DISPARITY-BASED PHOTOMETRIC STEREO

José R.A. Torreão and Cecilio J.L. Pimentel
Departamento de Informática, Universidade Federal de Pernambuco
50.739 Recife PE Brasil

Abstract

We point out a relationship between the processes of Stereoscopy and Photometric Stereo which can be exploited for the estimation of depth with single-camera imaging set-ups.

The displacement of pixel intensities observed when the illumination direction changes, in the formation of photometric stereo image pairs, can be treated as a disparity map arising in an equivalent convergent stereoscopic system. Such photometric disparity map can thus be estimated through the use of a stochastic approach which essentially matches the two images at pixel level, while the shape of the pictured surface can be simultaneously reconstructed by the usual photometric stereo technique. From the disparity and shape information, the parameters of the equivalent stereoscopic system can be estimated, and a depth field can thus be recovered by triangulation. This strategy allows the simultaneous shape and depth reconstruction from photometric stereo image pairs, which could be useful for simple industrial vision applications. Preliminary experimental results obtained with such scheme, for synthetic images, are presented.

KEYWORDS: Artificial Vision, Stereoscopy, Photometric Stereo

The present work has been supported by the grant APQ 171-1.03/90 from FACEPE (Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco).

1. Introduction

Stereoscopy and Photometric Stereo are two processes which are widely used for the estimation of geometric properties of imaged surfaces in Computational Vision. The first, also known as Geometric Stereo [**Grimson 86, Barnard 82**], extracts depth information from the displacement of features (which is called 'disparity') observed when two images of a scene are captured from different viewing positions. At the core of that process is the matching between points in the two images which are projections of the same location in the scene. Photometric Stereo [**Woodham 80**], on the other hand, is a Reflectance-Map technique employing two or more images, obtained from a single camera for different illumination directions, which allows the estimation of an orientation map for the viewed surfaces. The relevant information, in this case, is yielded by the multiple intensity values at each pixel site.

There is an obvious relationship between the two stereo modalities, since they are both multiple-image processes which yield the related properties of depth ($z(x,y)$) and surface orientation ($\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}$), but, beyond this, there is a less apparent and more useful duality between the two, as yet not fully exploited [**Torreão 91**]. This arises from the fact that in Stereoscopy, as well as in Photometric Stereo, both the geometric information (displacement of features) and the photometric information (multiple intensities) are available to be used for surface estimation purposes.

In Stereoscopy, for instance, once the correspondence of a pair of points in the images has been established, there will be two intensity values available, in the general case, for the recovery of surface orientation, once the Reflectance Map [**Horn 77, 79**] is known. On the other hand, for smooth surfaces, the change of illumination direction in Photometric Stereo entails a variation of pixel intensities in the image lattice, which can be treated as an intensity-displacement field akin to the disparity field resulting from the change of viewing position in Stereoscopy. Such displacement field (which will be referred to as the *photometric disparity* field) can, for example, be obtained through the stochastic algorithm for stereo correspondence proposed by Barnard [**Barnard 86**], where the matching between images is performed essentially at pixel level, yielding a dense map.

If we can determine a stereoscopic configuration which yields a disparity map similar to the one obtained from a given pair of photometric stereo images, we can employ such map for the inference of depth via simple triangulation. Here we propose a strategy for finding such equivalent (convergent) stereo system, when the displacement field and the surface shape have been recovered from a given photometric stereo pair. Our strategy is applicable for smooth surfaces, when the angle between

the illumination directions under which the images are captured is small. Under such conditions it allows the use of a single-camera set-up for the essentially simultaneous estimation of both depth and orientation. This could prove to be a very attractive alternative to the use of more expensive stereoscopic arrangements (involving Dual Photometric Stereo [Ikeuchi 87], for instance) in simple industrial vision applications.

2. Relating Photometric Stereo and Stereoscopic Systems

We refer to Figure 1, where a cross-sectional view of an imaging arrangement consisting of two cameras in a convergent stereoscopic configuration is shown. The coordinate axes associated with the scene and with each camera are denoted respectively by (x, z), (x_1, z_1), and (x_2, z_2), with the corresponding y-axes (not shown) being perpendicular to the paper. It should be noted that, for convenience, the axis z, in the scene system, points in the opposite direction to the axis z_1 in the first-camera system. With such orientation choice, the coordinates of a point, as given in these two systems, are the same. Two illumination directions and two viewing directions are represented, denoted respectively by $\hat{s}$ and $\hat{s}'$, and by $\hat{v}$ and $\hat{v}'$. Observe that the vectors $\hat{v}'$ and $\hat{s}'$ are obtained as rotations, in the xz plane, of the vectors $\hat{v}$ and $\hat{s}$, respectively. It is evident that, for such stereo configuration, the epipolar lines will lie along the x-direction. (We recall that, given a scene point, the associated epipolar lines in the images are those lines where the plane containing such point and the centers of projection of the two cameras cut the image planes. The search for matching pairs of points in the stereoscopic images needs to be performed only across the epipolar lines.)

