

**DETERMINING 3-D OBJECT POSE USING
THE COMPLEX EXTENDED GAUSSIAN IMAGE**

**Sing Bing Kang
Katsushi Ikeuchi**

Reprinted from **PROCEEDINGS OF THE 1991 IEEE COMPUTER SOCIETY
CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION,**
Lahaina, Maui, Hawaii, June 3-6, 1991

**Determining 3-D Object Pose Using the Complex
Extended Gaussian Image**

by

Sing Bing Kang and Katsushi Ikeuchi

Determining 3-D Object Pose Using the Complex Extended Gaussian Image

Sing Bing Kang and Katsushi Ikeuchi
The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

This paper describes a new method based on the Extended Gaussian Image (EGI) which can be used to determine the pose of a 3-D object. In this scheme, the weight associated with each outward surface normal is a complex weight. The normal distance of the surface from the predefined origin is encoded as the *phase* of the weight while the magnitude of the weight is the visible area of the surface. This approach decouples the orientation and translation determination into two distinct least-squares problems. Experiments involving synthetic data of two polyhedral and two smooth objects as well as real range data of the same smooth objects indicate the feasibility of this method.

Index Terms: Extended Gaussian Image, gaussian image, pose determination, object representation.

1 Introduction

A fundamental task in most 3-D computer vision and robotic systems is the determination of object pose in space. The pose of an object specifies completely its orientation and position with respect to a predefined frame or coordinate system.

The Extended Gaussian Image (EGI) representation provides a global method of extracting 3-D object orientation. The primary drawback of the EGI is the inability to determine the translation of a recognized 3-D object. This is because the weights in the EGI representation contain only area information and no positional data. One way to encode the positional information is to express the equation of the object face in dual space. This is the approach taken by Roach et al [13] who call the resulting encoded representation the *spherical dual image*. The dual space represents both the orientation and position of the planes or faces of the 3-D object; edges are explicitly described as connections between dual points. However, this scheme is primarily for object representation. Furthermore, planes passing near or through the designated origin cannot be dualized properly; they map to infinity or very large values.

This paper describes a new representation called the Complex Extended Gaussian Image (CEGI) from which both the orientation and translation of a given 3-D object can be determined. In addition, the CEGI has the desirable property of being able to differentiate larger classes of objects than the conventional EGI [8].

1.1 Organization of Paper

Section 2 gives a brief description of the Extended Gaussian Image (EGI) and presents the proposed variant of the EGI, namely the Complex EGI or CEGI. It shows how the distance information can be encoded in the CEGI representation.

The pose recovery strategy of a given object is subsequently presented in Section 3. Emphasis is especially made on how the translation parameters are determined, since this is the main advantage of CEGI over the conventional EGI representation.

Section 4 focuses on the simulations and results on the simulations. A set of simulations is made for both polyhedral and smooth objects to verify the feasibility of the CEGI as a means of extracting the displacement of the object.

Section 5 describes the experiments conducted using real range data and results obtained using the smooth models; final comments and conclusions are presented in Section 6.

1.2 Past Research on the EGI

The EGI has been applied [6, 4] to determine the object attitude, where the rotation in 3-D space brings a sample object into correspondence with a prototype. It has also been used as a means of object recognition [5] in an industrial environment. The EGI of the visible portion of an unknown object is formed by a constrained optimization method applied to data from photometric stereo [7]. The prototype EGI which best matches the partial EGI identifies the object. Little [10] uses a variant of the EGI method which employs the mixed volume as a basis of attitude determination of the sensed object. The mixed volume is a geometric construction used in Minkowski's [11] proof of existence of a convex object given a valid EGI. This paper addresses the deficiency faced by the normal EGI method, namely the inability to recover

