The Complex EGI: A New Representation for 3-D Pose Determination

Sing Bing Kang and Katsushi Ikeuchi, *Member, IEEE*

***Abstract*—This paper describes a new 3-D object representation that can be used to determine the pose of an object. In this representation (called the complex extended Gaussian image (CEGI)), 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, whereas 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. The justification for using such a scheme is twofold: It not only allows the pose of the object to be extracted, but it also distinguishes a convex object from a nonconvex object having the same EGI representation. The CEGI scheme has the advantage of not requiring explicit spatial object-model surface correspondence in determining object orientation and translation.**

Experiments involving synthetic data of two polyhedral and two smooth objects indicate the feasibility of this method. The best results are 4.7 and 1.5% (total distance error) for the polyhedral and smooth objects, respectively. The figures are quoted in terms of percentages of the maximum allowable displacement. Experiments using real range data for the two smooth objects yielded reasonably good results.

***Index Terms*—Extended Gaussian image, object pose estimation, object recognition, object representation, viewing sphere, view suitability index.**

I. 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.

Research on 3-D object pose determination has been intensive, but with varied results, as the comprehensive survey by Besl and Jain [2] indicates. Much of the work on 3-D pose extraction focuses on polyhedral objects. Pose determination of smooth objects is significantly more difficult. However, this problem is of particular interest because the world around us is composed of objects that are usually *not* piecewise planar but rather smooth and continuous in nature.

Nevatia and Binford [20] match curved objects using generalized cylinders. Faugeras and Hebert [9] and Faugeras [8] have proposed a 3-D object recognition algorithm based on geometrical matching between primitive surfaces. The primitive surface that was actually implemented is the plane, although quadric surface algorithms are presented as well. The best accuracy quoted in [8] is 2.3° for the rotation angle and 3 mm for the translation. The accuracy of the range data is 1 mm. Bhanu [3] uses planar polygonal faces to represent the object and employs a relaxation-based scheme called *stochastic face labeling* to match object shape. The primitives that Stein and Medioni [25] use to represent the object for recognition and pose estimation are small surface patches (which they term *splashes*) and 3-D curves corresponding to depth or orientation discontinuities. Bolle and Cooper [4] present a probabilistic approach to 3-D pose estimation of a known object. The surface of an object is modeled as a collection of planar, cylindrical, and spherical patches, which is used as the basis for their proposed locally optimum surface parameter estimation and roughly globally optimum object pose estimation.

Bolles *et al.* [5] developed 3DPO, which recognizes and locates 3-D objects using range data. The system's hypothesis generation and matching is based on matching of several features or feature clusters involving object-specific features such as a circular arc of a specific radius and edges. Fan *et al.* [7], on the other hand, use surface descriptions such as jump boundaries, creases, and limbs to match and locate 3-D objects. No accuracy figures were supplied in these two papers.

A recent paper by Ponce and Kreigman [21] describes a method to recognize and locate curved objects in a monocular intensity image. They consider the image contours (namely, the projections of surface discontinuities and occluding contours) as the basis for recognition and location of the object. However, this assumes that the contour equations are parametrically known. Furthermore, the location of the object is estimated in the image plane only.

Pose estimation of curved objects is usually achieved using local features such as the object boundary and edges. The object surface itself is usually not used in the pose-determination process. The EGI representation, on the other hand, uses surface information. It has been proposed to avoid the more difficult problem of local feature matching by directly extracting the object surface area distribution with the surface normal. Recognition of the object is then achieved by matching this distribution with those of database models.

A. Previous Work on the EGI

The principal advantage of the EGI is its position invariance. This reduces the object recognition and pose determination

Manuscript received April 14, 1991; revised April 23, 1992. This work was sponsored in part by the Avionics Laboratory, Wright Research and Development Center, Aeronautical Systems Division (AFSC), U.S. Air Force, Wright-Patterson AFB, OH 45433-6543 under F33615-90-C-1465, ARPA Order No. 7597. Recommended for acceptance by Associate Editor R. Bolle.

The authors are with The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213.

IEEE Log Number 9208012.

problems into two three degree-of-freedom problems, namely, orientation determination and translation recovery (which is the second problem to be solved using some other method if the conventional EGI is used). It dispenses with the more difficult problem of identifying and matching local features; techniques for this generally depend on whether the object of interest is smooth (e.g., [21]) or polyhedral (e.g., [10], [24]). On the other hand, many local methods of object recognition and pose determination do tolerate some degree of occlusion (e.g., [9], [10], [17], [24]). The recognition scheme involving the EGI (or CEGI) fails in the presence of significant occlusion by other objects.[1]

The EGI has been applied [6], [12], [11[, [14] 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 [13] in an industrial environment. In this approach, look-up tables of partial EGI's of each model at different viewpoints (which are suitably spaced apart on the Gaussian sphere) are computed to avoid errors due to object self occlusion. The prototype partial EGI that best matches the partial EGI of the unknown object identifies it and its orientation. Little [16] uses a variant of the EGI method that 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 [18] proof of existence of a convex object given a valid EGI.

The primary advantage of the EGI is also its primary drawback: Because of its translation invariance, the translation of a recognized 3-D object cannot be recovered. The EGI is translation invariant because the weights in its representation do not contain positional data.

To eliminate this deficiency, Nalwa [19] proposes that surfaces be represented by their Gaussian images augmented by the support function. This support function is the signed distance of the oriented tangent plane from a predefined origin. He proposes to ascribe to each different surface a separate support function value. This means that in general, the proposed variant of the Gauss map of a surface is not globally one to one. Although it is less compact, this support representation (which is dependent on the choice of the origin) can uniquely determine a surface. A method to determine object pose based on this representation was not presented in his paper. Roach *et al.* [22] encode positional information by expressing the equation of the object face in dual space. The resulting encoded representation is called the *spherical dual image*.[2] A point in the dual space represents both the orientation and position of a plane or face 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.

The proposed representation (CEGI) addresses the deficiency of the EGI by having distance encoded in its weights in a different manner. This representation allows both the orientation and translation of a given 3-D object to be determined separately. In addition, the CEGI has the desirable property of being able to differentiate larger classes of objects than the conventional EGI. Another attractive feature of the proposed representation is that the approach in extracting the pose of an object is identical for both polyhedral and smooth objects.

B. Brief Description of the Complex EGI

The weight at each discrete cell of the CEGI is a complex number. The complex weight associated with a particular surface patch is determined as follows: The magnitude of the weight is the visible surface area of the object associated with its surface normal.[3] The phase of each weight is the key to displacement determination; it is the (signed) distance of the surface patch from a designated origin in the direction of its normal. In summary, the support function associated with the CEGI is a complex number that encodes both the area and distance.

By encoding the distance information this way, it can be easily shown that the magnitude of the total weight at each cell of the CEGI is independent of object position. Thus, orientation can be determined in a manner identical to that for the conventional EGI. Once the orientation is found, the translation parameters can then be calculated by comparing the phases of the complex weights at matched cells of the model CEGI and the partial object CEGI.

C. Organization of Paper

Section II gives a brief description of the EGI and presents the proposed variant of the EGI, namely, the 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 III. Emphasis is especially made on how the translation parameters are determined since this is the main advantage of CEGI over the conventional EGI representation. In addition, error analysis is made on the recovered translation factors. We also justify using a view suitability index as a measure of the translation parameter error bound.

Section IV focuses on the results of simulations performed. There are two parts to this: In the first part, 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. This is subsequently followed by the results obtained for the simulations made using the two polyhedral objects to verify the linear upper bound of translation errors.

The following section (Section V) describes the experiments, which were conducted using real range data extracted from the smooth objects, and their results. These experiments were performed to further validate and test the concept of translation extraction using the CEGI.

Final comments and conclusions are presented in Section VI.

[1] Object self-occlusion, however, is *not* a problem.

[2] This is conventionally known as the tangent bundle.

[3] Note that this is true for *a single* surface patch only. One can easily see that this is no longer true in general if we add two complex weights representing two discrete surfaces having the same normal.

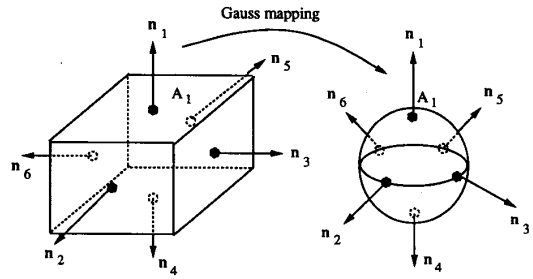

Fig. 1. Illustration of the EGI.