We will be considering two pairs of images obtained with such imaging arrangement. These correspond to a stereoscopic pair obtained from the viewing directions $\hat{v}$ and $\hat{v}'$, under illumination $\hat{s}$, and a photometric stereo pair obtained from the viewing direction $\hat{v}$ under the illuminations $\hat{s}$ and $\hat{s}'$. The two pairs, therefore, share a common image: the one obtained from $\hat{v}$ under illumination $\hat{s}$. With the use of the reflectance map function R [**Horn 77, 79**], those images can be expressed as

$$I_1(i,j) = R(\hat{n}_{ij}, \hat{s}, \hat{v}),\ I_2^S(i, j - D_{ij}) = R(\hat{n}_{ij}, \hat{s}, \hat{v}'),\ \text{and}\ I_2^P(i,j) = R(\hat{n}_{ij}, \hat{s}', \hat{v}), \qquad (1)$$

where D_{ij} and $\hat{n}_{ij}$ denote the disparity and the surface-orietation normal vector at site (i, j), which is given in the scene coordinate system, and where the superscripts S and P stand for stereoscopic and photometric stereo, respectively.

Applying the stochastic stereo algorithm by Barnard [**Barnard 86**], we can extract disparity

maps from each of the image pairs by minimizing, via simulated annealing relaxation [Kirkpatrick 83, Geman 84], a cost function of the form

$$E = U + \lambda V, \tag{2}$$

with

$$U = \sum_{i,j} |I_1(i,j) - I_2(i, j + D_{ij})|, \tag{3}$$

and

$$V = \sum_{i,j} |\nabla D_{ij}|. \tag{4}$$

The term U represents a *photometric constraint*, which dictates that the disparity assignments should map points in the first image to points in the second image with comparable intensity, while V encodes a *smoothness constraint*, which penalizes disparity maps with abrupt local variations, and which is implemented as the sum of the absolute differences between the disparity at site (i,j) and those on its eight neighboring sites. The constant λ is a weighting factor which determines the relative contribution of the two terms to E.

We now recall that our aim is to find a stereoscopic configuration which yields a pair of images similar to a given photometric stereo pair, in the sense that the intensity displacement arising from the change of viewing position, in the first case, should be the same as the one arising from the illumination change, in the second. By considering the form of equation (3) for the image pairs $\{I_1, I_2^S\}$ and $\{I_1, I_2^P\}$, when I_1, I_2^S and I_2^P are given in terms of the reflectance maps as in equation (1), it is easy to see that similar photometric and stereoscopic disparity maps will be obtained when the direction $\hat{v}'$ satisfies the relation

$$R(\hat{n}_{ij}, \hat{s}, \hat{v}') \approx R(\hat{n}_{i,j+D_{ij}}, \hat{s}', \hat{v}), \quad \forall i,j. \tag{5}$$

We thus suggest the following strategy for the parallel estimation of depth and shape in a photometric stereo configuration:

1) Barnard's algorithm can be employed for extracting a photometric disparity map through the minimization of the cost functional

$$E(D) = \sum_{ij} |I_1(i,j) - I_2^P(i, j + D_{ij})| + \lambda \sum_{ij} |\nabla D_{ij}|, \tag{6}$$

via stochastic relaxation, while shape estimates can be obtained through the traditional photometric stereo reconstruction.

2) From the results for $\hat{n}_{ij}$ and D_{ij}, the equivalent stereoscopic direction $\hat{v}'$ can be determined from equation (5) by the minimization of

$$E(\hat{v}') \equiv E(\theta_s) = \sum_{ij} |R(\hat{n}_{ij}, \hat{s}, \hat{v}') - R(\hat{n}_{i,j+D_{ij}}, \hat{s}', \hat{v})|, \tag{7}$$

where, from Figure 1, we see that $\hat{v}'$ has the form $\hat{v}' = (\sin\theta_s, 0, \cos\theta_s)$.

3) Once the angle θ_{eq} which minimizes the expression in (7) ($\theta_{eq} = \arg\min\{E(\theta_s)\}$) is available, relative depth estimates can be obtained from the photometric disparity field through the relation

$$z(i,j) = \frac{1}{\sin\theta_{eq}}[D_{ij} - X(i,j)(1 - \cos\theta_{eq})],$$

where we have taken $d = z_{max}\tan\theta_{eq}$ for the camera separation, and where $z(i,j) = z_{max} - z_{abs}(i,j)$ (see Figure 1).

In the following section, we present results of the aplication of the above strategy to some pairs of synthetically generated images.

