Massachusetts Institute of Technology

Artificial Intelligence Laboratory

A. I. Memo No. 744 August, 1983

Constructing A Depth Map from Images

Katsushi Ikeuchi

Abstract

This paper describes two methods for constructing a depth map from images. Each methods has two stages. First, one or more needle maps are determined using a pair of images. This process employs either the Marr-Poggio-Grimson stereo and shape-from-shading, or, instead, photometric stereo. Secondly, a depth map is constructed from the needle map or needle maps computed by the first stage. Both methods make use of an iterative relaxation method to obtain the final depth map.

Keywords: Shape from shading, Photometric Stereo, Surface reconstruction, Marr-Poggio-Grimson stereo, Boundary conditions, Depth map, Needle map, Relaxation.

Acknowledgment: This paper describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-80-C-0505 and in part by the Office of Naval Research under Office of Naval Research contract N00014-77-C-0389.

1. Introduction

This paper describes two methods for constructing a depth map from a pair of images. These new methods utilize shading information and depend on needle maps as intermediate representations of surface shape.

Binocular stereo provides depth information only on a sparse subset of the image plane, since it establishes correspondences between identifiable features in the left image and those in the right image. For example, the Marr-Poggio-Grimson stereo method matches zero crossing contours [Marr and Poggio 77], [Grimson 81]. Thus, this method provides depth information directly only along the zero crossing contours; an interpolation scheme is required to construct a continuous depth map over the whole region from this sparse information.

Some have proposed that the interpolated surface satisfy some smoothness criterion [Grimson 81]. In particular it might be the surface of least energy fitting the boundary conditions [Brady and Horn 83], [Horn 83]. These ideas were developed in the context of understanding of biological visual systems. The least energy surface is, however, only a guess. The real surface need not be the least energy surface.

Shading information provides constraints on surface orientation. This information may be used for construction of the surface. The real depth map can be obtained only by considering this shading information.

This paper explores methods for obtaining the exact surface shape using both depth information from a stereo system and surface orientation information from a shape-from-shading algorithm.

2. Depth Construction Schema Using one Needle Map

In this chapter we discuss a method for constructing a depth map from a pair of images using a needle map. The overall strategy is as follows:

1) Obtain depth information at the zero crossings from the Marr-Poggio-Grimson stereo algorithm,

2) Obtain surface orientation along the zero crossings by differentiation,

3) Obtain surface orientation over the whole region enclosed by the zero crossing contour using the Ikeuchi-Horn shape-from-shading algorithm,

4) Obtain depth information using the needle map computed above and the known depth information along the boundary, by means of an iterative reconstruction algorithm.

2.1. Marr-Poggio-Grimson Stereo Algorithm

The Marr-Poggio-Grimson stereo algorithm provides depth information along the zero crossing contours. Zero crossings are the places where the convolution of brightnesses with the difference of Gaussians (DOG) mask [Hildreth 80] becomes zero. See [Marr and Poggio 78], [Grimson 81] for further details.

2.2. Boundary Conditions for Shape-from-Shading

The iterative shape-from-shading algorithm requires boundary conditions. The surface orientation must be specified along a closed boundary. The Marr-Poggio-Gimson stereo algorithm provides depth information along the zero crossing boundaries. Derivatives of this depth information in turn provides surface orientation along the boundary, as we will show next.

Since the Marr-Poggio-Grimson stereo algorithm is based on zero crossing contours, the depth information is always determined along a closed curve. Let the curve be given by the vector $\mathbf{r}(s)$ where s is a parameter which varies continuously and monotonically along the curve (it could be arc length, for example).

We can determine a tangent vector $\mathbf{T}(s)$ at the point s on the curve, simply by differentiating,

$$\mathbf{T}(s) = \frac{d\mathbf{r}}{ds}.$$

We are interested in surface orientation, which can be represented by the surface normal. Since the curve lies on the surface of the object, the normal there must be perpendicular to the curve. Figure 1 shows the situation on the Gaussian sphere.

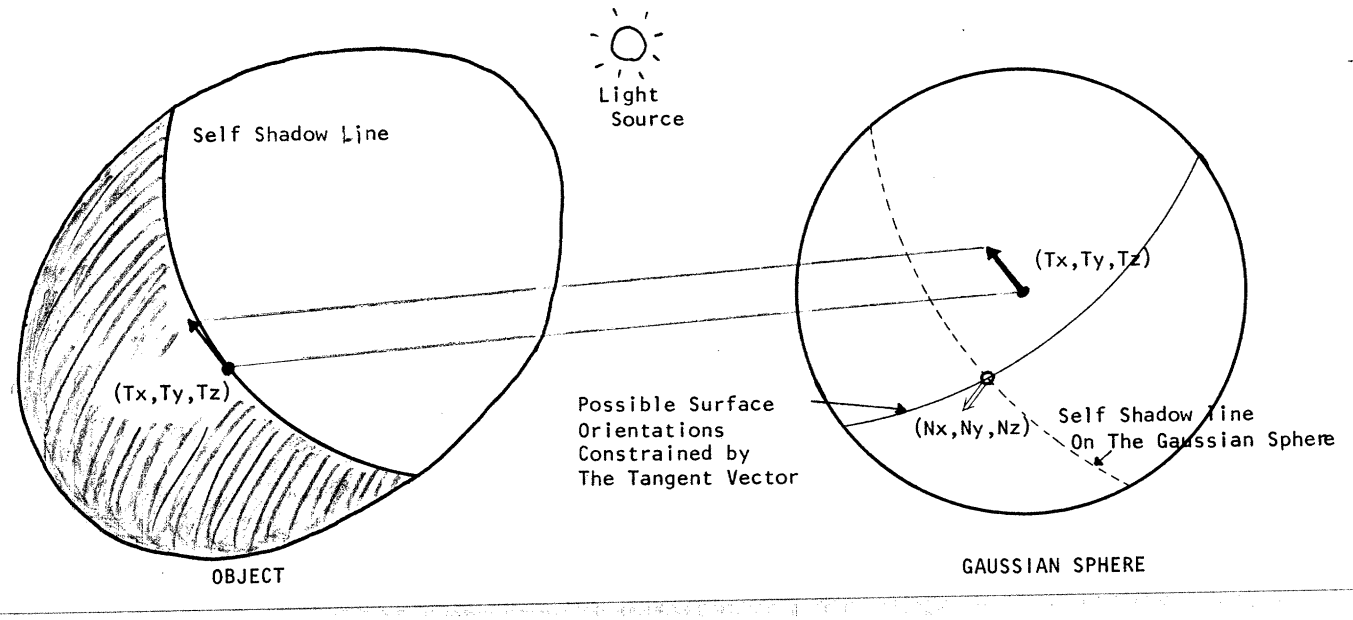

Figure 1. Two constraints on the Gaussian sphere define a unique surface orientation. One constraint comes from the known boundary curve, the other from the known brightness.