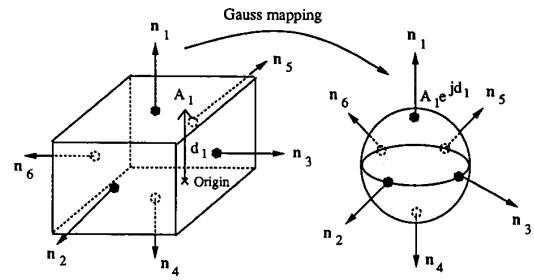

Fig. 2. Illustration of the CEGI.

II. The Complex EGI (CEGI)

A. Introduction to the EGI

The EGI of a 3-D object is a histogram that 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. The EGI of a cube is shown in Fig. 1.

B. Description of the CEGI

In the conventional EGI representation, each weight associated with the normals of the object face are scalars that represent the associated visible face area. The CEGI 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 in its support function. This is illustrated in Fig. 2. 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, in Fig. 3, the complex weight associated with face **A** is $A_{\hat{\mathbf{n}}_k} e^{jd_k}$, where $A_{\hat{\mathbf{n}}_k}$ is the area of face **A** with the outward normal $\hat{\mathbf{n}}_k$, and d_k is the normal distance of the plane Π_k (within which $A_{\hat{\mathbf{n}}_k}$ 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. The value of d_k is positive in this case (Fig. 3).

For any given point in the CEGI corresponding to normal $\hat{\mathbf{n}}_k$, the magnitude of the point's weight is $|A_{\hat{\mathbf{n}}_k} e^{jd_k}|$. $A_{\hat{\mathbf{n}}_k}$ is independent of the normal distance, and if the object is convex, the distribution of $A_{\hat{\mathbf{n}}_k}$ 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. This is an important attribute of the CEGI, which will be further described in Section II-C.

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 the surfaces whose normals are $\hat{\mathbf{n}}_k$ shown in Fig. 4 (the distribution $d_{1,k}, \ldots, d_{N_k,k}$ is henceforth referred to as the *surface normal distance distribution* or, simply, the *distance distribution*). Before translation, the corresponding complex weight is

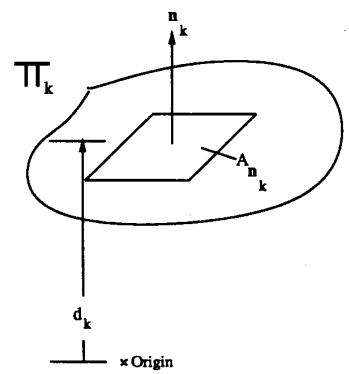

Fig. 3. Example of an oriented face with a complex weight $A_{\mathbf{n}_k} e^{jd_k}$.

$$\mathbf{P}_{\hat{\mathbf{n}}_k} = \sum_{l=1}^{N_k} A_{l,\hat{\mathbf{n}}_k} e^{jd_{l,k}}. \tag{1}$$

After a translation along a vector **T**, the complex weight becomes

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

Hence, for each point in the CEGI, the magnitude of the weight is independent of the translation. The complex number folds back onto itself for every translation distance of 2π. Consequently, there exists an ambiguity range beyond which errors would occur. In our method, all distances are normalized such that the greatest expected change in translation distance ($\max_k(\mathbf{T} \cdot \hat{\mathbf{n}}_k)$ in (2)) is π.

Note that the normal distance of the surface from a predefined origin is not constrained to lie in the interval $(-\pi, \pi]$, whereas the relative object translation is constrained to fall in this interval. To illustrate, refer to Fig. 5. The detected object is shown in solid lines; the dashed lines indicate the initial estimation of the object pose (model pose) in the neighborhood of the detected object. For any arbitrary object surface A_k and model surface $A_{M,k}$ whose normal is $\hat{\mathbf{n}}_k$, let the respective normal distances be d_k and $d_{M,k}$. Although d_k and $d_{M,k}$ are

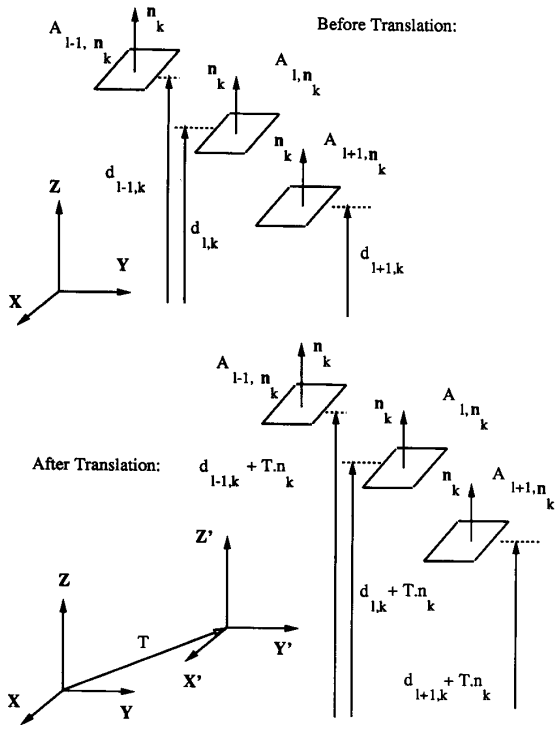

Fig. 4. Effect of translating object on the complex weight for particularly oriented faces.

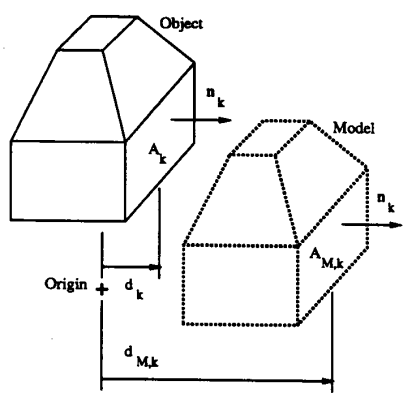

Fig. 5. Object translation relative to the model.

not restricted to lie in the interval $(-\pi, \pi]$, their difference $d_k - d_{M,k}$ is constrained within this interval. This constraint does not constitute a problem for the CEGI representation because an approximate location of the object can be easily determined from the image itself (e.g., range image).

C. Some Properties of the CEGI Representation

Property 1: The magnitude distribution of the CEGI representation is translation invariant.

Proof: This has been shown in Section II-B. □

Property 2: Suppose a convex object (denoted by $\mathcal{O}_C$) and a nonconvex object (denoted by $\mathcal{O}_{NC}$) have identical EGI representations. The CEGI representations of $\mathcal{O}_C$ and $\mathcal{O}_{NC}$ are different.

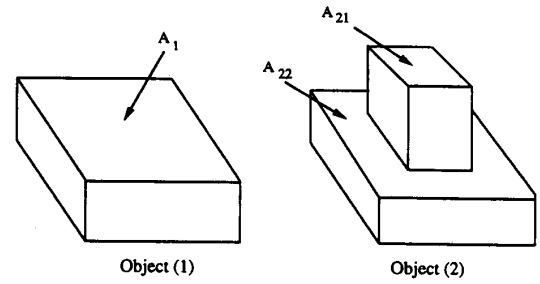

Fig. 6. Example of two objects having identical EGI's.

Proof: It is possible for a convex object and a class of nonconvex objects to have identical EGI representations. The CEGI representation has the property of being able to differentiate the convex object from that class of nonconvex objects whose EGI representations are identical. Examples of two such objects are shown in Fig. 6.

