (first partial derivatives of the altitude z). $\vec{L} = (\alpha_L, \beta_L, \gamma_L)$ and $\vec{S} = (\alpha_S, \beta_S, \gamma_S)$ are the unit normals along the light source and the viewing direction respectively. Since in the present case, digitization occurs on the surface rather than in the image plane, the number of elementary patches in the receptive field of the camera will be inversely proportional to the cosine of the reflection angle and therefore we have included $\vec{N}.\vec{S}$ in (1). Note that . is refering to the dot product of two vectors. Now assuming that the Point Spread Function(PSF) of the camera $H_{\bar{k}}$ can be modeled by a gaussian blurring kernel, we can obtain the image intensity at any

1063-6919/94 $3.00 © 1994 IEEE

point by convolution:

$$I_{\vec{k}} = \sum_{\vec{i} \in w} H((u-k_1),(v-k_2)) g_{\vec{i}} R_{\vec{i}} \qquad (2)$$

where $g_{\vec{i}}$ is the surface albedo, $R_{\vec{i}}$ is the reflectance given by (1), w is the support of the PSF and $\vec{i} = (i_1, i_2)$ and $\vec{k} = (k_1, k_2)$ are the surface and shifted image coordinates, respectively. We have used a generalized camera coordinate system:

$$u = a_{11}x + a_{12}y + a_{13}z + a_{14} \qquad (3)$$

$$v = a_{21}x + a_{22}y + a_{23}z + a_{24} \qquad (4)$$

where (x, y, z) are the world coordinates, $a_{13} = a_{23} = 0$, $a_{11} = a_{22} = S_u cos\tau$, $a_{12} = a_{21} = S_v sin\tau$ with τ the tilt angle of the camera and S_u and S_v the sampling rates along x and y axes respectively. a_{14} and a_{24} are given by the translations of the camera along x and y axes. These translational parameters should be in such way that the overlap between two image frames is at subpixel accuracy and hence image frames do not contain redundant information [8][10].

3 An Algorithm For Super-Resolution

We will now consider the problem of refining photometric and structural information extracted from a sequence of intensity images. We will assume that the imaging process is well known and that a sequence of low resolution observations are available with pixels on each frame registered at subpixel accuracy by camera jittering [2]. Initializing the super-resolution image to an arbitrary estimate, we will then simulate the imaging process to obtain a set of estimated low resolution pictures. Comparing these with the observed sequence of low resolution frames will allow us to minimize a penalty function iteratively and hence update the initial guess until a stop criterion is met(see Figure 1).

The penalty function contains two terms: the estimation error and a smoothness constraint to ensure a stable convergence to a unique solution. If the brightness error term is set aside then the solution degenerates to pure interpolation. However, in practice, the solution is regularized by including some terms which reflect the constraints together with the estimation error. The degree of regularization required can depend on the actual image being processed, and hence, Some authors [5][9] advocate empirical tuning at each iteration. We will use the gradient of the penalty function to specify pixel value modifications for the albedo. For

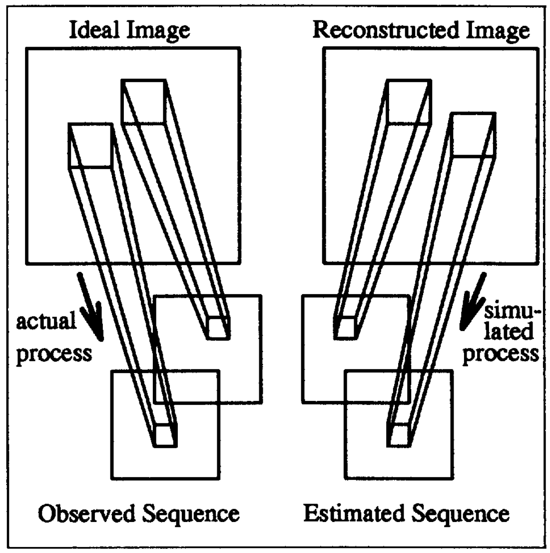

Figure 1: Schematic diagram of the iterative super-resolution algorithm

z, however, the penalty function is minimized locally while updating pixels in small overlapping windows by fixed small values at each iteration.