The bold line denotes the locus of possible surface normals N determined from a particular tangent vector T. Note that the possible surface orientations lie on a great circle on the Gaussian sphere. The axis of this great circle corresponds to the tangent vector, and we have $(\mathbf{N} \cdot \mathbf{T}) = 0$. This however is not sufficient to pin down the surface orientation uniquely.

Let us assume that we also know what the surface material is and where the light sources are positioned; that is, we are given the reflectance map. Then, the brightness at a point on the contour provides another constraint. The surface orientation must lie on a second one-dimensional locus on the Gaussian sphere. This locus is shown as a broken line in Figure 1. Actually, Figure 1 shows a special case where the brightness is zero. The locus in this case is the self-shadow line, a great circle whose axis points towards the light source. This is the most important case for us here, but the same idea applies in the general case.

The locus from the reflectance map intersects the locus from the tangent vector on the Gaussian sphere. The surface orientation is at the point of intersection.

2.3. Constructing a Needle Map: Shape-from-Shading

The apparent brightness of a surface patch is determined by four factors:

1) Viewer direction,

2) Surface orientation,

3) Surface material, and

4) Light source direction.

Given the surface material, viewer direction, and light source direction, the apparent brightness can be related to surface orientation. This relationship can be expressed by the image irradiance equation. Since surface orientation has two degrees of freedom, the equation can be given in the form

$$E(x,y) = R(f,g),$$

where $E(x,y)$ denotes the brightness observed at the point (x,y) in the image and $R(f,g)$ denotes the brightness expected of a surface patch with orientation (f,g). Here, (f,g) is the orientation at the point on the object corresponding to the image point (x,y). Surface orientation is express in stereographic coordinates [Ikeuchi and Horn 83] so that orientations of surface patches on the occluding boundary correspond to finite values of the parameters. The occluding contour is one of the important sources of boundary conditions.

Since the image irradiance equation provides only one constraint, an additional constraint is necessary to determine surface orientation. We assume that the object surface is smooth. One way to express this constraint is to try and minimize the following measure of the departure from smoothness:

$$f_x^2 + f_y^2 + g_x^2 + g_y^2,$$

where f_x denotes the partial derivative of f with respect to x, and so on. Overall then, the integral to be minimized is

$$\int\int (r + \lambda s)\, dx\, dy,$$

where

$$r = (E(x,y) - R(f,g))^2 \qquad \text{and} \qquad s = f_x^2 + f_y^2 + g_x^2 + g_y^2$$

The parameter λ depends on the relative importance of matching image irradiance versus that of creating a smooth surface. It should be chosen by considering the noise in image brightness measurement, which will affect the accuracy of the term r in the integrand.

This calculus of variation problem [Courant and Hilbert 53] can be solved using Euler's differential equations. If one introduces the Laplacian operator

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

then the solution can be expressed in the form

$$\lambda \nabla^2 f = -(E - R) R_f \qquad \text{and} \qquad \lambda \nabla^2 g = -(E - R) R_g,$$

where R_f and R_g are the partial derivatives of $R(f, g)$ with respect to f and g respectively.

This leads to a simple iterative solution scheme, based on a discrete approximation of the Laplacian. At the point (i, j) on a discrete grid one has

$$(\nabla^2 f)_{i,j} \approx (\overline{f}_{i,j} - f_{i,j})/\rho,$$

where $\overline{f}$ is the average of f computed over a neighborhood of the point (i, j), and ρ is a proportionality constant which depends on the size of this neighborhood.

In the iterative solution scheme, the new values of f and g are computed from the old values of f and g as follows:

$$f^{n+1} = \overline{f}^n + (\rho/\lambda)(E - R^n) R_f^n \qquad \text{and} \qquad g^{n+1} = \overline{g}^n + (\rho/\lambda)(E - R^n) R_g^n,$$

where $\overline{f}$, $\overline{g}$ are local averages of f, and g, respectively [Ikeuchi and Horn 81].

2.4. Constructing a Depth Map from a Needle Map

The second process constructs a depth map, $z(x, y)$, from the needle map, $(p(x, y), q(x, y))$. The problem is overdetermined, since we are given both partial derivatives of the depth. Previously, various *ad hoc* schemes for recovering depth

from needle diagrams have been in vogue [Ikeuchi and Horn 81]. We chose to use a least squares approach instead. The components p and q of the surface gradient are the first partial derivatives of depth, z. The desired depth map thus should minimize the following integral:

$$\int\int (z_x - p)^2 + (z_y - q)^2 \, dx \, dy,$$

where (p, q) denotes surface orientation expressed in the gradient space and z is the surface depth [Horn 80], [Terzopolous 83].

The solution to this calculus of variation problem [Courant and Hilbert 53] can also be obtained using Euler's differential equation,

$$\nabla^2 z = p_x + q_y.$$

This leads to a simple iterative scheme of the form

$$z^{n+1} = \overline{z}^n - \rho(p_x + q_y)$$

where $\overline{z}$ is a local average of the depth z. Again, the parameter ρ depends on how this average is computed.

This process can generate *relative* depth over a connected region. Figure 2 shows the relative depth map generated from a needle map of a vase. The needle diagram was obtained using the photometric stereo method [Woodham 78], [Ikeuchi *et al* 83]. The photometric stereo method was applied to three images of a real vase.