3. Obtaining Shape and Depth from the Disparity-Based Photometric Stereo

We have considered Reflectance Maps as below [**Horn 77**], involving both a diffuse reflection component, and a quasi-specular component given by the Phong model [**Phong 75**],

$$R(\hat{n}, \hat{s}, \hat{v}) = \rho[(1-\sigma)\hat{n}.\hat{s} + \frac{1}{2}\sigma(m+1)(2(\hat{n}.\hat{s})(\hat{n}.\hat{v}) - \hat{v}.\hat{s})^m], \tag{8}$$

where ρ is the albedo, σ is the specular weight, and m is a constant. When $\sigma = 0$, and therefore only the diffuse or Lambertian reflection is present, the Reflectance Map carries no information about the observation direction, and therefore equation (7) can not be used for the estimation of $\hat{v}'$. We remark that we are, nevertheless, still able to employ our strategy in such case, by considering equation (5) for a Reflectance Map as in (8) in the limit of $\sigma \to 0$. In such limit, the minimization of (6) matches points for which $\hat{n}_{ij}.\hat{s} \approx \hat{n}_{i,j+D_{ij}}.\hat{s}'$. Employing this relation, together with the form (8), in equation (5), it is easy to obtain the condition $(2(\hat{n}_{ij}.\hat{s})(\hat{n}_{ij}.\hat{v}') - \hat{v}'.\hat{s}) \approx (2(\hat{n}_{i,j+D_{ij}}.\hat{s}')(\hat{n}_{i,j+D_{ij}}.\hat{v}) - \hat{v}.\hat{s}')$, for all sites (i,j). We can thus use for $E(\hat{v}')$, in the Lambertian case, the expression

$$E(\hat{v}') \equiv E(\theta_s) = \sum_{ij} |(2(\hat{n}_{ij}.\hat{s})(\hat{n}_{ij}.\hat{v}') - \hat{v}'.\hat{s}) - (2(\hat{n}_{i,j+D_{ij}}.\hat{s}')(\hat{n}_{i,j+D_{ij}}.\hat{v}) - \hat{v}.\hat{s}')|, \tag{9}$$

which is a purely geometric relation, independent of σ.

The photometric stereo images employed in our experiments have been generated through the use of the Image Irradiance Equation [Horn 77, 79]

$$I^{(k)}(i,j) = R(\hat{n}, \hat{s}_k, \hat{v}), \quad k = 0, 1, 2$$

for Lambertian reflectance maps, and illumination directions $\hat{s}_0 = (0, 0, 1)$, $\hat{s}_1 = (\sin\theta_p, 0, \cos\theta_p)$ and $\hat{s}_2 = (\sin(-\theta_p), 0, \cos(-\theta_p)) \equiv (-\sin\theta_p, 0, \cos\theta_p)$. From the pairs of images $\{I^{(0)}, I^{(1)}\}$ and $\{I^{(0)}, I^{(2)}\}$, two disparity fields, $D_1(i,j)$ and $D_2(i,j)$ (which are taken here in absolute value), were obtained. From these, a combined photometric disparity map was constructed as $D(i,j) = D_1(i,j) - D_2(i,j)$, which, together with the surface orientation estimates yielded by photometric stereo, has been employed for the determination of the equivalent stereoscopic angle θ_{eq}, through (9). We have used this combination of disparity maps obtained for illumination directions symmetric about the z-axis, in order to avoid the unmatchable regions which occur when only one pair of photometric stereo images is employed (such regions, where no disparity information is available, are the equivalent of the occluded areas for stereoscopic systems). Depth estimates can be obtained from such combined maps, as below:

As we have seen, relative depth can be expressed in terms of disparity, for the equivalent convergent stereoscopic system, by (see Figure 1)

$$z(i,j) = \frac{D_1(i,j) - (1 - \cos\theta_{eq}).X(i,j)}{\sin\theta_{eq}}.$$

For an angle $-\theta_{eq}$, this expression becomes

$$z(i,j) = \frac{-D_2(i,j) + (1 - \cos\theta_{eq}).X(i,j)}{\sin\theta_{eq}},$$

and from the two expressions above we can get

$$z(i,j) = \frac{D_1(i,j) - D_2(i,j)}{2.\sin\theta_{eq}} = \frac{D(i,j)}{2.\sin\theta_{eq}}, \tag{10}$$

which is the relation we have employed in our estimation of depth from the photometric disparity data.

In figures 2 through 4, we show results of depth reconstruction from the above scheme. Letters (a) in those figures depict input photometric stereo pairs (for angles θ_p equal to 15°, 10°, and 12°, respectively), along with the recovered disparity fields. Letters (b) show the respective tridimensional depth reconstructions obtained from (10), where the angles θ_{eq} have been estimated from

the curves in Figure 5, which represent the form of $E(\theta_s) \times \theta_s$ for each of the imaged surfaces. From Figure 5, we find the minimizing angles θ_{eq} of 15°, 10°, and 20°, respectively, and we observe that, for surfaces which show axial symmetry about the y-axis, one has simply $\theta_{eq} = \theta_p$, as should have been expected [**Torreão 91**]. The depth values obtained in those reconstructions have been consistent with the synthetic surfaces employed.