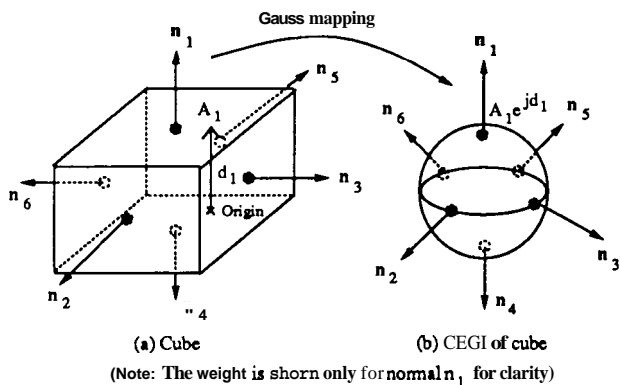


Figure 1: Illustration of the Complex Extended Gaussian Image

translation of objects.

In addition to recognizing objects and determining object attitude, the EGI has also been used to reconstruct convex polyhedra. Ikeuchi [7] has proposed a reconstruction procedure which minimizes the sum of the square differences between the calculated areas of the polyhedron and the given area in the EGI. Little's iterative scheme [9] minimizes the error in the area of the faces as well as in the location of the centroid of the reconstructed figure. More recently, Moni [12] has proposed a reconstruction method which involves determining the adjacency of faces and the length of edges of the polyhedron from the EGI.

Dane and Bajcsy [2] make use of the Gaussian Image to spatially segment a group of range points lying on a surface of a 3-D object into planar and quadric surfaces. The method proposed by Hebert and Ponce [3] segments depth maps into plane, cylindrical and conical primitives. This is done by mapping the estimated surface normals to the EGI and subsequently using the Hough transform to characterize the surfaces.

2 The Complex EGI (CEGI)

2.1 Introduction to the EGI

The EGI of a 3-D object is a histogram which records the variation of surface area with surface orientation. The weights in the EGI representation do not contain any direct distance information. As such, it is translation invariant, and it is easy to see that the EGI representation rotates in the exact manner as the object in space.

2.2 Description of the CEGI

In the conventional EGI representation, each weight associated with the normal of the object face is a scalar which represent the associated visible face area. The CEGI (com-

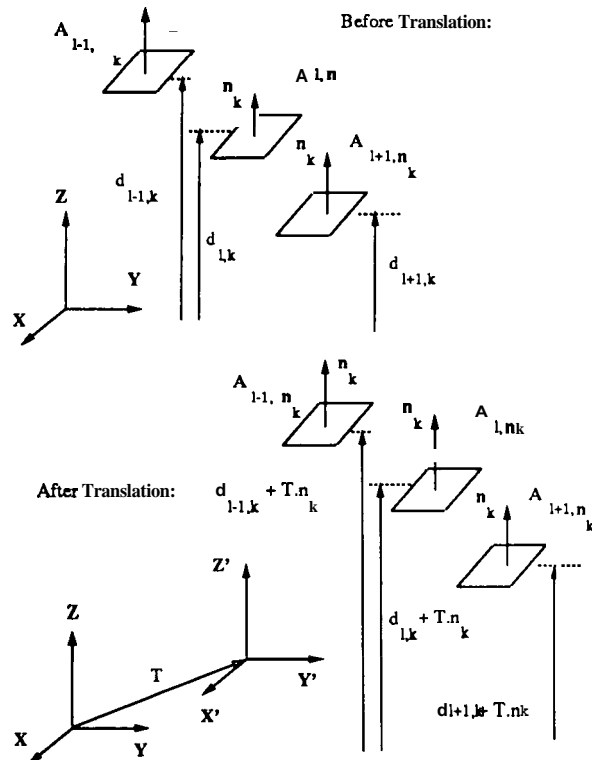


Figure 2: Effect of translating object on the complex weight for particularly oriented faces

plex EGI) concept extends such a representation by adding the normal distance of that face to the origin (in the direction of the normal) as the *phase* component. This is illustrated in Fig. 1. In other words, the weight associated with a particular normal in the CEGI is a complex number whose magnitude is the corresponding visible face area and whose (signed) phase is the normal distance of the face from the designated origin in the direction of the normal. To illustrate further, the complex weight associated with face A is $A_{\hat{n}_k} e^{j d_k}$, where $A_{\hat{n}_k}$ is the area of the plane k with the outward normal \hat{n}_k , and d_k is the normal distance of the plane Π_k (within which $A_{\hat{n}_k}$ lies) to an assigned origin. d_k is positive if the perpendicular vector from the origin to the face is in the same direction as the outward facing normal of the face.