Suppose these two objects (one convex and the other nonconvex) have identical EGI's. Comparing the weights associated with normal $\hat{\mathbf{n}}$ (e.g., the positive z direction normal in Fig. 6), we have for the convex object (e.g., object (1) in Fig. 6)

$$P_{1,\hat{\mathbf{n}}} = A_1 e^{jd_1}. \tag{3}$$

For the nonconvex object in general ($n = 2$ for object (2) in Fig. 6)

$$P_{2,\hat{\mathbf{n}}} = \sum_{i=1}^{n} A_{2,i} e^{jd_{2,i}}. \tag{4}$$

Since they have identical EGI's, $A_1 = \sum_{i=1}^{n} A_{2,i}$. If these CEGI weights (specified by (3) and (4)) were to be equal, then

$$\begin{aligned}(\sum_{i=1}^{n} A_{2,i}) e^{jd_1} &= \sum_{i=1}^{n} A_{2,i} e^{jd_{2,i}} \\ &= \sum_{i=1}^{n} A_{2,i} \cos d_{2,i} + j \sum_{i=1}^{n} A_{2,i} \sin d_{2,i}. \end{aligned} \tag{5}$$

For real values of d_1 and $d_{2,i}$'s, taking the square of the magnitudes,

$$\begin{aligned}(\sum_{i=1}^{n} A_{2,i})^2 &= (\sum_{i=1}^{n} A_{2,i} \cos d_{2,i})^2 + (\sum_{i=1}^{n} A_{2,i} \sin d_{2,i})^2 \\ &= \sum_{i=1}^{n} A_{2,i}^2 + 2 \sum_{i \neq k} A_{2,i} A_{2,k} \cos(d_{2,i} - d_{2,k}) \end{aligned} \tag{6}$$

i.e.

$$\sum_{i \neq k} A_{2,i} A_{2,k} = \sum_{i \neq k} A_{2,i} A_{2,k} \cos(d_{2,i} - d_{2,k}). \tag{7}$$

Hence, for $|d_{2,i} - d_{2,k}| < 2\pi$,[4] $i, k \leq n$ and $i \neq k$, $d_{2,i} = d_{2,k}$; in addition, from (5), $d_{2,i} = d_{2,k} = d_1$. In other words, the

[4] This inequality can always be satisfied by scaling appropriately the object dimensions.

CEGI's of these two objects are different since there is no real solution for d_1 when $d_{2,i} \neq d_{2,k}$ $(i \neq k)$. This means that the CEGI is able to differentiate between convex and nonconvex objects with identical EGI's. □

Property 3: If the object is convex, then the mass center of the magnitude distribution of its CEGI representation coincides with the Gaussian sphere center.

Proof: For a given conventional EGI representation of an object, the center of mass coincides with the sphere center. If $P_{\hat{\mathbf{n}}_i}^{EGI}$ is the weight at normal $\hat{\mathbf{n}}_i$, then the previous statement can be mathematically expressed as

$$\sum_{i=1}^{N_{faces}} P_{\hat{\mathbf{n}}_i}^{EGI} \hat{\mathbf{n}}_i = \mathbf{0} \tag{8}$$

where N_{faces} is the number of flat surfaces. Since for a convex object $|P_{\hat{\mathbf{n}}_i}^{CEGI}| = P_{\hat{\mathbf{n}}_i}^{EGI}$ (where $P_{\hat{\mathbf{n}}_i}^{CEGI}$ is the complex weight associated with normal $\hat{\mathbf{n}}_i$), the result follows. □

Corollary: If, for an object

$$\sum_{i=1}^{N_{faces}} |P_{\hat{\mathbf{n}}_i}^{CEGI}| \hat{\mathbf{n}}_i \neq \mathbf{0} \tag{9}$$

then that object is *not* convex.

Note that the converse of Property 3 is not necessarily true. If $\sum |P_{\hat{\mathbf{n}}_i}^{CEGI}| \hat{\mathbf{n}}_i = \mathbf{0}$, then it does not necessarily mean that the object is convex. This relation is a necessary but not sufficient condition for convexity.

Property 4: If, for a *nonconvex* object

$$\sum_{i=1}^{N_{faces}} |P_{\hat{\mathbf{n}}_i}^{CEGI}| \hat{\mathbf{n}}_i = \mathbf{0} \tag{10}$$

then it satisfies one of the following conditions:

1. The object is point symmetric (e.g., Fig. 7).
2. The nonconvex parts of the object possess face gradients not found elsewhere. In other words, each face must have a unique outward normal. Examples of such objects are shown in Fig. 8.
3. If the nonconvex parts of the object contain faces with the same gradients as those at other parts, then these faces must lie on the same plane, i.e., they have the same normal distance from the origin.

Although it is clear from the above analysis that different nonconvex objects having similar EGI's would have distinct CEGI's, the following open questions remain:

1. Can several nonconvex objects that are distinct under the EGI have the same CEGI?
2. Can several nonconvex objects in general have the same CEGI?

Although it is very possible that the answers to both these questions are in the affirmative, in such cases, it is probably much less likely for two objects to possess similar CEGI's than for them to have similar EGI's. Intuitively speaking, the level of discrimination between different objects under the CEGI is higher due to the addition of the support function in the form of the surface normal distance.

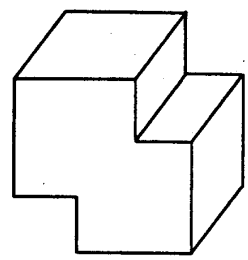

Fig. 7. Point symmetric object.

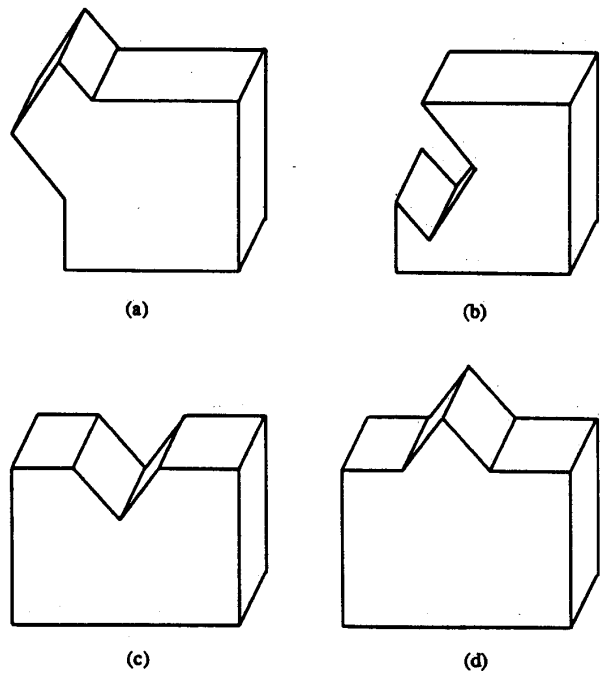

Fig. 8. Examples of nonconvex objects satisfying the relation $\sum |P_{\hat{\mathbf{n}}_i}^{\mathrm{CEGI}}| \hat{\mathbf{n}}_i = \mathbf{0}$.

D. Alternative Representations

There are several other ways of incorporating the distance information in the EGI representation. Two of the alternatives are described and compared with the CEGI representation.

1) Representation with Simple Ordered Pair: One possibility is to use a simple ordered pair of numbers[5] that are functions of the surface area and the surface normal distance from the origin. A problem that immediately arises is the choice of mixing functions for both these numbers that allows consistent merging of representations of discrete surfaces with the same normal. Let $A_{i,\mathbf{n}_k}$ and $d_{i,\mathbf{n}_k}$ be the area of the ith discrete surface area and surface normal distance, respectively. The second term in the subscripts indicate that its surface normal is $\mathbf{n}_k$. If the ordered pair is represented by $(A_{i,\mathbf{n}_k}, d_{i,\mathbf{n}_k})$ for a solitary surface patch, one possible set of mixing

[5] The complex representation can be viewed as an ordered pair of the magnitude and phase values. What we meant by a simple ordered pair here is that both the numbers in that pair are created independently of each other.

functions for a collection of discrete surface patches with normals $\mathbf{n}_k$ is

$$P_{\mathbf{n}_k} = \left(\sum_i A_{i,\mathbf{n}_k}, \frac{\sum_i A_{i,\mathbf{n}_k} d_{i,\mathbf{n}_k}}{\sum_i A_{i,\mathbf{n}_k}} \right). \tag{11}$$

This representation works because if $P_{\mathbf{n}_k}$ is the initial ordered pair and $P_{\mathbf{n}_k}'$ is the ordered pair after the object undergoes translation $\mathbf{T}$, then it is easy to see that $P_{\mathbf{n}_k}' - P_{\mathbf{n}_k} = (0, \mathbf{T} \cdot \mathbf{n}_k)$. $\mathbf{T}$ can then be determined from simple least square fitting.

The approach for determining object pose is to recognize the object and determine its orientation before finally recovering its translation. The first two steps in this approach necessarily involve *only* the first value in each pair of numbers expressed by (11). This is because the surface normal distances are initially unknown and only are recovered once these first two steps are successfully taken. As a result, such a representation possesses the same ambiguity problem as the EGI since only area-normal information is used (as in the case of the EGI) in recognizing the object and determining its orientation.

Even though the mixing function associated with the second number is directly affected by the surface normal distance, there is no invariance with respect to object translation that can be used in orientation recovery.