4. Concluding Remarks

In this article, we have proposed a single-camera depth evaluation scheme, based on an analogy between the processes of Stereoscopy and Photometric Stereo, which allows the extraction of a disparity map from a photometric stereo image pair. Such disparity map, together with an orientation field obtained from the traditional photometric stereo technique, can be employed for the determination of the vergence angle of an equivalent convergent stereoscopic set-up. In this way, depth estimates can be recovered from the obtained photometric disparity field. Here we have considered only the estimation of relative depth, but we are presently trying to extend our framework for the treatment of images generated under a perspective projection geometry, which would allow the inference of absolute depth information.

5. References

[**Barnard 86**] Barnard, S.T., *A Stochastic Approach to Stereo Vision*, in Procs. 5th National Conf. on AI, pp. 676-680, Philadelphia, PA, 1986.

[**Barnard 82**] Barnard, S.T., and M.A. Fischler, *Computational Stereo*, Computing Surveys, vol.14(4), pp. 553–572, 1982.

[**Geman 84**] Geman S., and D. Geman, *Stochastic Relaxation, Gibbs Distributions, and The Bayesian Restoration of Images*, IEEE Trans. Patt. Anal. Mach. Intell. vol.6, pp. 721–741, 1984.

[**Grimson 86**] Grimson, W.E.L., *From Images to Surfaces*, The MIT Press, Cambridge, MA, 1986.

[**Horn 77**] Horn, B.K.P., *Understanding Image Intensities*, Artificial Intell., vol.8(11), pp. 201–231, 1977.

[**Horn 79**] Horn, B.K.P., and R.W. Sjoberg, *Calculating the Reflectance Map*, Applied Optics,

vol.18(11), pp. 1770–1779, 1979.

[Ikeuchi 87] Ikeuchi, K., *Determining a Depth Map Using a Dual Photometric Stereo*, Int. J. Robotics Research, vol. 6(1), 1987.

[Kirkpatrick 83] Kirkpatrick, S., C.D. Gelatt, Jr., and M.P. Vecchi, *Optimization by Simulated Annealing*, Science, 220(4598), pp. 671–680, 1983.

[Phong 75] Phong, B.T., *Illumination for Computer Generated Pictures*, Commun. Ass. Comput. Mach., vol. 18(6), 1975.

[Torreão 89] Torreão, J.R.A., *A Bayesian Approach to 3-D Shape Estimation for Robotic Vision*, Ph.D. Thesis, Department of Physics, Brown University, September 1989.

[Torreão 90a] Torreão, J.R.A., *Novel Applications of the Photometric Stereo Set-Up for Robotic Vision*, in Procs. X Congress of the Brazilian Computing Society, Vitória, Brazil, July 1990.

[Torreão 90b] Torreão, J.R.A., *A Bayesian Framework for the Estimation of Shape and Shape Discontinuities in 3-D*, in Procs. XVI Latin-American Conference in Informatics – CLEI, Asunción, Paraguay, September 1990.

[Torreão 91] Torreão, J.R.A., *Towards a Single-Camera Stochastic Stereoscopy*, in Procs. VIII Brazilian Symposium on Artificial Intelligence, Brasília, Brazil, November 1991.

[Woodham 80] R.P. Woodham, *Photometric Method for Determining Surface Orientation from Multiple Images*, Optical Engineering, vol.19(1), pp. 139–144, 1980.

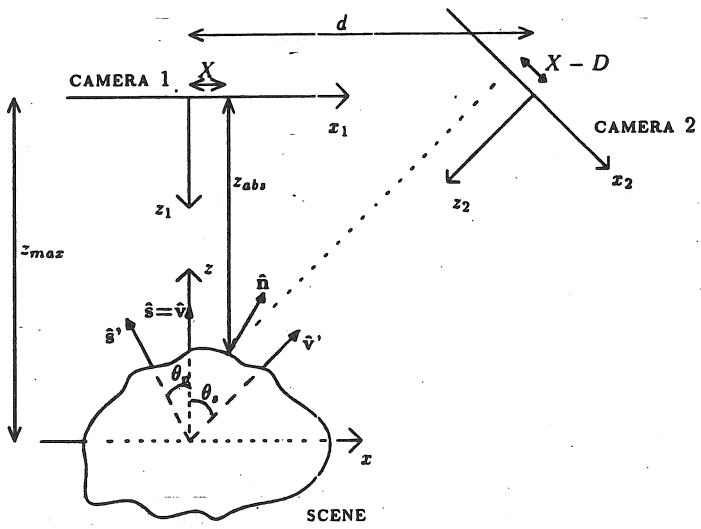
d
X − D
CAMERA 1
X
CAMERA 2
x_1
x_2
z_1
z_2
z_{abs}
z_{max}
z
$\hat{n}$
$\hat{s}=\hat{v}$
$\hat{s}'$
$\hat{v}'$
θ_i
θ_e
x
SCENE