For any given point in the CEGI corresponding to normal \hat{n}_k , the magnitude of the point's weight is $|A_{\hat{n}_k} e^{j d_k}| = A_{\hat{n}_k}$, and thus is independent of the normal distance. If the object is convex, the distribution of $A_{\hat{n}_k}$ corresponds to the conventional EGI representation. If the object is not convex, the magnitude of each weight will not necessarily be equal to those of the corresponding conventional EGI. The translation invariance property of the weight magnitude applies even if

there are more than one contiguous surface patches with the same outward normal. Consider surfaces whose normals are \hat{n}_k (Fig. 2). Before translation, the corresponding complex weight is

$$\mathbf{P}_{\hat{n}_k} = \sum_{l=1}^{N_k} A_{l, \hat{n}_k} e^{j d_{lk}} \quad (1)$$

After a translation along a vector \mathbf{T} , the complex weight becomes

$$\mathbf{P}'_{\hat{n}_k} = \sum_{l=1}^{N_k} A_{l, \hat{n}_k} e^{j(d_{lk} + \mathbf{T} \cdot \hat{n}_k)} = e^{j \mathbf{T} \cdot \hat{n}_k} \mathbf{P}_{\hat{n}_k} \quad (2)$$

For each point in the CEGI, the magnitude of the weight is thus independent of the translation. However, there exists an ambiguity range $(-\mathbf{a}, \mathbf{a}]$ beyond which errors would occur. In our method, all distances are normalized such that the greatest expected change in distance is π .

In short, a translational change in object position does not affect the weight magnitude distribution of its CEGI. Hence by comparing the magnitudes, we can then recognize objects and determine their orientations as we would for a conventional EGI. Subsequently, by comparing the differences in the complex weight phases, we can proceed to calculate the distance change along the normals.

3 Pose Determination Strategy

3.1 Methodology

Given a prototype CEGI and a partial CEGI of an unknown object, we can recognize the object and determine its orientation by the following: first, calculate the magnitude distributions of both CEGI's and second, proceed as one would for the conventional EGI's. Once both the object and its orientation with respect to the stored model are recognized, the object translation can be calculated by using the suitably oriented CEGI's.

The translation parameters can be determined by applying a least-squares technique as follows: Suppose the object has been translated by δx , δy and δz in the x -, y - and z - directions respectively (in model world coordinates). Then for each surface whose surface normal is \hat{n}_k , and whose complex weight is originally $A_{\hat{n}_k} e^{j d_k}$, after undergoing translation $\delta \mathbf{d}$, the complex weight becomes $A_{\hat{n}_k} e^{j(d_k + \delta \mathbf{d} \cdot \hat{n}_k)}$, where

$$\begin{aligned} \delta \mathbf{d} &= \delta x \hat{\mathbf{i}} + \delta y \hat{\mathbf{j}} + \delta z \hat{\mathbf{k}} \\ \hat{n}_k &= n_{kx} \hat{\mathbf{i}} + n_{ky} \hat{\mathbf{j}} + n_{kz} \hat{\mathbf{k}} \end{aligned} \quad (3)$$

Then for each matched weight, let

$$\begin{aligned} \omega_i &= \arg\left(\frac{\mathbf{P}'_{\hat{n}_k}}{\mathbf{P}_{\hat{n}_k}}\right) = \arg\left(\frac{A_{\hat{n}_k} e^{j(d_k + \delta \mathbf{d} \cdot \hat{n}_k)}}{A_{\hat{n}_k} e^{j d_k}}\right) \\ &= \delta x n_{ix} + \delta y n_{iy} + \delta z n_{iz} \\ &\text{for } i = 1, \dots, N_{\text{visible}} \end{aligned} \quad (4)$$