In contrast, in the CEGI weighting scheme, the magnitude of the complex weight is invariant with respect to object translation. Referring to (2), for a given object, while the CEGI weight itself is a function of surface distances from the origin *and* their distribution (e.g., compare Fig. 3 with Fig. 4), the magnitude of that weight is invariant with respect to the collective change in the distance distribution.[6] For a convex object $\mathcal{O}_C$ and a nonconvex object $\mathcal{O}_{NC}$ with identical EGI representations, the magnitudes of their CEGI weights are *not* identical for all the surface directions because there exist different distance distributions in at least one surface normal direction. (In Fig. 6, Object (1) is an instance of $\mathcal{O}_C$, and Object (2) is an instance of $\mathcal{O}_{NC}$.) Thus, the complex weight helps in differentiating $\mathcal{O}_C$ from $\mathcal{O}_{NC}$ in addition to enabling object pose determination (Property 2 in Section II-C).

The representation given by (11) does not have the desirable property of being able to simultaneously discriminate between $\mathcal{O}_C$ and $\mathcal{O}_{NC}$ *and* maintain invariance with respect to the collective change in the distance distribution of each object. However, it imposes no restrictions on the magnitude of object translation (as compared with the CEGI representation).

2) Representation with Maximum Distance Support Function: Another possibility is to use the maximum distance support function with the sum of the area given by

$$P_{\mathbf{n}_k} = \left(\sum_i A_{i,\mathbf{n}_k}, \max_i (d_{i,\mathbf{n}_k}) \right). \tag{12}$$

Although this representation is theoretically correct and probably is intuitively more appealing, it may not be robust in practice. The maximum distance support function would be prone to sensor errors (e.g., in a range image) since a glitch that is not large enough to be filtered out entirely could result in a significant error in the support function at a point in the augmented EGI. Including surface area-normal information of all the surface patches (as in the case of the CEGI representation) reduces the effects of errors such as these. The representation with the maximum distance support function also faces the ambiguity problem described previously.

As with the previous representation, this representation imposes no restrictions on the magnitude of object translation (as compared with the CEGI representation). On the other hand, the CEGI representation is able to differentiate between convex and nonconvex objects having equivalent EGI representations. As explained in Section II-B, the translation magnitude restriction is not a significant problem of the CEGI representation because the vicinity of the object is known from the image.

[6] By collective change, we mean that all distances in the distance distribution change by the same amount equal to the object translation in the direction of the surface normal.

III. Pose Determination Strategy Using CEGI

A. Methodology

Given a prototype CEGI and a partial CEGI of an unknown object, we can recognize the object and determine its orientation as follows: First, calculate the magnitude distributions of both CEGI's, and second, match the resulting distribution with that of the prototype. 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 that the object has been translated by δx, δy, and δz in the x, y and z directions, respectively (in the model world coordinates). Then, for each surface whose surface normal is $\hat{\mathbf{n}}_k$ and whose complex weight is originally $A_{\hat{\mathbf{n}}_k} e^{jd_k}$, after translation, the complex weight becomes $A_{\hat{\mathbf{n}}_k} e^{j(d_k + \delta\mathbf{d} \cdot \hat{\mathbf{n}}_k)}$, where

$$\begin{aligned} \delta\mathbf{d} &= \delta x \hat{\mathbf{i}} + \delta y \hat{\mathbf{j}} + \delta z \hat{\mathbf{k}} \\ \hat{\mathbf{n}}_k &= \mathrm{n}_{kx} \hat{\mathbf{i}} + \mathrm{n}_{ky} \hat{\mathbf{j}} + \mathrm{n}_{kz} \hat{\mathbf{k}}. \end{aligned} \tag{13}$$

Then, for each matched weight $P'_{\hat{\mathbf{n}}_i}$ in the object CEGI corresponding to the weight $P_{\hat{\mathbf{n}}_i}$ in the model CEGI, let

$$\begin{aligned} \omega_i &= \arg\left(\frac{P'_{\hat{\mathbf{n}}_i}}{P_{\hat{\mathbf{n}}_i}}\right) \\ &= \arg\left(\frac{A_{\hat{\mathbf{n}}_k} e^{j(d_k + \delta\mathbf{d} \cdot \hat{\mathbf{n}}_k)}}{A_{\hat{\mathbf{n}}_k} e^{jd_k}}\right) \\ &= \delta x \mathrm{n}_{ix} + \delta y \mathrm{n}_{iy} + \delta z \mathrm{n}_{iz} \\ &\quad \text{for } i = 1, \ldots, N_{visible} \end{aligned} \tag{14}$$

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

$$\mathcal{E} = \sum_{i=1}^{N_{visible}} (\omega_i - \mathrm{n}_{ix}\delta x - \mathrm{n}_{iy}\delta y - \mathrm{n}_{iz}\delta z)^2. \tag{15}$$

This is done by differentiating $\mathcal{E}$ with respect to the three unknown translation parameters and equating them to zero, i.e.

$$\frac{\partial \mathcal{E}}{\partial \delta x} = \frac{\partial \mathcal{E}}{\partial \delta y} = \frac{\partial \mathcal{E}}{\partial \delta z} = 0 \tag{16}$$

which yield the following system of equations:

$$\begin{aligned}
\delta x \sum \mathrm{n}_{ix}^2 + \delta y \sum \mathrm{n}_{ix}\mathrm{n}_{iy} + \delta z \sum \mathrm{n}_{ix}\mathrm{n}_{iz} &= \sum \omega_i \mathrm{n}_{ix} \\
\delta x \sum \mathrm{n}_{ix}\mathrm{n}_{iy} + \delta y \sum \mathrm{n}_{iy}^2 + \delta z \sum \mathrm{n}_{iy}\mathrm{n}_{iz} &= \sum \omega_i \mathrm{n}_{iy} \\
\delta x \sum \mathrm{n}_{ix}\mathrm{n}_{iz} + \delta y \sum \mathrm{n}_{iy}\mathrm{n}_{iz} + \delta z \sum \mathrm{n}_{iz}^2 &= \sum \omega_i \mathrm{n}_{iz}.
\end{aligned} \tag{17}$$

Using Cramer's rule to solve for δx, δy, and δz, we get

$$\delta x = \frac{1}{\mathbf{D}} \begin{vmatrix} \sum \omega_i \mathrm{n}_{ix} & \sum \mathrm{n}_{ix}\mathrm{n}_{iy} & \sum \mathrm{n}_{ix}\mathrm{n}_{iz} \\ \sum \omega_i \mathrm{n}_{iy} & \sum \mathrm{n}_{iy}^2 & \sum \mathrm{n}_{iy}\mathrm{n}_{iz} \\ \sum \omega_i \mathrm{n}_{iz} & \sum \mathrm{n}_{iy}\mathrm{n}_{iz} & \sum \mathrm{n}_{iz}^2 \end{vmatrix} \tag{18}$$

$$\delta y = \frac{1}{\mathbf{D}} \begin{vmatrix} \sum \mathrm{n}_{ix}^2 & \sum \omega_i \mathrm{n}_{ix} & \sum \mathrm{n}_{ix}\mathrm{n}_{iz} \\ \sum \mathrm{n}_{ix}\mathrm{n}_{iy} & \sum \omega_i \mathrm{n}_{iy} & \sum \mathrm{n}_{iy}\mathrm{n}_{iz} \\ \sum \mathrm{n}_{ix}\mathrm{n}_{iz} & \sum \omega_i \mathrm{n}_{iz} & \sum \mathrm{n}_{iz}^2 \end{vmatrix} \tag{19}$$

$$\delta z = \frac{1}{\mathbf{D}} \begin{vmatrix} \sum \mathrm{n}_{ix}^2 & \sum \mathrm{n}_{ix}\mathrm{n}_{iy} & \sum \omega_i \mathrm{n}_{ix} \\ \sum \mathrm{n}_{ix}\mathrm{n}_{iy} & \sum \mathrm{n}_{iy}^2 & \sum \omega_i \mathrm{n}_{iy} \\ \sum \mathrm{n}_{ix}\mathrm{n}_{iz} & \sum \mathrm{n}_{iy}\mathrm{n}_{iz} & \sum \omega_i \mathrm{n}_{iz} \end{vmatrix} \tag{20}$$

where

$$\mathbf{D} = \begin{vmatrix} \sum \mathrm{n}_{ix}^2 & \sum \mathrm{n}_{ix}\mathrm{n}_{iy} & \sum \mathrm{n}_{ix}\mathrm{n}_{iz} \\ \sum \mathrm{n}_{ix}\mathrm{n}_{iy} & \sum \mathrm{n}_{iy}^2 & \sum \mathrm{n}_{iy}\mathrm{n}_{iz} \\ \sum \mathrm{n}_{ix}\mathrm{n}_{iz} & \sum \mathrm{n}_{iy}\mathrm{n}_{iz} & \sum \mathrm{n}_{iz}^2 \end{vmatrix}. \tag{21}$$

Note that this yields either a global minimum or a saddle point (depending on the sign of the right-hand expression in (23)) since[7]

$$\frac{\partial^2 \mathcal{E}}{\partial \delta p^2} = 2 \sum n_{ip}^2 > 0 \tag{22}$$

and

$$\frac{\partial^2 \mathcal{E}}{\partial \delta p \delta q} = 2 \sum n_{ip} n_{iq} \tag{23}$$

(p and q can be x, y, or z, but they are different axis components.)