Figure 1

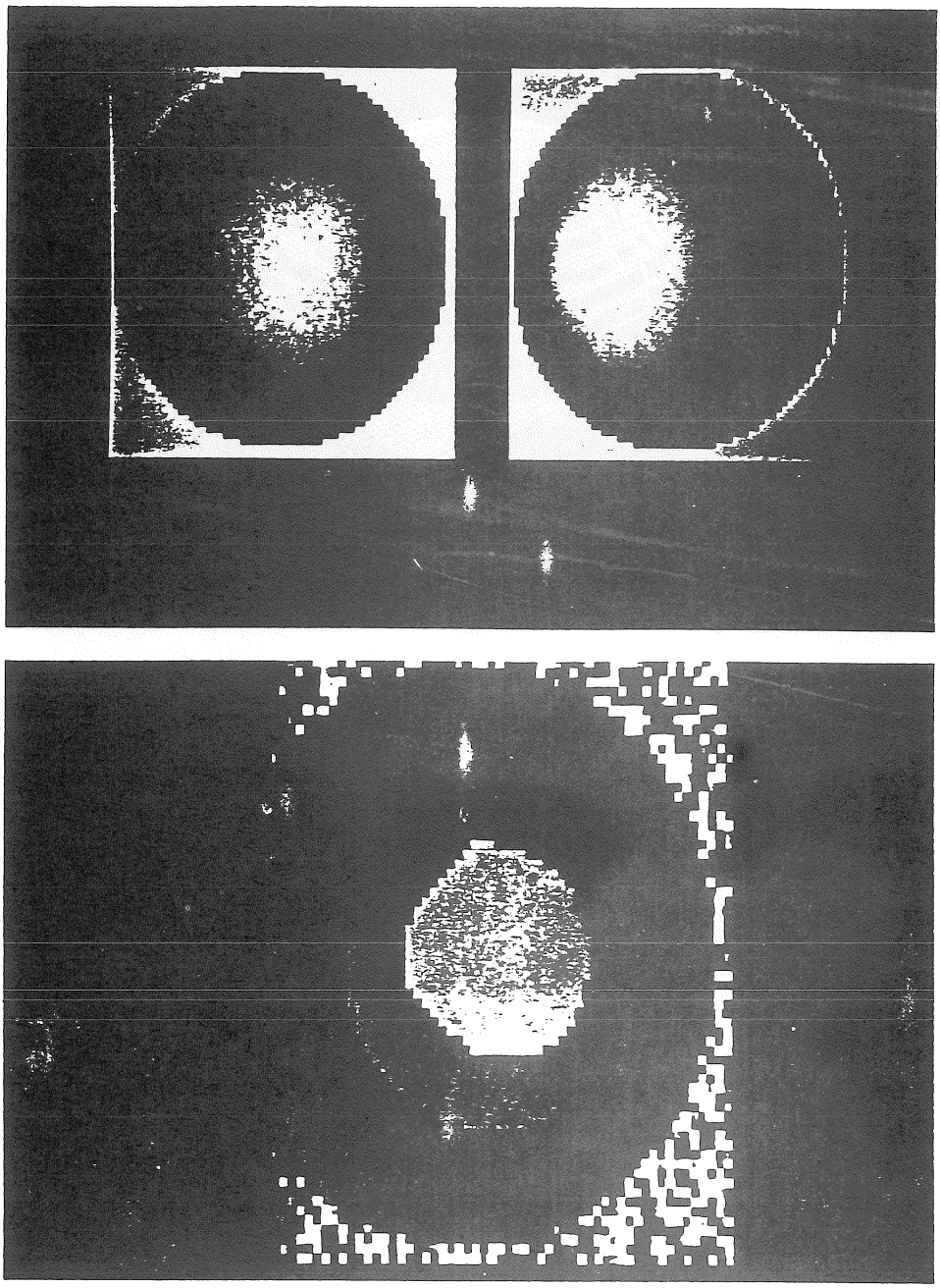

Figure 2a

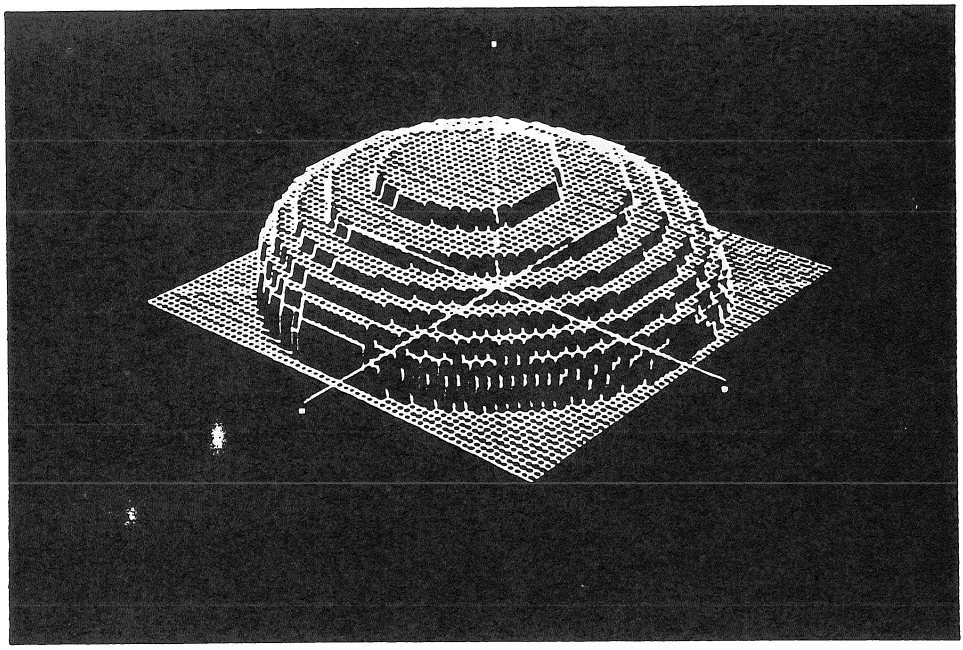

Figure 2b