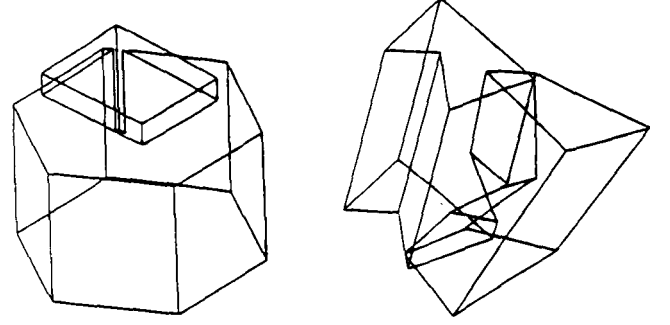


Figure 3: Left: First composite object used for testing; Right: Second composite object used for testing

where N_{visible} is the total number of visible faces on the object. We try to minimize the total squared error given by

$$\mathcal{E} = \sum_{i=1}^{N_{\text{visible}}} (\omega_i - n_{ix} \delta x - n_{iy} \delta y - n_{iz} \delta z)^2 \quad (5)$$

The optimal values of δx , δy and δz are

$$\delta x = \frac{1}{\mathbf{D}} \begin{vmatrix} \sum \omega_i n_{ix} & \sum n_{ix} n_{iy} & \sum n_{ix} n_{iz} \\ \sum \omega_i n_{iy} & \sum n_{iy}^2 & \sum n_{iy} n_{iz} \\ \sum \omega_i n_{iz} & \sum n_{iy} n_{iz} & \sum n_{iz}^2 \end{vmatrix} \quad (6)$$

$$\delta y = \frac{1}{\mathbf{D}} \begin{vmatrix} \sum n_{ix}^2 & \sum \omega_i n_{ix} & \sum n_{ix} n_{iz} \\ \sum n_{ix} n_{iy} & \sum \omega_i n_{iy} & \sum n_{iy} n_{iz} \\ \sum n_{ix} n_{iz} & \sum \omega_i n_{iz} & \sum n_{iz}^2 \end{vmatrix} \quad (7)$$

$$\delta z = \frac{1}{\mathbf{D}} \begin{vmatrix} \sum n_{ix}^2 & \sum n_{ix} n_{iy} & \sum \omega_i n_{ix} \\ \sum n_{ix} n_{iy} & \sum n_{iy}^2 & \sum \omega_i n_{iy} \\ \sum n_{ix} n_{iz} & \sum n_{iy} n_{iz} & \sum \omega_i n_{iz} \end{vmatrix} \quad (8)$$

where

$$\mathbf{D} = \begin{vmatrix} \sum n_{ix}^2 & \sum n_{ix} n_{iy} & \sum n_{ix} n_{iz} \\ \sum n_{ix} n_{iy} & \sum n_{iy}^2 & \sum n_{iy} n_{iz} \\ \sum n_{ix} n_{iz} & \sum n_{iy} n_{iz} & \sum n_{iz}^2 \end{vmatrix} \quad (9)$$

In order for this scheme to work, the magnitude of the translation (i.e the normalized translation) must be less than as given in (10).

$$A = \sqrt{(\delta x)^2 + (\delta y)^2 + (\delta z)^2} < \mathbf{a} \quad (10)$$

This is because the complex weight is unique as long as the change in phase (and hence the change in surface normal distance from the origin) lies within the principal interval $(-\pi, \pi]$.