In order for this scheme to work, the magnitude of the translation (i.e., the normalized translation) must be less than π as given in (24).

$$\Delta = \sqrt{(\delta x)^2 + (\delta y)^2 + (\delta z)^2} < \pi. \tag{24}$$

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

[7] We are assuming a nontrivial normal distribution (corresponding to the visible faces of the object) whose normals span $\Re^3$.

B. Analysis of Error in Translation Parameters

The upper bound error depends on a variety of parameters, with the main factors being the normal and normal distance errors. Curiously enough, for a *convex* object, the magnitude of the weight does not contribute to the distance error. This is because only the *phase* components of the complex weight are used in the least square formulation, and for a convex object, no complex weight associated with a particular normal is a combination of at least two weights from two different disjoint object faces.

Each normal is represented by its directional cosines

$$\hat{\mathbf{n}}_l = (\cos\theta_{x,l} \;\; \cos\theta_{y,l} \;\; \cos\theta_{z,l})^T \tag{25}$$

and is subject to the constraint

$$\cos^2\theta_{x,l} + \cos^2\theta_{y,l} + \cos^2\theta_{z,l} = 1. \tag{26}$$

Suppose the exact translation is $\delta\mathbf{d} = (\delta x \;\; \delta y \;\; \delta z)^T$ and the approximate translation obtained using the least square formulation is $\delta\mathbf{d}' = (\delta x' \;\; \delta y' \;\; \delta z')^T$. In addition, suppose that the perturbed lth normal is indicated by $\hat{\mathbf{n}}'_l$. Let

$$\omega_l = \arg\left(\frac{P_{\hat{\mathbf{n}}_{l,obj}}}{P_{\hat{\mathbf{n}}_{l,model}}}\right) \tag{27}$$

i.e.,

$$\begin{aligned}
\omega_l &= \delta\mathbf{d} \cdot \hat{\mathbf{n}}_l \\
&= \delta x \cos\theta_{x,l} + \delta y \cos\theta_{y,l} + \delta z \cos\theta_{z,l}.
\end{aligned} \tag{28}$$

For an imperfectly extracted object

$$\begin{aligned}
\omega_l + \epsilon_l &= \delta\mathbf{d}' \cdot \hat{\mathbf{n}}'_l \\
&= \delta x' \cos\theta'_{x,l} + \delta y' \cos\theta'_{y,l} + \delta z' \cos\theta'_{z,l} \\
&= (\delta x + \epsilon_x)\cos(\theta_{x,l} + \eta_{x,l}) + (\delta y + \epsilon_y)\cos(\theta_{y,l} + \eta_{y,l}) \\
&\quad + (\delta z + \epsilon_z)\cos(\theta_{z,l} + \eta_{z,l}).
\end{aligned} \tag{29}$$

Subtracting (28) from (29) and ignoring the second-order terms, we have

$$\begin{aligned}
\epsilon_l = \epsilon_x \cos\theta_{x,l} - \eta_{x,l}\delta x \sin\theta_{x,l} + \epsilon_y \cos\theta_{y,l} - \eta_{y,l}\delta y \sin\theta_{y,l} \\
+ \epsilon_z \cos\theta_{z,l} - \eta_{z,l}\delta z \sin\theta_{z,l}.
\end{aligned} \tag{30}$$

Hence

$$\begin{aligned}
&\epsilon_x \cos\theta_{x,l} + \epsilon_y \cos\theta_{y,l} + \epsilon_z \cos\theta_{z,l} \\
&\quad = \epsilon_l + \eta_{x,l}\delta x \sin\theta_{x,l} + \eta_{y,l}\delta y \sin\theta_{y,l} + \eta_{z,l}\delta z \sin\theta_{z,l}.
\end{aligned} \tag{31}$$

Suppose that the maximum positive error is $|\eta_{x,l}|_{max} = |\eta_{y,l}|_{max} = |\eta_{z,l}|_{max} = \eta_{max}$. For the normalized distances, $|\delta x|_{max} = |\delta y|_{max} = |\delta z|_{max} = \pi$. The following simplifying assumptions are made:

1. The normals can be "mapped" into the first octant; x, y, and z values are positive such that the smaller angle subtended with the x, y, and z axes are still the same. This means that $0 \le \theta_{x,l}, \theta_{y,l}, \theta_{z,l} \le \frac{\pi}{2}$.
2. $\epsilon_x, \epsilon_y, \epsilon_z$, and ϵ_l are all positive.

Then

$$|\epsilon_x|\cos\theta_{x,l} + |\epsilon_y|\cos\theta_{y,l} + |\epsilon_z|\cos\theta_{z,l} \leq |\epsilon_l| + \eta_{max}\pi(\sin\theta_{x,l} + \sin\theta_{y,l} + \sin\theta_{z,l}). \quad (32)$$

Summing this over all the visible normals and averaging, we obtain

$$|\epsilon_x|\kappa_{\theta x} + |\epsilon_y|\kappa_{\theta y} + |\epsilon_z|\kappa_{\theta z} \leq \xi + \eta_{max}\pi(\tau_{\theta x} + \tau_{\theta y} + \tau_{\theta z}) \quad (33)$$

where

$$\kappa_{\theta x} = \frac{1}{N_{visible}} \sum_{l=1}^{N_{visible}} \cos\theta_{x,l}$$

$$\kappa_{\theta y} = \frac{1}{N_{visible}} \sum_{l=1}^{N_{visible}} \cos\theta_{y,l}$$

$$\kappa_{\theta z} = \frac{1}{N_{visible}} \sum_{l=1}^{N_{visible}} \cos\theta_{z,l}$$

$$(0 \leq \theta_{x,l}, \theta_{y,l}, \theta_{z,l} \leq \frac{\pi}{2})$$

$$\xi = \frac{1}{N_{visible}} \sum_{l=1}^{N_{visible}} |\omega_l - \hat{\mathbf{n}}_l' \cdot (\delta x' \delta y' \delta z')^T| \quad (34)$$

$$\tau_{x,l} = \frac{1}{N_{visible}} \sum_{l=1}^{N_{visible}} \sin\theta_{x,l}$$

$$\tau_{y,l} = \frac{1}{N_{visible}} \sum_{l=1}^{N_{visible}} \sin\theta_{y,l}$$

$$\tau_{z,l} = \frac{1}{N_{visible}} \sum_{l=1}^{N_{visible}} \sin\theta_{z,l}.$$

The value of ξ is normally small when compared with the other terms (at least an order of magnitude smaller) and thus can be ignored. From (33), we can deduce the following set of constraints on the maximum error in each direction:

$$|\epsilon_x| \leq \frac{\eta_{max}\pi(\tau_{\theta x} + \tau_{\theta y} + \tau_{\theta z})}{\kappa_{\theta x}}$$
$$|\epsilon_y| \leq \frac{\eta_{max}\pi(\tau_{\theta x} + \tau_{\theta y} + \tau_{\theta z})}{\kappa_{\theta y}} \quad (35)$$
$$|\epsilon_z| \leq \frac{\eta_{max}\pi(\tau_{\theta x} + \tau_{\theta y} + \tau_{\theta z})}{\kappa_{\theta z}}.$$

From the set of inequalities in (35), the constraint on the error in the total effective distance is

$$\epsilon_{total} = \sqrt{\epsilon_x^2 + \epsilon_y^2 + \epsilon_z^2} \leq \eta_{max}\pi(\tau_{\theta x} + \tau_{\theta y} + \tau_{\theta z})\sqrt{\frac{1}{\kappa_{\theta x}^2} + \frac{1}{\kappa_{\theta y}^2} + \frac{1}{\kappa_{\theta z}^2}}. \quad (36)$$