An additional constraint is necessary to generate an *absolute* depth map, however. Fortunately, on the boundary, the Marr-Poggio-Grimson stereo can provide depth information. Actually, we can use this information all the way around the boundary for better accuracy, even though, in theory, we need only find one offset.

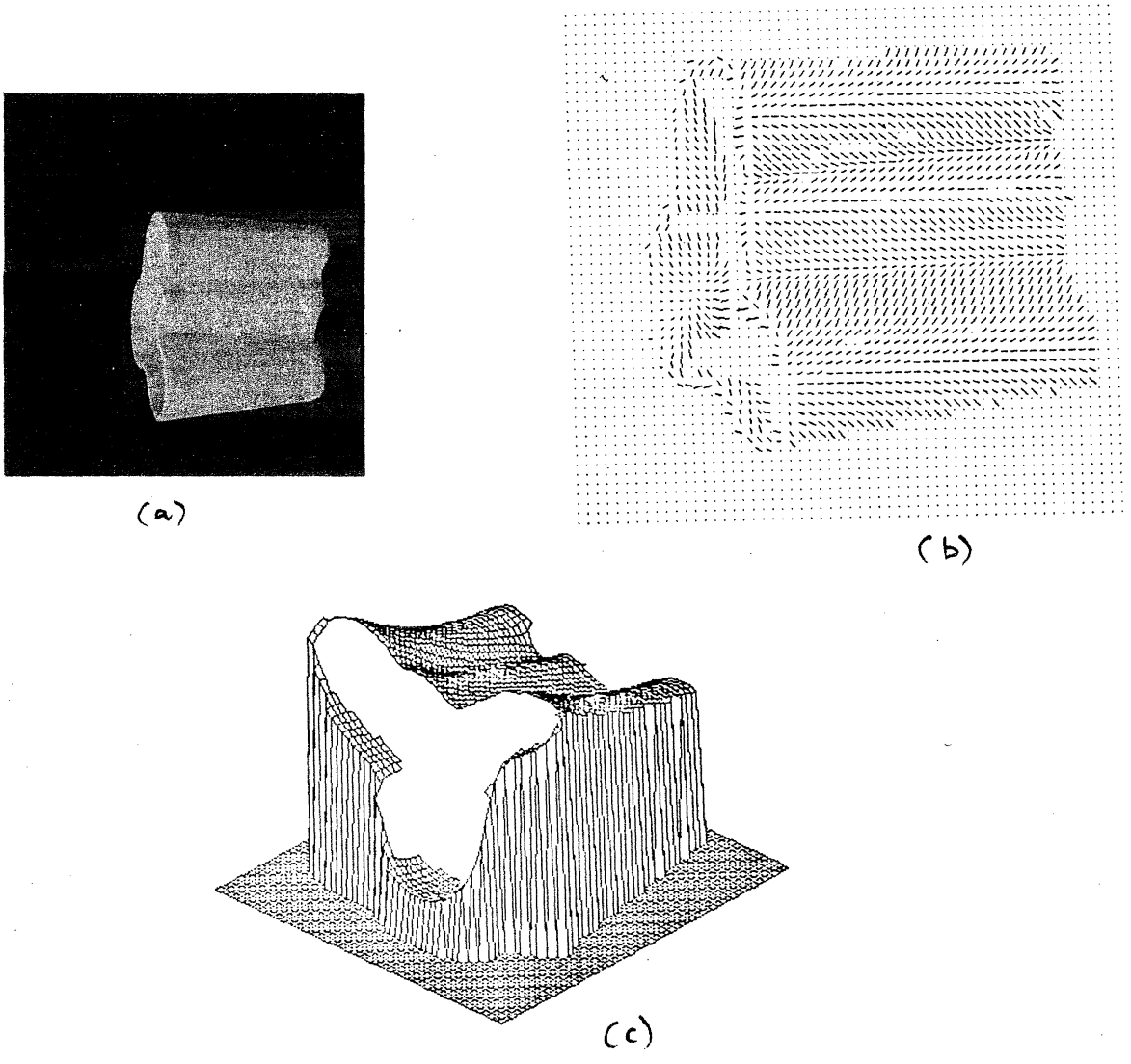

Figure 2. The relative depth map generated from a needle map: (a) Photograph of the original object, (b) The needle map obtained using photometric stereo, (c) The relative depth map computed from the needle map.

2.5. Examples

2.5.1. Synthetic images

A pair of synthetic images was generated to help judge the performance of the algorithm (see Figure 3). In the figure, the x-axis goes to the right and the y-axis upward, while the z-axis comes outward towards the viewer. Two cylinders lie in a plane parallel to the images. Both are 40 units in height and 10 units in radius.

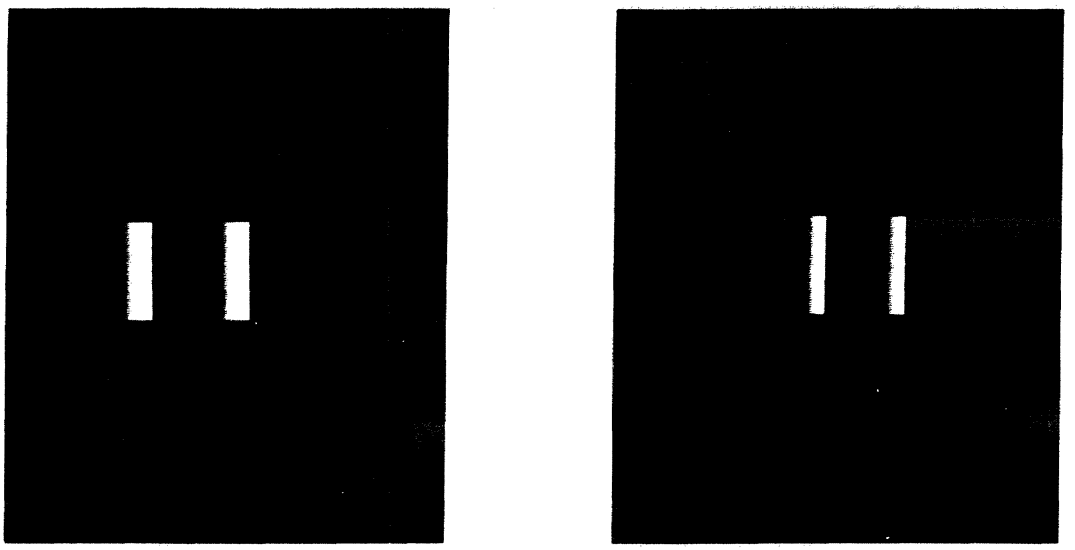

Figure 3. Left and right synthetic images of two cylinders at different depths.