M#	$\overline{\epsilon_{d_x}}$	$\sigma_{\epsilon_{d_x}}$	$\overline{\epsilon_{d_y}}$	$\sigma_{\epsilon_{d_y}}$	$\overline{\epsilon_{d_z}}$	$\sigma_{\epsilon_{d_z}}$	$\overline{\epsilon_{d_{\text{syn}}}}$
1	1.6	1.2	3.4	3.8	12.2	15.9	14.1
2	2.5	1.8	2.6	1.8	2.1	1.7	4.7

Table 1: Results for Simulations using Polyhedral Objects

4 Simulations

4.1 Experiments with Polyhedral Objects

Two models are used to test the concept of translation parameter extraction using the CEGI representation; they are shown in Fig. 3.

4.1.1 Implementational Issues

The scheme was implemented in Lisp, with the models generated using VANTAGE. VANTAGE is a geometric/sensor modeler developed at Carnegie Mellon University [1]. The CEGI viewing sphere is discretized into 240 sampling view directions located at the center of each face of the tessellated pentakis dodecahedron. The normal direction space is discretized into 240 cells as well. The CEGI weights are recomputed for each discrete view direction. This is to compensate for the varying degrees of self-occlusion which causes non-convex objects to register different weights at different view directions for the same surface normal.

4.1.2 Results

Table 1 summarizes the simulation results for the two polyhedral objects. All the errors are expressed in percentages of the maximum displacement. (Note: $\overline{\epsilon_K}$ denotes the average error in K while σ_{ϵ_K} is the standard deviation of the error distribution.)

It can be readily seen from Table 1 that Model 1 yields significantly higher predicted distance errors than Model 2. For Model 1, if the viewpoint subtends a very small angle to the either the z-axis or the x-y plane, the errors incurred in the recovery will be very high; this is attributed to the sparse and uneven distribution of Model 1's normals. The lower errors for Model 2 are attributed to its higher number of different surface normals and their more even distribution.

4.2 Experiments with Smooth Objects

In addition to the two polyhedral objects, two smooth objects were also used in simulations and experiments involving real range data. The two chosen smooth objects are the torus and the ellipsoid (Fig. 4).

4.2.1 Simulation

Implementational Issues

The torus is modeled parametrically as $[(R+r \cos \phi) \cos \phi, (R+r \cos \phi) \sin \phi, r \sin \theta]^T$ while the ellipsoid is represented parametrically as $[a \sin \theta \cos \phi, a \sin \theta \sin \phi, b \cos \theta]^T$. The parametric values used are: $r=20, R=40, a=20, b=40, |O_I O|=100$,

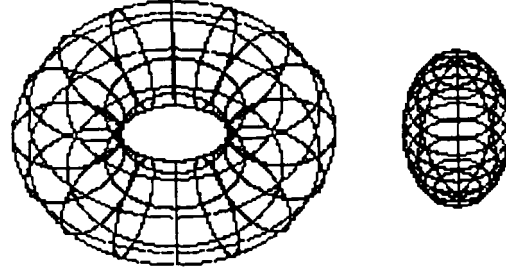


Figure 4: Models used in experiments: torus (left) and ellipsoid (right)

and $|V_c O|=1000$. O, O_I and V_c are the object, image and focal centers respectively. These centers are colinear. Simulations are performed for image resolutions $n \times n$ for $n = 32, 64$ and 128 . To achieve a certain degree of realism, a simple ray-tracing technique is employed to estimate the object surface area projected onto each pixel as well as the surface normal to be attributed to that surface.

Results

The simulation results are graphically depicted in Fig. 5 and listed in Table 2. The number of runs for each resolution per object is 250. For both of these models, it is apparent that the errors in the predicted displacement decreases monotonically as the resolution increases.

We also observe that the displacement errors incurred for the torus are significantly higher (a factor of about 2-3) than those incurred for the ellipsoid. This is because, for the ellipsoid, the resultant complex weight in each CEGI cell is the sum of those weights corresponding to surface patches which are either contiguous or spatially close to each other. On the other hand, for the non-convex torus, this is not true. Two spatially distinct groups of surface patches, whose normal distances differ greatly, may contribute to a cell in the CEGI. The phase of the resultant complex weight is thus expected to have higher variability than that for the ellipsoid, for the same amount of phase error per surface patch.