Define the view suitability index

$$\Omega_\theta = (\tau_{\theta x} + \tau_{\theta y} + \tau_{\theta z})\sqrt{\frac{1}{\kappa_{\theta x}^2} + \frac{1}{\kappa_{\theta y}^2} + \frac{1}{\kappa_{\theta z}^2}} \quad (37)$$

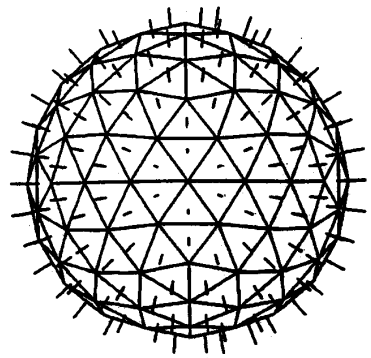

Fig. 9. Tesselated pentakis dodecahedron.

so that

$$\epsilon_{total} \leq \eta_{max}\pi\Omega_\theta. \quad (38)$$

This index shows us that minimizing the distance error *bound* requires us to find a viewpoint having it as its minimum. Equation (38) indicates that the error bound increases linearly as a function of the view suitability index Ω_θ.

IV. Simulations

A. Experiments Involving Polyhedral Objects

Two models were used to test the concept of translation parameter extraction using the CEGI representation; they are shown in Figs. 10 (Model 1) and 11 (Model 2). The maximum displacement was chosen to be 100 mm. The circumscribing cuboids of both these models have dimensions close to this figure. Fifty runs were made for each model; at each run, a random viewpoint and object translation were chosen. The viewpoint is assumed to be known initially, but it is subject to discretization error.

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 two-frequency dodecahedron (tesselated pentakis dodecahedron) as shown in Fig. 9. The normal direction space is discretized into 240 cells as well.

It is important to note that the CEGI weights are recomputed for each discrete view direction. This is done to compensate for the varying degrees of self occlusion that causes nonconvex objects to register different weights at different view directions for the same surface normal.

2) Results: The results of the simulation are depicted in Figs. 12 and 13. (The errors have been normalized such that the value π corresponds to the maximum allowable displacement of 100 mm.) As can be seen, Model 1 evinces a higher variability in errors than those demonstrated with Model 2. The extremely high peaks in Fig. 12 correspond to errors detected when the actual viewpoint subtends a very small angle to the x-y plane. Since the planes that yield the z-direction information (i.e., perpendicular to the z axis) for Model 1 (shown in Fig. 10) subtend almost 90° to these viewpoints,

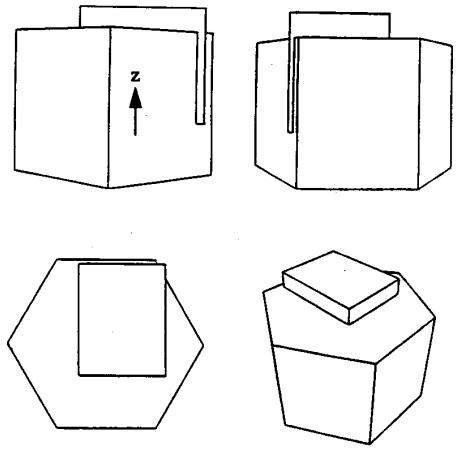

Fig. 10. First composite object used for testing (Model 1) shown from four different viewing directions.

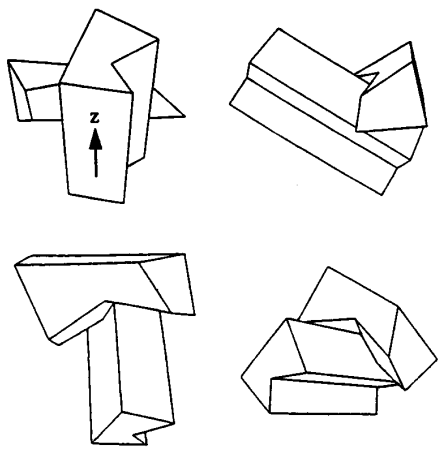

Fig. 11. Second composite object used for testing (Model 2) shown from four different viewing directions.

errors predicted in the z direction are not reliable and are prone to high errors. In addition, Model 1 has fewer faces than Model 2 (shown in Fig. 11). Furthermore, the number of differently oriented normals in Model 1 is even smaller than those in Model 2.

Table I 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 absolute error in K, whereas σ_{ϵ_K} is the standard deviation of the absolute error distribution. $\overline{\epsilon_{d_{xyz}}}$ is the mean absolute total error in position.)

It can be readily seen from Table I 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. On the other hand, the number of different surface normals and their more even distribution result in lower errors in the recovered parameters, as evidenced in the results for Model 2.

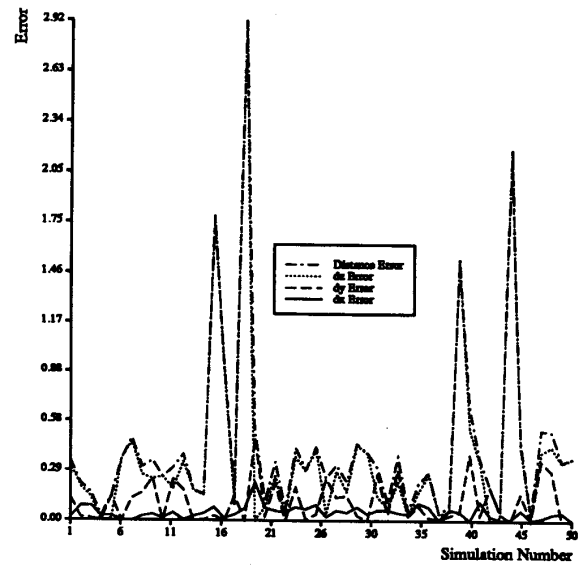

Fig. 12. Errors in dx, dy, dz, and total distance for Model 1.

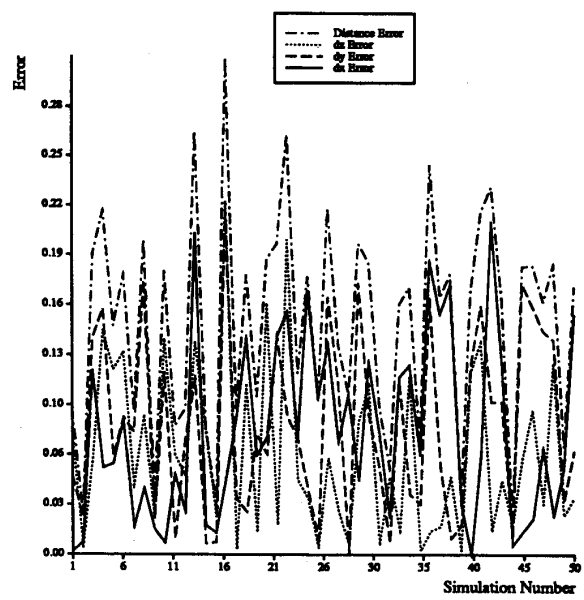

Fig. 13. Errors in dx, dy, dz, and total distance for Model 2.

TABLE I
RESULTS FOR SIMULATIONS INVOLVING POLYHEDRAL OBJECTS

Model #	$\overline{\epsilon_{d_x}}$	$\sigma_{\epsilon_{d_x}}$	$\overline{\epsilon_{d_y}}$	$\sigma_{\epsilon_{d_y}}$	$\overline{\epsilon_{d_z}}$	$\sigma_{\epsilon_{d_z}}$	$\overline{\epsilon_{d_{xyz}}}$	$\sigma_{\epsilon_{d_{xyz}}}$
1	1.6	1.2	3.4	3.8	12.2	15.9	14.1	16.3
2	2.5	1.8	2.6	1.8	2.1	1.7	4.7	2.3

B. Experiments Involving Smooth Objects

In addition to the two polyhedral objects, two smooth and continuous objects were also used in simulations and

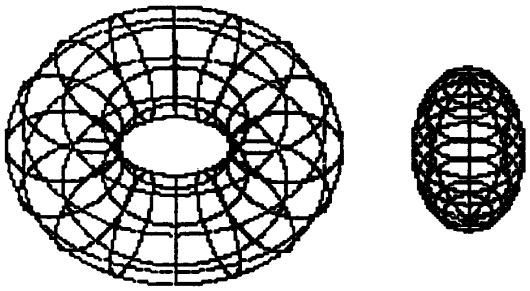

Fig. 14. Models used in experiments: torus (left) and ellipsoid (right).