4 A Probabilistic Framework

At this stage we will sample both g and z (ie. the albedo and the altitude respectively) on a super-resolution grid size and impose a probabilistic model on them. We will, therefore, try to solve a classical Expectation Maximization (EM) problem, directly in a finite dimensional space, but possibly with real values. We will use Maximum A Posteriori (MAP) [1] for estimating the most probable configuration of the reconstructed image. The class of stochastic models (ie. random fields) that meet our criteria is the Markov Random Fields (MRF) on finite lattices [1][3][4]. Let (g, z) depict the pair of independent random fields to which our observed images belong. Determination of the MAP estimate requires the knowledge of constructing the a posteriori density function given a sequence of n observations: $(I_1 ... I_n)$.

Using Bayes law gives:

$$p((g,z)/I_1...I_n) = Kp(I_1...I_n/(g,z))p(g,z) \qquad (5)$$

where K is a constant. Assuming that the noise process is independent from one view to another:

$$p(I_1 ... I_n/(g,z)) = \prod_n p(I_n/(g,z)) \tag{6}$$

Since the probability distribution of a MRF can be expressed in the form of a Gibbs distribution [3][4]:

$$p(g) = \frac{1}{Z} \exp\left(-\frac{1}{2} {U_g}^2 {C_{gg}}^{-1} T^{-1}\right) \tag{7}$$

$$\text{and :} \quad p(I_n/g) = \left(\frac{1}{2\pi C_{\epsilon\epsilon}}\right)^{1/2} \exp\left(-\frac{1}{2} F^2 {C_{\epsilon\epsilon}}^{-1}\right) \tag{8}$$

where Z is a normalizing constant called the partition function, $U_g = \sqrt{\Sigma_{\vec{i}} \Sigma_{\vec{j} \in \nu_{\vec{i}}} (g_{\vec{i}} - g_{\vec{j}})^2}$, T is a temperature parameter set to 1, $F = \sqrt{\Sigma_n \Sigma_{\vec{k}} \Sigma_{\vec{l} \in w_k} (I_{n,\vec{l}} - \hat{I}_{n,\vec{l}})^2}$, w_k are the neighbouring pixels of $\vec{k}$ whose PSF include $\vec{i}$, and C_{gg} and $C_{\epsilon\epsilon}$ are the covariances of g and the noise process, respectively. Extending to the general case, including z, we can see that maximizing (5) will lead to minimizing the following energy function:

$$E = \sum_{\vec{i}} \left(\frac{\sum_{\vec{j} \in \nu_{\vec{i}}} (g_{\vec{i}} - g_{\vec{j}})^2}{2{\sigma_g}^2} + \frac{\sum_{\vec{j} \in \nu_{\vec{i}}} (z_{\vec{i}} - z_{\vec{j}})^2}{2{\sigma_z}^2} \right) + \sum_n \sum_{\vec{k}} \frac{\sum_{\vec{l} \in w_k} (I_{n,\vec{l}} - \hat{I}_{n,\vec{l}})^2}{2{\sigma_\epsilon}^2} \tag{9}$$