5 Experiments Using Range Data

To confirm the validity of the simulations, experiments involving real range data of the torus and ellipsoid were conducted.

5.1 Implementational Issues

A torus and an ellipsoid were crafted out of clay to resemble the models whose databases were created. A light-stripe range finder was used to produce a range image of these two objects, each in three different poses at two different resolutions (a total of 12 images). The lower resolution images (Set

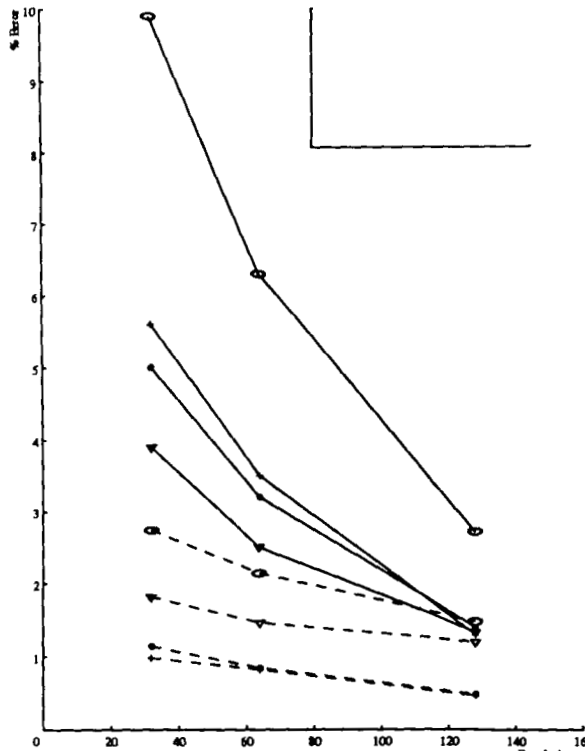


Figure 5: Simulation Results for Torus and Ellipsoid

1) have spacings of about 3mm in the x- and y- directions. The resolution of Set 2 range images is twice that of Set 1 images. The accuracy in the z-direction is 0.1mm.

5.2 Results

The experimental results are summarized in Table 3. The figures are quoted in percentages of the maximum allowable displacement. In this case, all the numbers are in millimeters.

Again, as for the simulation results, the displacement errors are smaller for Set 2 experiments (which feature higher resolution range images) for both objects. The errors in the translation parameters are significantly higher (this time by a smaller factor of 1.5-2) for the torus than those for the ellipsoid. Despite the fact that the clay models are not exactly the same as the models created in the database, reasonable accuracy could still be attained. Figs. 6 and 7 depict calculated model positions superimposed on actual object positions for the torus and ellipsoid respectively.

6 Conclusions

A new variant of the EGI representation which encodes object face position has been described. Known as the Complex Extended Gaussian Image (CEGI), it is a histogram of spatial orientation in which each weight associated with a normal is a complex number. The normal distance of the face from the predefined origin is encoded as the phase of the weight while

Res.	$\overline{\epsilon_{d_x}}$	$\sigma_{\epsilon_{d_x}}$	$\overline{\epsilon_{d_y}}$	$\sigma_{\epsilon_{d_y}}$	$\overline{\epsilon_{d_z}}$	$\sigma_{\epsilon_{d_z}}$	$\overline{\epsilon_{d_{xyz}}}$	$\sigma_{\epsilon_{d_{xyz}}}$
32 ²	5.0	5.5	5.6	6.1	3.9	3.3	9.9	7.3
64 ²	3.2	3.2	3.5	3.7	2.5	2.0	6.3	4.1
128 ²	1.4	1.5	1.3	1.2	1.3	1.1	2.7	1.7
32'	1.1	1.1	1.0	0.8	1.8	1.5	2.7	1.5
64'	0.8	0.8	0.8	0.7	1.5	1.1	2.1	1.1
128'	0.5	0.4	0.5	0.3	1.2	1.0	1.5	0.9