The center of the right one is at $(-10.0, 0.0, 0.0)^T$ and the center of the left one is at $(10.0, 0.0, 30.0)^T$. Both cameras are assumed to perform orthographic projection.

The distance between the two camera is 60.0 units and the distance between the cameras and the origin is 200.0 units. The light source direction is 70.0 degrees in zenith angle and 10.0 degrees in azimuth angle.

Figure 3 shows the images generated. These images are fed to the Marr-Poggio-Grimson stereo algorithm.

2.5.2. Output from Marr-Poggio-Grimson stereo

Figure 4 shows the zero-crossing contours obtained by the stereo system. The Marr-Poggio-Grimson stereo algorithm generates disparity values based on these zero-crossing contours as shown in Figure 5.

2.5.3. Determining Surface Orientation along the Boundary

The disparity values and camera parameters allow one to determine depth information along the zero-crossing contours. Surface orientation can be determined based on this depth information.

Let us suppose that T, N and S are the tangent of the zero crossing contour, the surface normal, and the light source direction, respectively. Since we assume

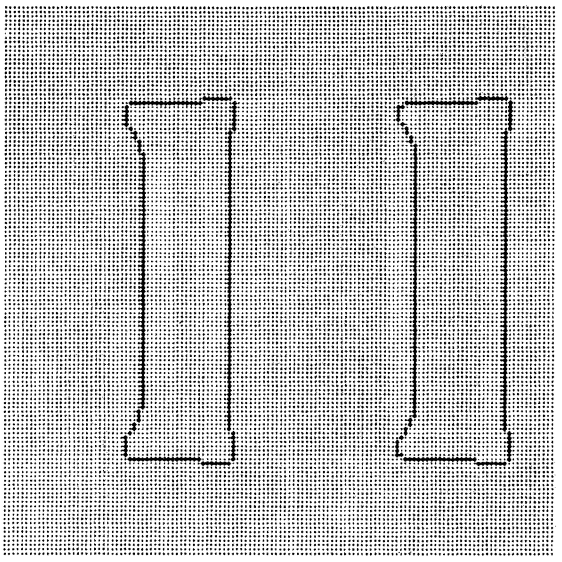

Figure 4. Zero crossings obtained from the stereo pair of synthetic images.

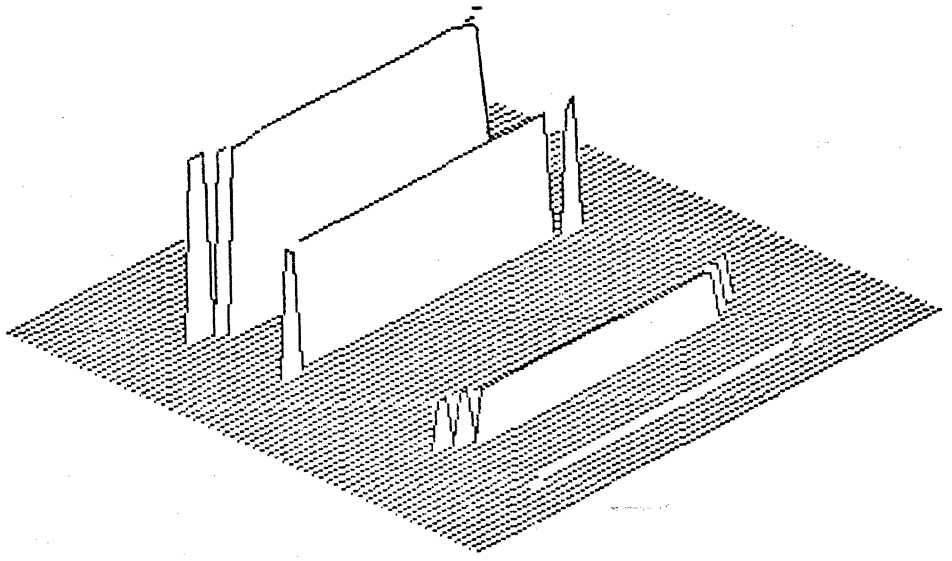

Figure 5. Disparity value obtained using the stereo algorithm.

that the surface of the object here has Lambertian reflectance properties, the image irradiance equation gives the following:

$$\mathbf{S} \cdot \mathbf{N} = E$$

where E is the observed brightness. The tangent is perpendicular to the surface normal,

$$\mathbf{T} \cdot \mathbf{N} = 0$$

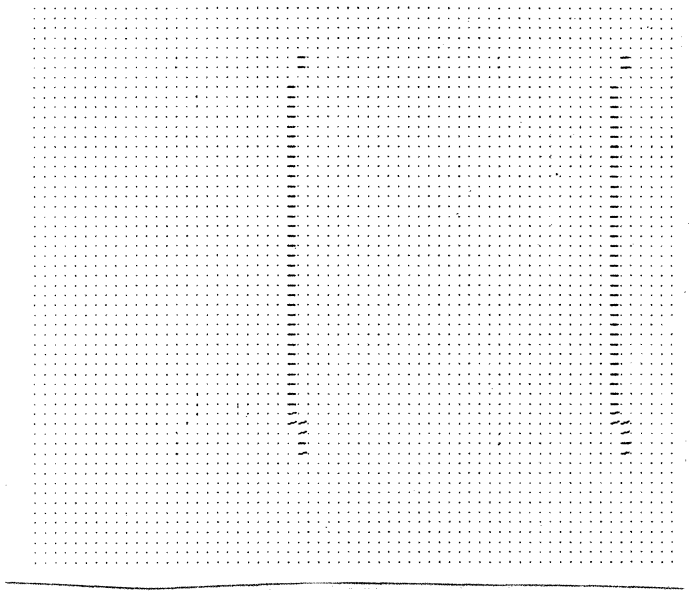

Figure 6. Surface orientations along the boundary.