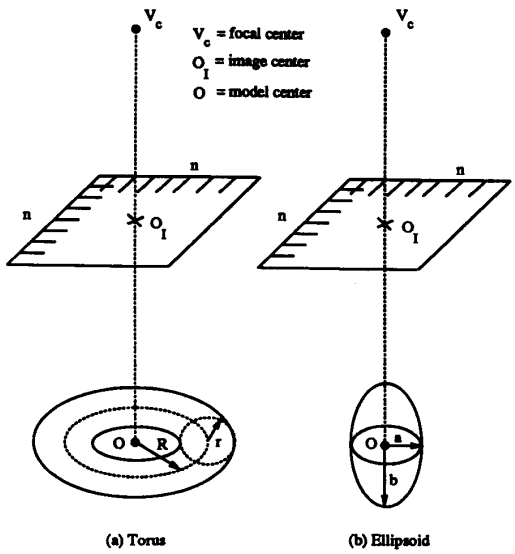

Fig. 15. Simulation using the torus (left) and ellipsoid (right). n denotes the resolution in the x and y directions of the simulated image plane.

experiments involving real range data. The two chosen smooth objects are the torus and the ellipsoid (Fig. 14).

1) Simulation: Implementational Issues: The torus is modeled parametrically as $[(R+r\cos\theta)\cos\phi, (R+r\cos\theta)\sin\phi, r\sin\theta]^{\mathrm{T}}$, whereas the ellipsoid is represented parametrically as $[a\sin\theta\cos\phi, a\sin\theta\sin\phi, b\cos\theta]^{\mathrm{T}}$. The ideal setup is shown in Fig. 15. The parametric values used are $r = 20, R = 40, a = 20, b = 40, |O_I O| = 100$, and $|V_c O| = 1000$. 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. This is illustrated in Fig. 16. The surface normal is approximated to be that at point P_{ck}, whereas the surface area is estimated by the sum of the area of the triangles $\Delta P_{ck}P_{1k}P_{2k}$, $\Delta P_{ck}P_{2k}P_{3k}$, $\Delta P_{ck}P_{3k}P_{4k}$, and $\Delta P_{ck}P_{4k}P_{1k}$. If the surface normal is $\mathbf{n}_k$ and the area is A_k, then the complex weight attributed to this surface is simply $A_k e^{j\mathbf{n}_k \cdot P_{ck}}$. This is done for all the pixels in the image plane.

2) Simulation: Results: The simulation results are graphically depicted in Fig. 17 and listed in Tables II and III.

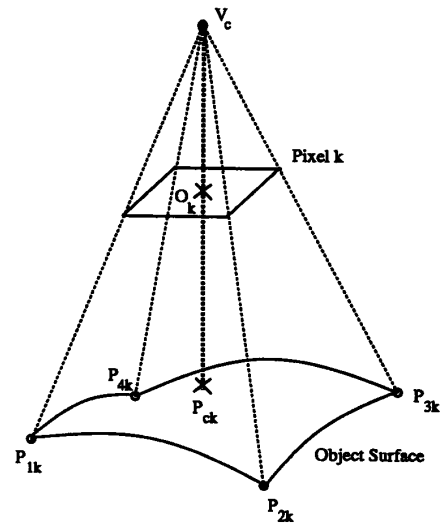

Fig. 16. Approximation of area and surface normal in simulation.

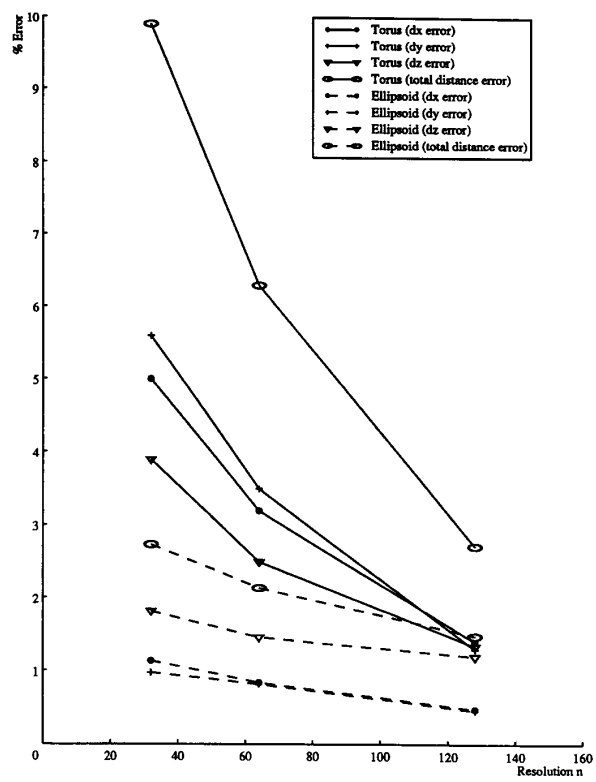

Fig. 17. Simulation results for torus and ellipsoid.

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

TABLE II
SIMULATION RESULTS FOR TORUS

Resolution	$\overline{\epsilon_{d_x}}$	$\sigma_{\epsilon_{d_x}}$	$\overline{\epsilon_{d_y}}$	$\sigma_{\epsilon_{d_y}}$	$\overline{\epsilon_{d_z}}$	$\sigma_{\epsilon_{d_z}}$	$\overline{\epsilon_{d_{xyz}}}$	$\sigma_{\epsilon_{d_{xyz}}}$
32×32	5.0	5.5	5.6	6.1	3.9	3.3	9.9	7.3
64×64	3.2	3.2	3.5	3.7	2.5	2.0	6.3	4.1
128×128	1.4	1.5	1.3	1.2	1.3	1.1	2.7	1.7

TABLE III
SIMULATION RESULTS FOR ELLIPSOID

Resolution	$\overline{\epsilon_{d_x}}$	$\sigma_{\epsilon_{d_x}}$	$\overline{\epsilon_{d_y}}$	$\sigma_{\epsilon_{d_y}}$	$\overline{\epsilon_{d_z}}$	$\sigma_{\epsilon_{d_z}}$	$\overline{\epsilon_{d_{xyz}}}$	$\sigma_{\epsilon_{d_{xyz}}}$
32×32	1.1	1.1	1.0	0.8	1.8	1.5	2.7	1.5
64×64	0.8	0.8	0.8	0.7	1.5	1.1	2.1	1.1
128×128	0.5	0.4	0.5	0.3	1.2	1.0	1.5	0.9

those incurred for the ellipsoid. This phenomena can be readily explained as follows:

Since the ellipsoid is convex, the resultant complex weight in each CEGI cell is the sum of those weights corresponding to surface patches that are either contiguous or spatially close to each other. (The number of contributing surface patches to a cell is inversely proportional to the local Gaussian curvature.) As such, the phases of these contributing complex weights are not expected to be very different from each other, despite errors in the phases. Since the resultant phase must lie between the minimum and maximum of these phases, the variance of the resultant phase error is expected to be small.

On the other hand, for the nonconvex 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.

C. Verification of the View Suitability Index

Section III-B establishes the upper bound relationship between errors in the distances δ_x, δ_y, δ_z, the total effective distance, and the view suitability index. Simulations were made to verify this linear relationship. To simplify the analysis, we considered the case where the normals are not discretized according to the pentakis dodecahedron. Instead, normals are represented exactly. Random positional errors, which are subject to certain specified angular errors between the actual and randomized normals, are created to simulate the errors that could occur in a real situation. Since Figs. 19–21 and 23 and 24 show that the distance errors do indeed increase in a linear fashion with the angular error, we are able to use the view suitability index as a means of comparison. Furthermore, as Figs. 18 and 22 show, the view suitability index Ω_θ does appear to be a reasonable means of determining the upper bound of the distance errors in the effective total distances. Similar graphs [15] are obtained for the distance errors in the x, y, and z distances for both models.

V. EXPERIMENTS USING REAL RANGE DATA

To further confirm the validity of the pose recovery technique, experiments involving real range data of two smooth objects were conducted. The two smooth objects are the torus and ellipsoid, whose databases have been created in the simulations described earlier.