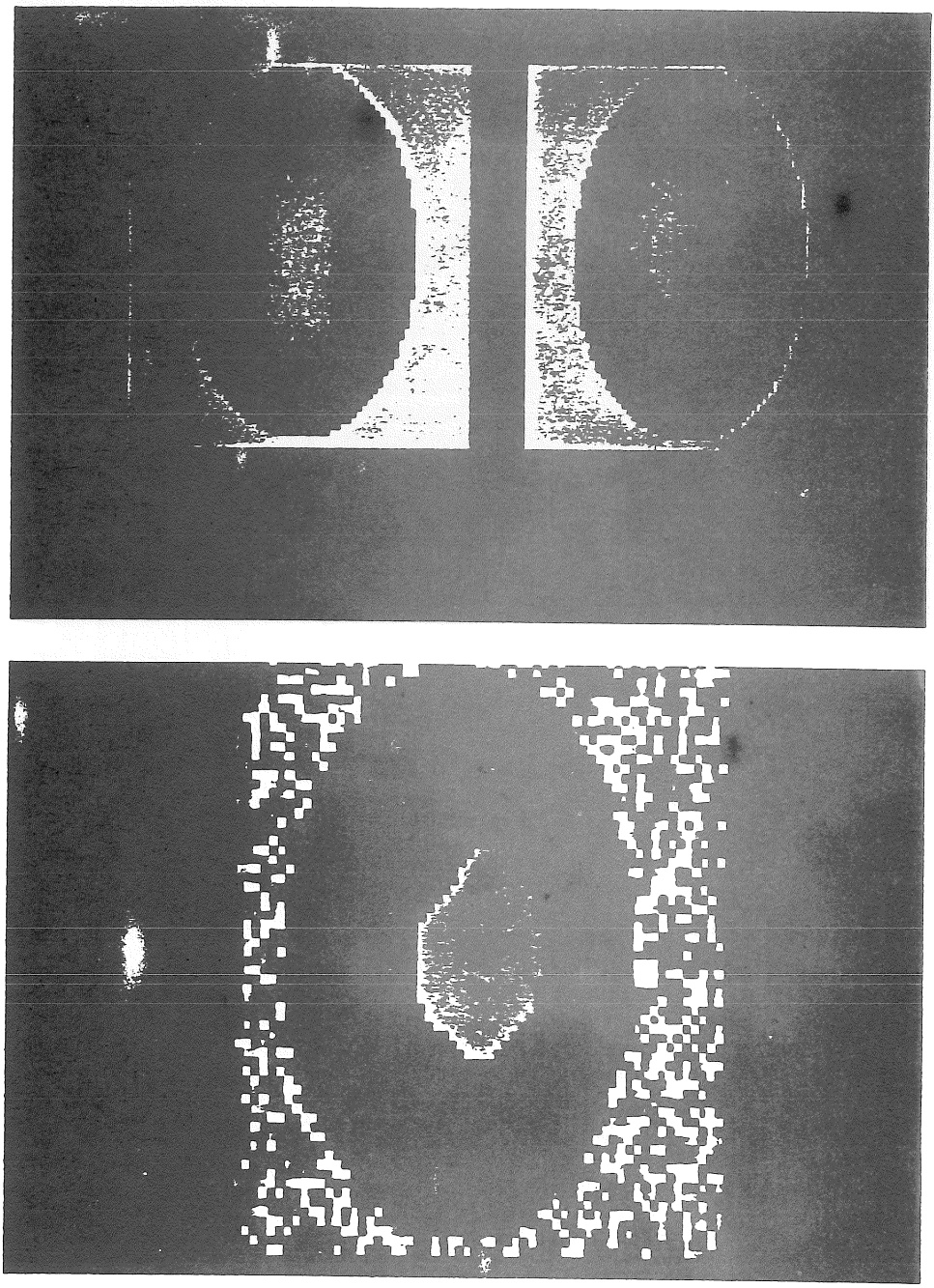

Figure 3a

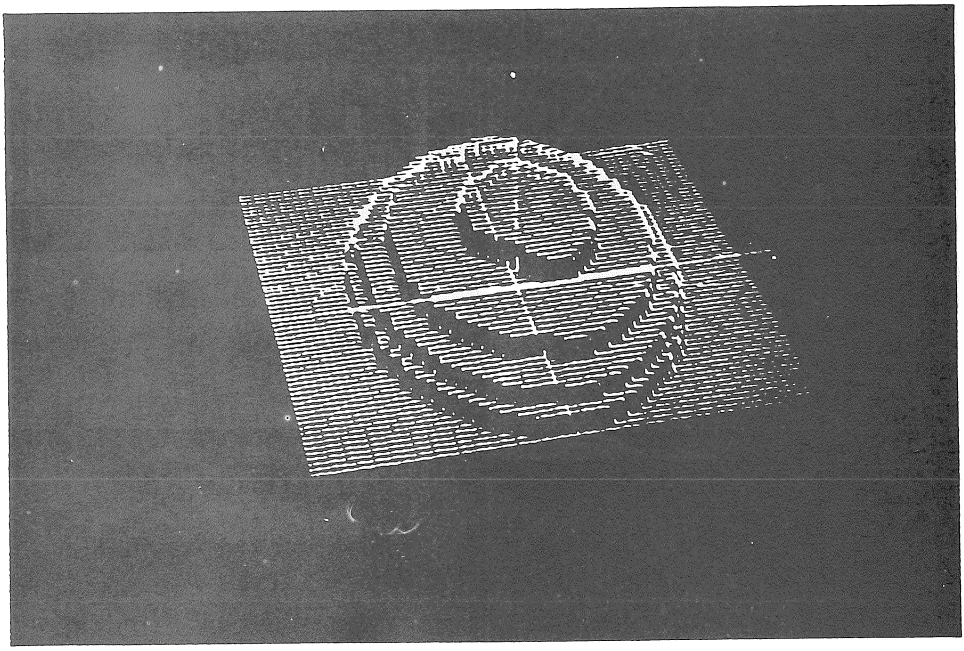

Figure 3b