---

and N is a unit vector. Thus,

$$\mathbf{N} \cdot \mathbf{N} = 1.$$

Using these three equations, we can solve for N,

$$\mathbf{N} = \frac{1}{(\mathbf{T} \times \mathbf{S})^2}\left[E(\mathbf{T} \times \mathbf{S}) \pm \sqrt{(\mathbf{T} \times \mathbf{S})^2 - E^2\,\mathbf{T}^2\mathbf{S}}\right] \times \mathbf{T}.$$

This solution can be verified readily by taking dot-products with the vectors **T**, **S**, and **N**.

We usually need to determine surface orientation only on the self shadow lines. There, $E = 0$, and we get the much simpler expression:

$$\mathbf{N} = \pm\frac{\mathbf{T} \times \mathbf{S}}{|\mathbf{T} \times \mathbf{S}|}.$$

We chose the sign so as to obtain an outward pointing normal. Along occluding boundaries the surface normal can be found just as simply:

$$\mathbf{N} = \pm\frac{\mathbf{T} \times \mathbf{V}}{|\mathbf{T} \times \mathbf{V}|},$$

where **V** is a vector pointing towards the viewer. Figure 6 shows the surface orientations obtained along the self-shadow lines and occluding boundaries.

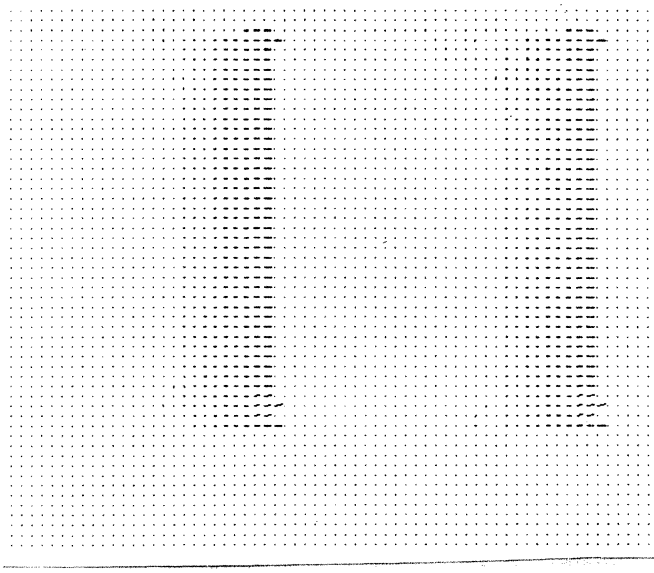

Figure 7. The needle map computed by the shape-from-shading algorithm.

The reflectance map, the image, and the needle map along the boundaries are input information for the shape-from-shading algorithm. Figure 7 shows the final needle map obtained by the shape-from-shading algorithm.

2.5.4. Constructing a Continuous Depth Map

The process for generating a depth map from the needle map is executed next. Figure 8a shows the depth map obtained by the construction process. Figure 8b shows a cross section at the middle of the cylinders. The significant differences between the reconstructed depth profiles and the profiles of the original objects appear to be in part due to errors in the depth values obtained by the stereo algorithm on the boundary (These arise to some extend because of the fact that the left camera sees a different curve on the surface as occluding boundary than does the right).

3. Depth Construction Schema Using a Pair of Needle Maps

Sometimes a pair of needle maps is provided. For example, two photometric stereo systems can provide a pair of needle maps directly. Another possibility depends on a feature of the stereo algorithm: The depth information it produces applies to *both* the left and the right image. Thus, we can obtain a pair of