Partial derivatives of E with respect to g_i and z_i are:

$$\frac{\partial E}{\partial g_{\vec{i}}} = \sum_{\vec{i}} \frac{\sum_{\vec{j} \in \nu_{\vec{i}}} (g_{\vec{i}} - g_{\vec{j}})}{{\sigma_g}^2} + \sum_n \sum_{\vec{k}} \frac{\sum_{\vec{l} \in w_k} (I_{n,\vec{l}} - \hat{I}_{n,\vec{l}}) H_{\vec{k}} R_{\vec{i}}}{{\sigma_\epsilon}^2} \tag{10}$$

$$\frac{\partial E}{\partial z_{\vec{i}}} = \sum_{\vec{i}} \Bigg(\frac{\sum_{\vec{j} \in \nu_{\vec{i}}} (z_{\vec{i}} - z_{\vec{j}})}{{\sigma_z}^2} + \sum_n \sum_{\vec{k}} \frac{\sum_{\vec{l} \in w_k} (I_{n,\vec{l}} - \hat{I}_{n,\vec{l}})}{{\sigma_\epsilon}^2} \sum_{\vec{j} \in \nu_{\vec{i}}} g_{\vec{i}} H_{\vec{k}} \frac{\partial R_{\vec{j}}}{\partial z_{\vec{i}}} \Bigg) \tag{11}$$

$$\text{where :} \quad \frac{\partial R_{\vec{j}}}{\partial z_{\vec{i}}} = -(\alpha_L P_{ji} + \beta_L Q_{ji}) \frac{\vec{N}.\vec{S}}{\sqrt{p_{\vec{i}}^2 + q_{\vec{i}}^2 + 1}} - (\alpha_S P_{ji} + \beta_S Q_{ji}) \frac{\vec{N}.\vec{L}}{\sqrt{p_{\vec{i}}^2 + q_{\vec{i}}^2 + 1}} - (p_{\vec{i}} P_{ji} + q_{\vec{i}} Q_{ji}) \frac{\vec{N}.\vec{L} \times \vec{N}.\vec{S}}{(p_{\vec{i}}^2 + q_{\vec{i}}^2 + 1)^{3/2}} \tag{12}$$

($p_{\vec{i}} = \sum_{\vec{j} \in w_p} P_{ij} z_j$, $q_{\vec{i}} = \sum_{\vec{j} \in w_p} Q_{ij} z_j$ and w_p =support of the Prewitt filter)

The minimum value of E is obtained when these partial derivatives vanish. The annealing process, however, can be problematic when E is not convex as in the case for z. The method used for g comprises of an iterative gradient descent algorithm using Successive Over Relaxation(SOR). As for z we have tackled the problem by minimizing E locally under the positivity constraint for z.

5 Experimental Results

Figure 2(a) and 2(b) show the synthetic high resolution albedo and altitude, respectively. Figure 2(c) is one of the 9 simulated low resolution images obtained by the camera. Assuming that the high resolution altitude is known, the albedo has been recovered at high resolution(Figure 2(d) to 2(f)).

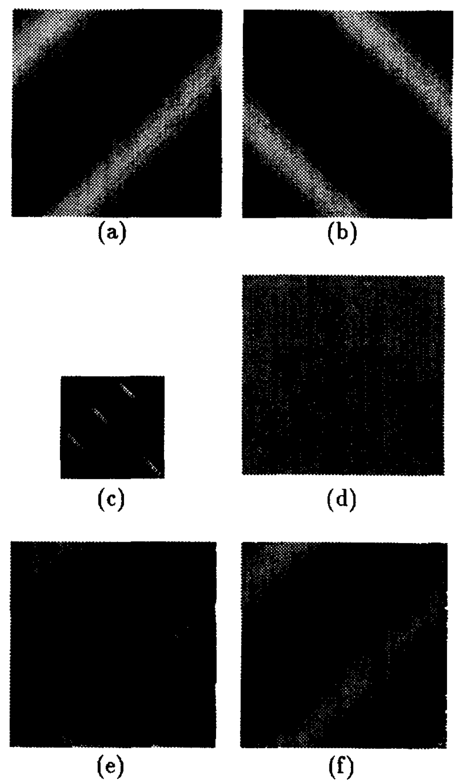

Figure 2: (a) Ideal albedo, (b) ideal altitude, (c) one of the 9 low resolution simulated camera images , (d) Initialization, (e) and (f) reconstructed albedo after the 1st and the 4th iteration.