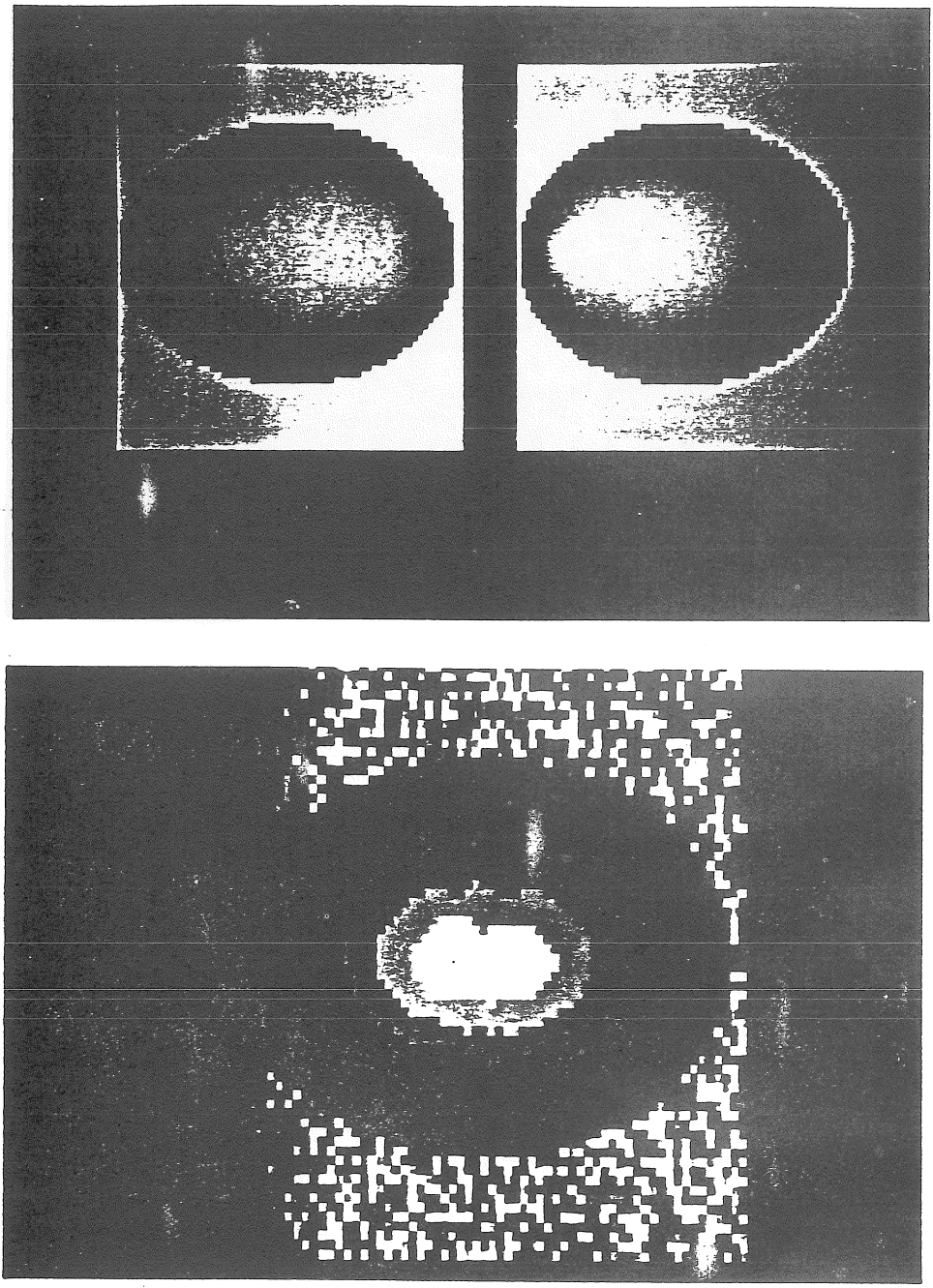

Figure 4a

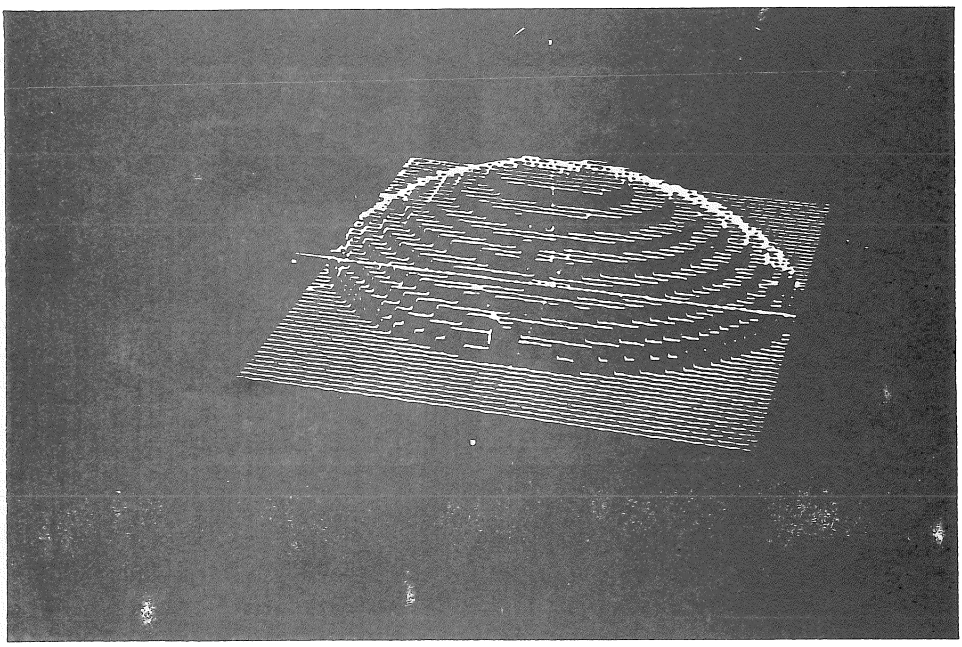

Figure 4b