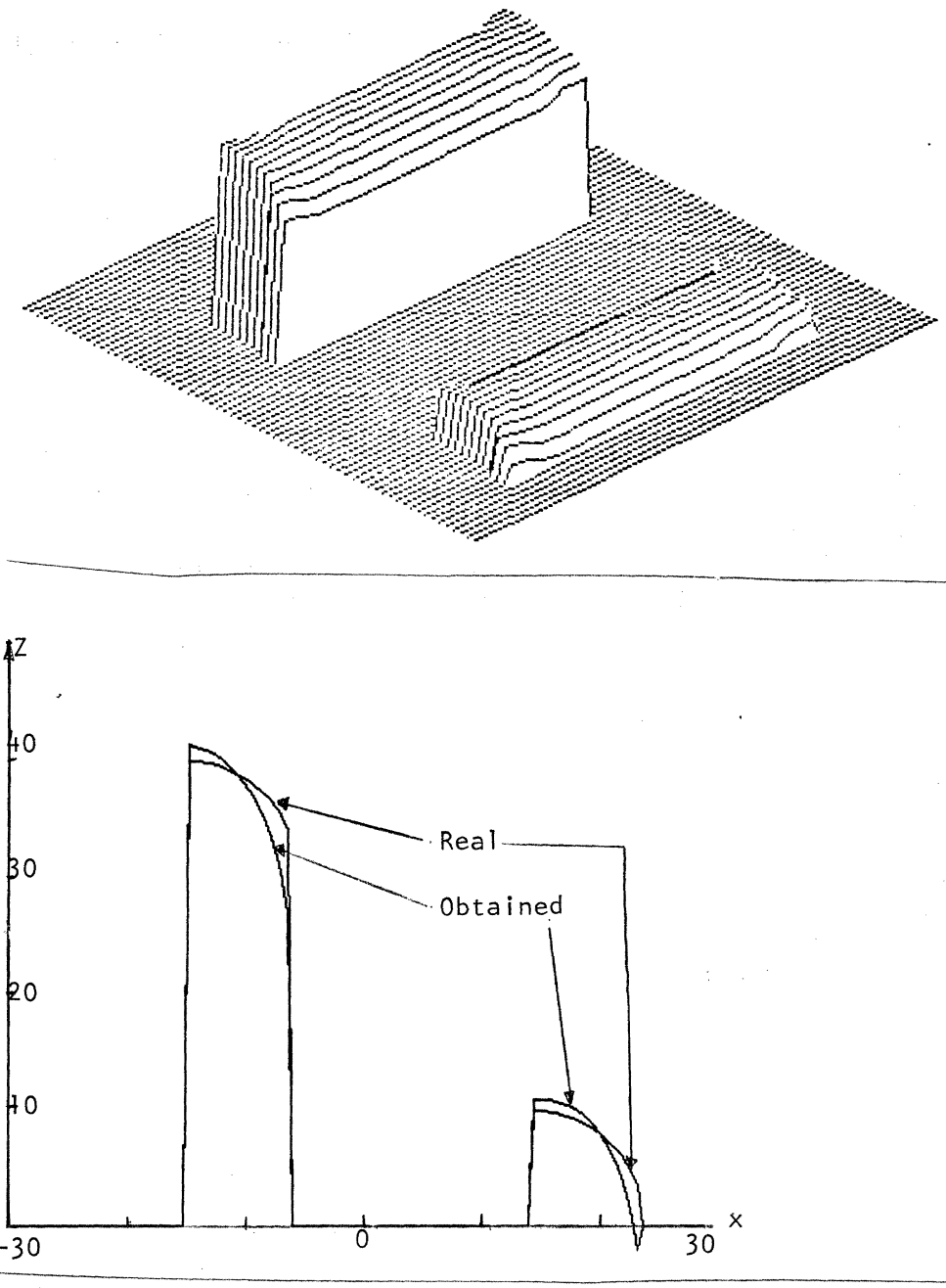

Figure 8. The resulting depth map and a cross-section through it.

needle maps by applying the shape-from-shading algorithm to the left and right images independently, using surface orientation information on the zero crossing boundaries.

This method has two potential advantages:

1) The surface normal has two degrees of freedom, while brightness has only one (Thus, the surface normal has less ambiguity than brightness),

2) Even when the reflectance map is not accurately known, the constructed result does not suffer, because surface depth is generated by comparing the disparity of surface normals.

The overall strategy is as follows:

1) Obtain a pair of needle maps using either the photometric stereo system or the method described in the first part of this article, and

2) Construct the depth map using an iterative method.

3.1. Iterative Construction using two Needle Maps

In this case a pair of needle maps are the input to the system. Let the superscripts l and r denote the left and right images respectively. We arbitrarily pick the left image as a reference (It might have been better to develop a more symmetrical form of the equations).

We will minimize two factors: The first one is the measure of departure from correct surface orientation used already above

$$s = (z_x - p^l)^2 + (z_y - q^l)^2.$$

Suppose that the point (x, y) in the left image corresponds to the point $(ax+bz+c, y)$ in the right image, where a, b, and c are parameters determined from the relative orientation of the two cameras. Then the second factor is the correspondence requirement between orientations,

$$d_p = (p^r(ax + bz + c, y) - p^l(x, y))^2,$$

$$d_q = (q^r(ax + bz + c, y) - q^l(x, y))^2,$$

This term only applies where there actually *is* a correspondence between the two images.

We will minimize the following integral:

$$e = \int\int (s + \lambda(d_p + d_q))\, dx\, dy$$

Using Euler's differential equation, we get

$$\nabla^2 z = (p_x^l + q_y^l) + (\lambda\, b)[(p^r - p^l)p_x^r + (q^r - q^l)q_x^r].$$

From this, one can obtain the following iterative scheme:

$$z^{n+1} = \bar{z}^n - \rho\,(p_x^l + q_y^l) - (\rho\,\lambda\, b)[(p^r - p^l)p_x^r + (q^r - q^l)q_x^r].$$

This iterative algorithm works well near the optimal value. In areas where there is no correspondence, we set $(\rho\,\lambda\, b) = 0$. Namely, only simple integration is executed in that area.

3.2. Example 1: Synthetic Image

3.2.1. Output from Marr-Poggio-Grimson Stereo

We can generate a pair of needle maps from the synthetic images in Figure 3. Since depth information obtained by the Marr-Poggio-Grimson stereo algorithm applies to both the left image and the right image, a pair of needle maps along the zero-crossing contours can be obtained. Applying the shape-from-shading algorithm to the original image pair with these boundary conditions independently, we obtain a pair of needle maps as shown in Figure 9.

3.2.2. Result of Shape from Shading Algorithm

Figure 10a shows the depth map obtained by shape-from-shading. Figure 10b shows a cross section at the middle of the cylinders. In order to judge the basic performance of this solution schema, we did *not* use the absolute depth given by the Marr-Poggio-Grimson stereo along the zero-crossing contours. Instead, we allowed the depth values to float free at the zero-crossing contours. For comparison, the surface generated by the method described earlier is also presented in the same figure. It is not clear why there are still significant errors in the result.