Table 2: Simulation Results (Top three lines - torus; Bottom three lines - ellipsoid)

C#	Lower Resolution				Higher Resolution			
	$\overline{\epsilon_{d_x}}$	$\overline{\epsilon_{d_y}}$	$\overline{\epsilon_{d_z}}$	$\overline{\epsilon_{d_{xyz}}}$	$\overline{\epsilon_{d_x}}$	$\overline{\epsilon_{d_y}}$	$\overline{\epsilon_{d_z}}$	$\overline{\epsilon_{d_{xyz}}}$
T1	1.7	5.9	5.0	8.5	1.4	2.9	4.0	5.4
T2	2.6	4.1	7.3	9.5	4.2	2.5	5.6	7.7
T3	3.1	4.2	3.6	6.6	3.2	1.4	2.9	5.5
E1	0.9	1.4	3.2	3.7	0.9	3.1	1.6	3.6
E2	0.4	2.1	4.9	5.4	0.9	4.1	0.2	4.2
E3	1.3	2.7	2.1	3.9	1.6	0.5	1.9	2.6

Table 3: Experimental Results (T1-T3, E1-E3 - cases involving the torus and ellipsoid respectively)

the magnitude of the weight is the visible area of the face.

The CEGI provides a global method of extracting the pose of a detected object with respect to a stored model. It effectively decouples the orientation and translation determination into two separate problems. The orientation of the object can be determined by first calculating the magnitude distribution of its CEGI before matching the resulting distribution with those in the database. This operation is exactly the same as that using the conventional EGI. The translation parameters can subsequently be estimated by comparing the complex weight phases.

A significant advantage of this scheme is that it works for both polyhedral and smooth objects. It can be used without the need to know the type of object a priori. In addition, it is capable of distinguishing between the convex object and most other concave objects whose EGI's are identical to that of the convex object [8] (not shown in this paper due to lack of space).

Acknowledgments

Many thanks to Martial Hebert, Mark Wheeler and Kathryn Porsche for proofreading draft copies of this paper and for their constructive comments. This research was sponsored by DARPA under Contract F33615-87-C-1499.

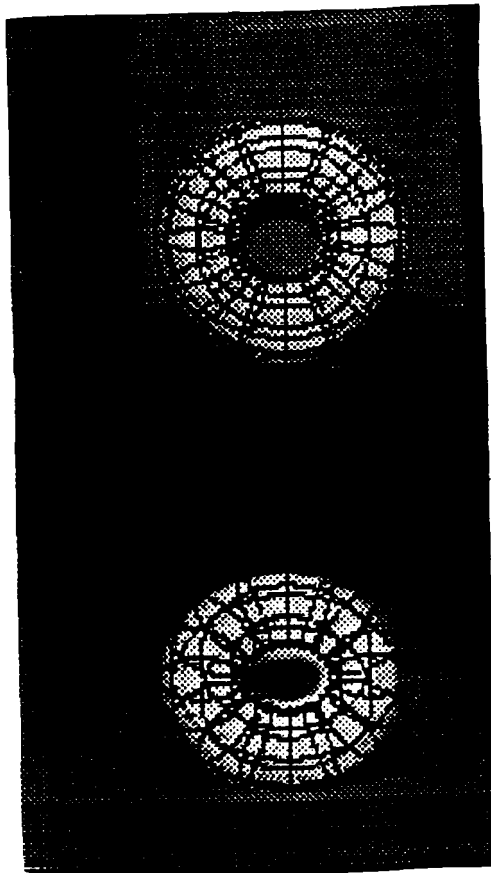


Figure 6: Top: Superimposed model (Set 2, Case 1); Bottom: Superimposed model (Set 2, Case 2)