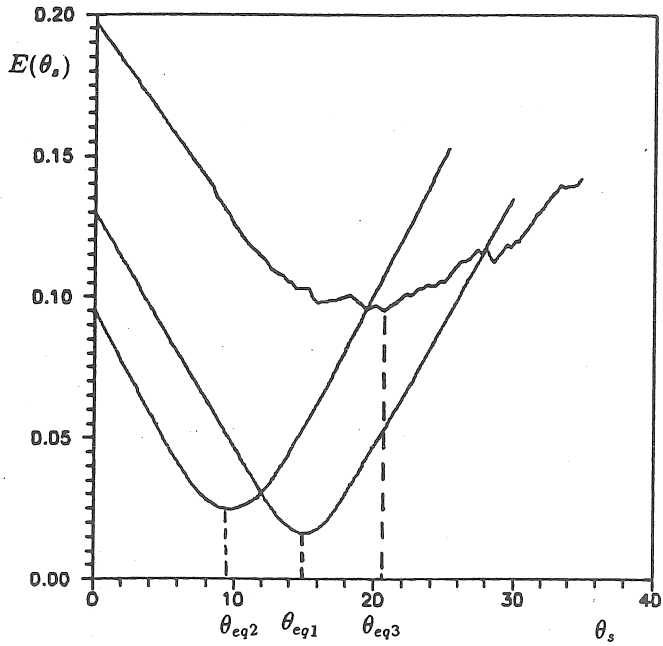
0.20
$E(\theta_s)$
0.15
0.10
0.05
0.00
0
10
20
30
40
θ_{eq2}
θ_{eq1}
θ_{eq3}
θ_s

Figure 5