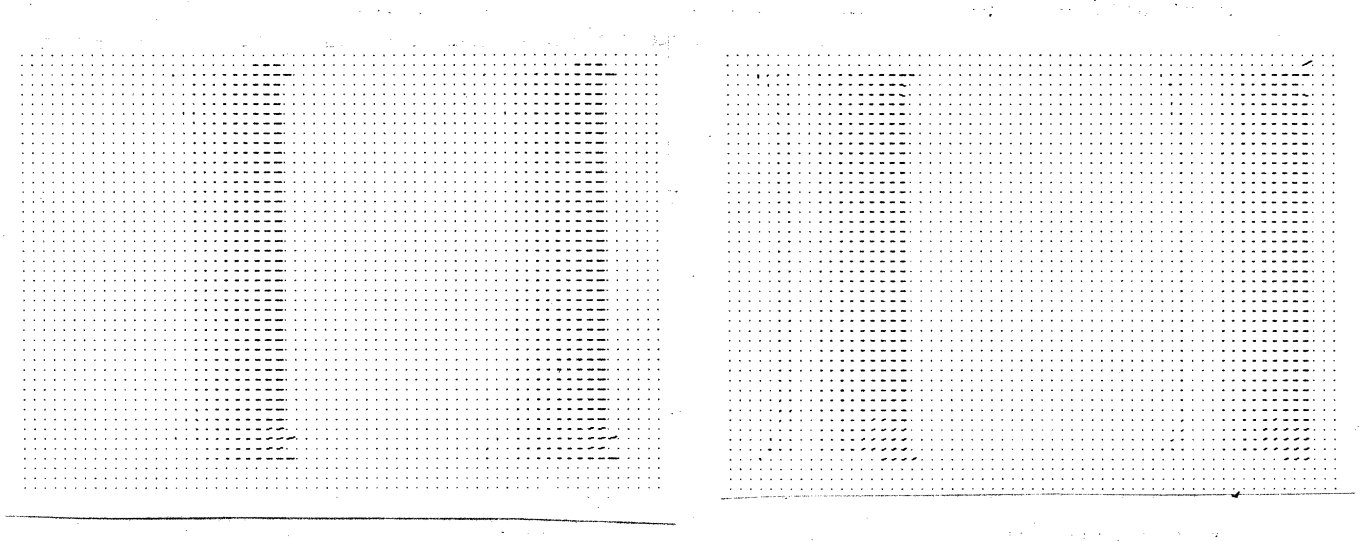

Figure 9. A pair of needle maps obtained by applying the shape-from-shading algorithm to left and right images independently.

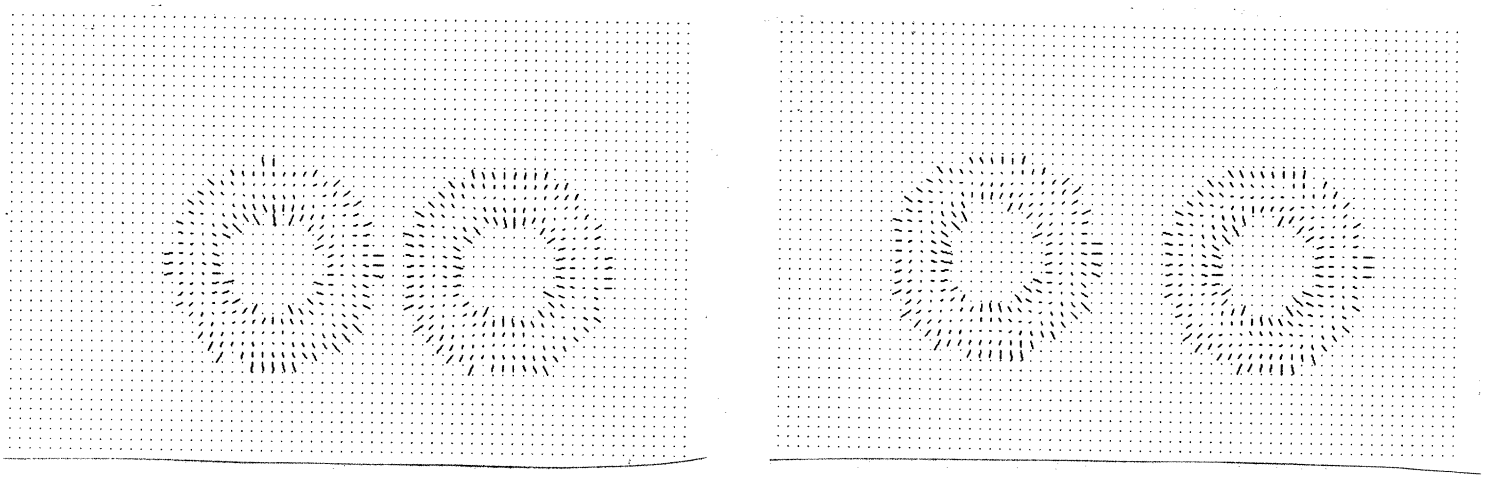

Figure 11. A pair of needle maps of two torii obtained from real images.

3.3. Example 2: Real Image

3.3.1. Using Output from Photometric Stereo

The method was next applied to two real images of two donut-shaped objects. Two photometric stereo systems provide the pair of needle maps shown in Figure 11. Note that the left donut is higher than the right one.