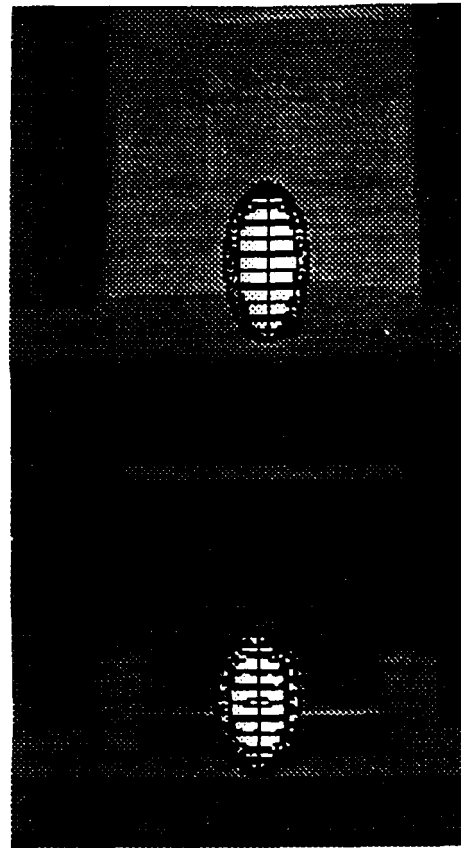


Figure 7: Top: Superimposed model (Set 2, Case 1); Bottom: Superimposed model (Set 2, Case 2)

References

- [1] Balakumar, P. et al. *VANTAGE: A Frame-Based Geometric Modeling System - Programmer/User's Manual*. Carnegie Mellon Univ., December 1988.
- [2] C. Dane and R. Bajcsy. Three-Dimensional Segmentation Using the Gaussian Image and Spatial Information. *Proc. Int'l Conf. on Pattern Recog. and Image Proc.*, pages 54–56, August 1981.
- [3] M. Hebert and J. Ponce. A New Method for Segmenting 3-D Scene into Primitives. *Proc. Int'l Conf. on Pattern Recog.*, pages 836–838, October 1982.
- [4] B.K.P. Horn and K. Ikeuchi. The Mechanical Manipulation of Randomly Oriented Parts. *Scientific American*, August 1984.
- [5] K. Ikeuchi. Recognition of 3-D Objects Using the Extended Gaussian Image. In *Proc. of Seventh IJCAI*, pages 595–600, 1981.
- [6] K. Ikeuchi. Determining Attitude of Object from Needle Map Using Extended Gaussian Image. Technical Report AI Memo 714, MIT, 1983.
- [7] K. Ikeuchi and B.K.P. Horn. Numerical Shape from Shading and Occluding Boundaries. *Artificial Intelligence*, 17:141–185, 1981.
- [8] S.B. Kang and K. Ikeuchi. 3-D Object Pose Determination Using Complex EGI. Technical Report CMU-RI-TR-90-18, Carnegie Mellon Univ., October 1990.
- [9] J.J. Little. An Iterative Method for Reconstructing Convex Polyhedra from Extended Gaussian Images. *Proc. of the National Conf. on AI*, pages 247–250, August 1983.
- [10] J.J. Little. Recovering Shape and Determining Attitude from Extended Gaussian Image. Technical Report TR 85-2, Univ. of British Columbia, April 1985.
- [11] H. Minkowski. Allgemeine Lehrsätze über die konvexe Polyeder. *Nachr. Ges. Wiss. Göttingen*, 1897.
- [12] S. Moni. A Closed-Form Solution for the Reconstruction of a Convex Polyhedron from its Extended Gaussian Image. *Int'l Conf. on Pattern Recog.*, 1:223–226, June 1990.
- [13] J.W. Roach, J.S. Wright, and V. Ramesh. Spherical Dual Images: A 3D Representation Method for Solid Objects that Combines Dual Space and Gaussian Spheres. *Proc. CVPR*, pages 236–241, June 1986.