Figure 3 shows an aerial image of Paris which we have used as our ideal albedo. Using this ideal albedo and the altitude map, we have simulated a set of nine low resolution observed camera images. We have run the algorithm first for recovering the albedo using these observed images and assuming that the high resolution altitude map is known. Next, we carried out the same procedure for the altitude, with the knowledge of the high resolution albedo while minimizing the energy function locally(see [2] for details):

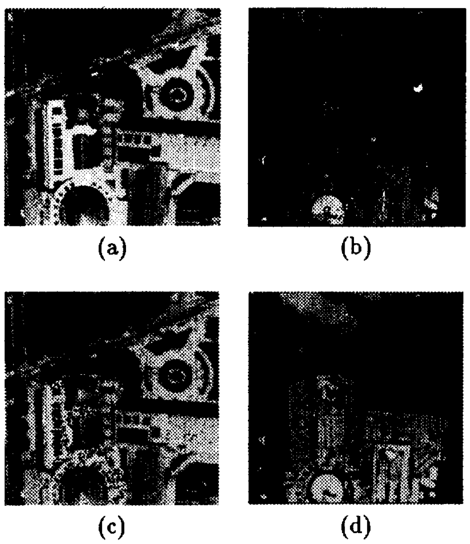

Figure 3: (a) ideal albedo, (b) ideal altitude, (c) reconstructed albedo after 4 iterations, (d) reconstructed altitude after 10 iterations

6 Concluding Remarks

We presented herein an algorithm for recovering 3D information from a set of low resolution intensity images. It is important, however, to have interframe subpixel displacements [7][8][10] so that pixel information incorporated in our reconstructed high resolution images are determined by an oversampled surface. The advantage of the technique to pure interpolation is therefore the fact that the reconstructed image embodies interpixel values obtained directly from the actual scene using an error criterion. Also, change of spatial resolution using pure interpolation is usually obtained by trading the gray level quantization[2].

In the case of g, the degree of relaxation can highly affect the speed of convergence and the quality of final results. Discretization prior to convolution can produce problems around the borders of the reconstructed images. These can be handled by means of extrapolation methods. Note that our method can be used for direct recovery of height from shading [9] with varying albedo. Results for both g and z are indeed very much promising.

Acknowledgements

The financial support of this project was partially provided by AFIRST.

References

[1] R. Azencott, "Image Analysis and Markov Fields", *Proc. of Int. Conf. Ind. & Appl. Math*, SIAM, Paris, 1987.

[2] M. Berthod, H. Shekarforoush, M. Werman, J. Zerubia, "Reconstruction of High Resolution 3D Visual Information" *INRIA Research Report N° 2142*, 1993.

[3] J. E. Besag, "On the Statistical Analysis of Dirty Pictures" *Jl. Roy. Statis. Soc.*, B.48, 1986.

[4] S. Geman, D. Geman, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images", *IEEE Trans. PAMI 6*, 1984.

[5] B. K. P. Horn, "Height and Gradient From Shading", *Proc. of Image Understanding Workshop*, DARPA, 1989.

[6] T. Huang,R. Tsai, "Multiframe Image Restoration and Registration", *Advances in Computer Vision and Image Processing*, vol. 1, pp 317-339, JAI Press Inc, 1984.

[7] M. Irani, S. Peleg, "Super-resolution from Image Sequences", *Technical Report 89-7, Dept. of Comp. Sci.*, The Hebrew Uni. of Jerusalem, 1989.

[8] D. Keren, S. Peleg, R. Brada, "Image Sequence Enhancement Using Sub-pixel Displacement", *Proc. of CVPR*, pp 742-746, 1988.

[9] Y. G. Leclerc, A. F. Bobick, "The direct Computation of Height from shading", *Proc. of CVPR*, 1991.

[10] S. Peleg, D. Keren, L. Schweitzer, "Improving Image Resolution Using Subpixel Motion", *Pattern Recognition Letters*, 1987.