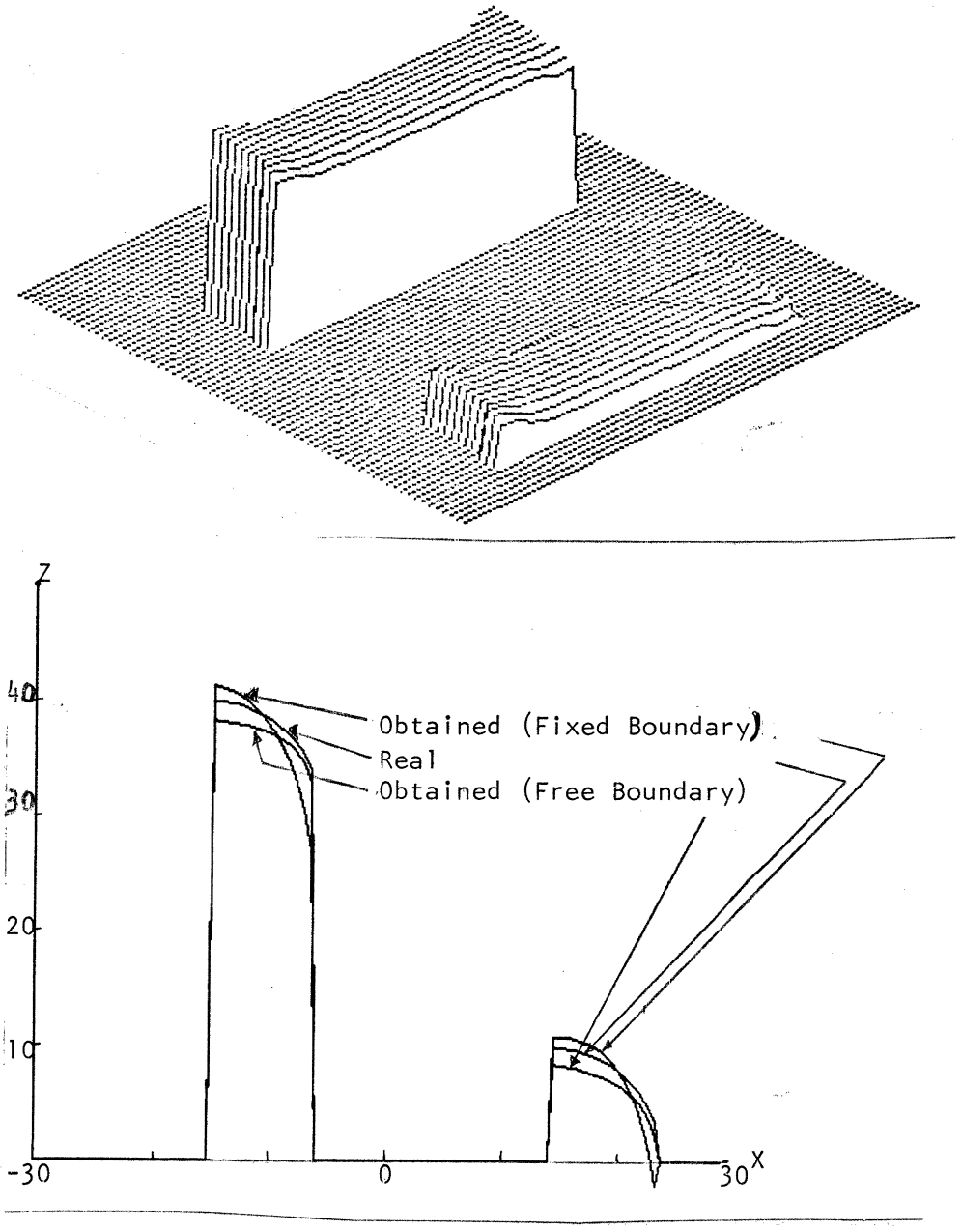

Figure 10. The surface generated using the pair of needle maps.

3.3.2. Result from Shape from Shading Algorithm

The shape-from-shading process generated the depth map shown in Figure 12.

4. Conclusions

This paper describes two methods for constructing the depth map of a surface.

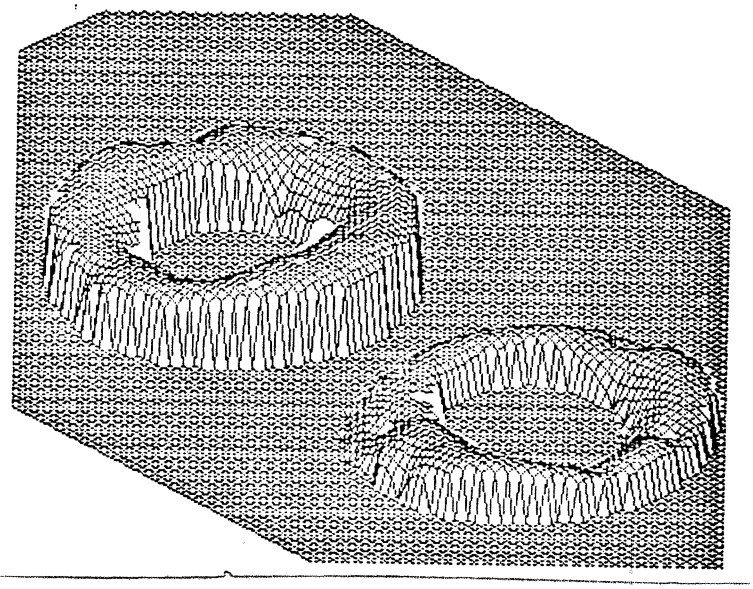

Figure 12. The surface generated from the pair of needle maps.

The first one uses the needle map obtained from shading information and the depth information obtained by the Marr-Poggio-Grimson stereo algorithm. The second one uses two needle maps.

The second scheme works well when the disparity is relatively small. Future research will address the development of a robust technique based on the result presented here.

5. Acknowledgements

Drafts of this article were read by K. Nishihara and also by B. K. P. Horn, who used vector notation to simplify the mathematics in section 2.5.3 and rewrote the solutions of the calculus of variation problems and their discrete approximation. E. Grimson permitted me to use the Marr-Poggio-Grimson stereo software. Their comments are appreciated.

6. References

Brady, M. and Horn, B. K. P., "Rotationally Symmetric Operators for Surface Interpolation," *Computer Vision, Graphics, and Image Processing*, Vol. 21, No. 4, April 1983.

Courant, R. and Hilbert, D., *Methods of Mathematical Physics—Volume I*, Interscience, New York, 1953.

Grimson, W. E. L., *From Images to Surface: A Computational Study of the Human Early Visual System*, M. I. T. Press, Cambridge, 1981.

Grimson, W. E. L., "A Computational Theory of Visual Surface Interpolation," *A. I. Memo. No. 613*, M. I. T., June 1891.

Grimson, W. E. L., "Binocular Shading and Visual Surface Reconstruction," *A. I. Memo. No. 697*, M. I. T., August, 1982.

Hildreth, E. "A Computer Implementation of a Theory of Edge Detection," *A. I. TR-579*, M. I. T., 1980.

Horn, B. K. P. Personal communication, 1980.

Horn, B. K. P. "The Least Energy Curve," *A. C. M. Transactions on Mathematical Software*, Vol. 9, No. 4, December 1983.

Ikeuchi, K. and Horn, B. K. P., "Numerical Shape from Shading and Occluding Boundaries," *Artificial Intelligence*, Vol. 17 ,1981.

Ikeuchi, K., Horn, B. K. P., Nagata, S., Callahan, T., and Feingold, O., "Picking up an Object from a Pile of Objects," *A. I. Memo. No. 726*, M. I. T., May, 1983.

Marr, D. and Poggio, T., "A Theory of Human Stereo Vision," *A. I. Memo. No. 451*, M. I. T., Nov., 1977.

Marr, D., *Vision*, Freeman, 1982.

Terzopoulos, D., "Multi-Level Reconstruction of Visual Surface," *A. I. Memo. 671*, M. I. T., 1982.

Terzopoulos, D., "The Role of Constraints and Discontinuities in Visible-Surface Reconstruction," *Proc. IJCAI-8*, Karlsruhe, pp. 1073-1077.

Woodham, R. J., "Reflectance Map Techniques for Analyzing Surface Defects in Metal Castings," *A. I. TR-457*, M. I. T., 1978.