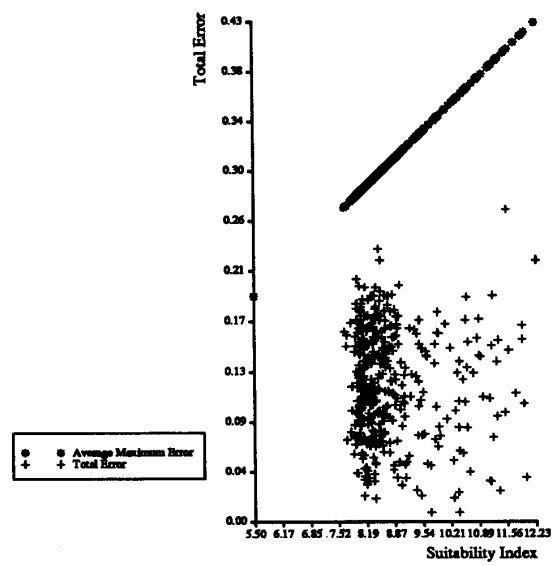

Fig. 18. Total error versus view suitability index for Model 1 (angular error = 3^0).

A. Implementational Issues

The torus and ellipsoid were crafted out of clay to resemble the models whose databases were created. A light stripe range finder [23] was used to produce a range image of these two objects in various poses and at two different resolutions. The experimental scheme is shown in Fig. 25. The resolution of Set 2 range images is twice that for Set 1 images.

B. Results

The surface maps of the range images taken are shown in Figs. 26 (torus-Set 1), 27 (ellipsoid-Set 1), 28 (torus-Set 2), and 29 (ellipsoid-Set 2). The experimental results for these images are summarized in Tables IV and V. Note that all the figures are quoted in percentages of the maximum allowable displacement (which is 100 mm). In this case, all the numbers

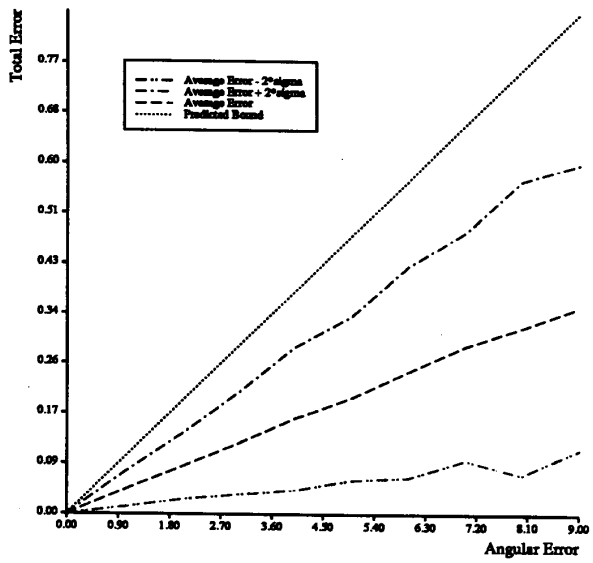

Fig. 19. Total error versus angular error for Model 1 ($7.50 \leq \Omega_\theta < 8.50$).

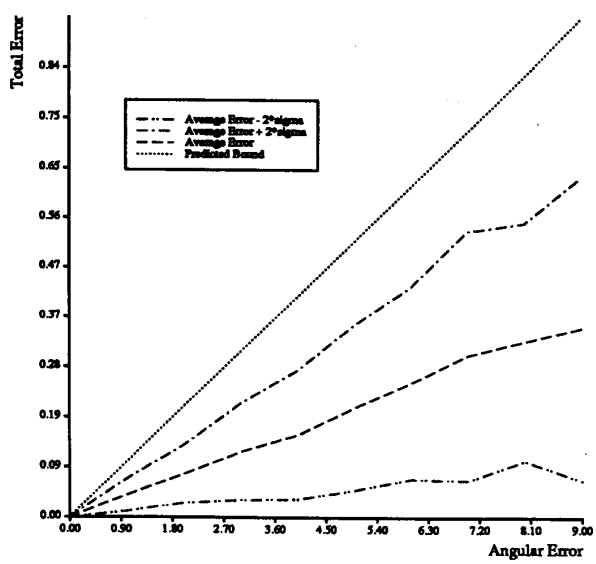

Fig. 20. Total error versus angular error for Model 1 ($8.50 \leq \Omega_\theta < 9.50$).

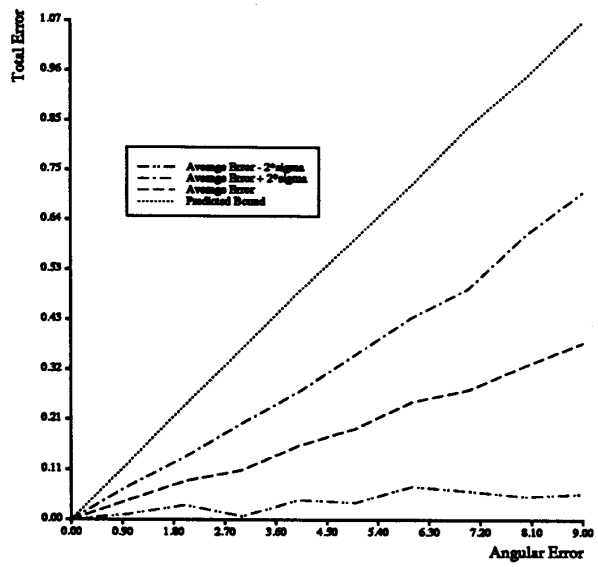

Fig. 21. Total error versus angular error for Model 1 ($9.50 \leq \Omega_\theta < 11.50$).

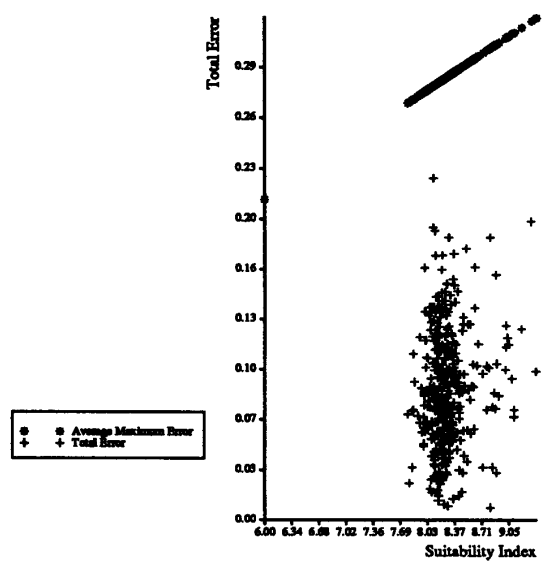

Fig. 22. Total error versus view suitability index for Model 2 (angular error = 3^0).

are in millimeters. (Note: ϵ_K denotes the absolute error in K; $\epsilon_{d_{xyz}}$ is the absolute total error in position.)

Again, as for the simulation results, the displacement errors are smaller for Set 2 experiments (which feature higher resolution range images) for both objects. Again, the error 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. 30 and 31 compare the actual positions with the calculated positions of the torus and ellipsoid in the range images.

VI. Conclusions

A new 3-D object representation that encodes object face position has been described. Called 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, whereas the magnitude of the weight is the visible area of the face.

The CEGI representation enables estimation of both the orientation and translation of a detected object with respect to a stored model or prototype. It effectively decouples the orientation and translation determination into two separate

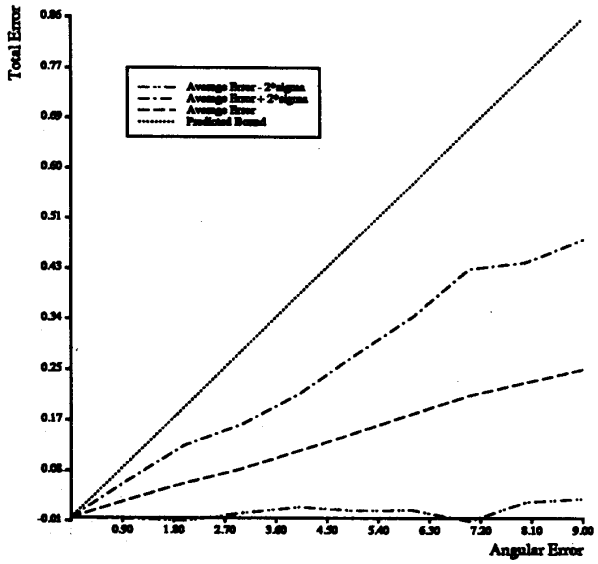

Fig. 23. Total error versus angular error for Model 2 ($7.50 \leq \Omega_\theta < 8.50$).

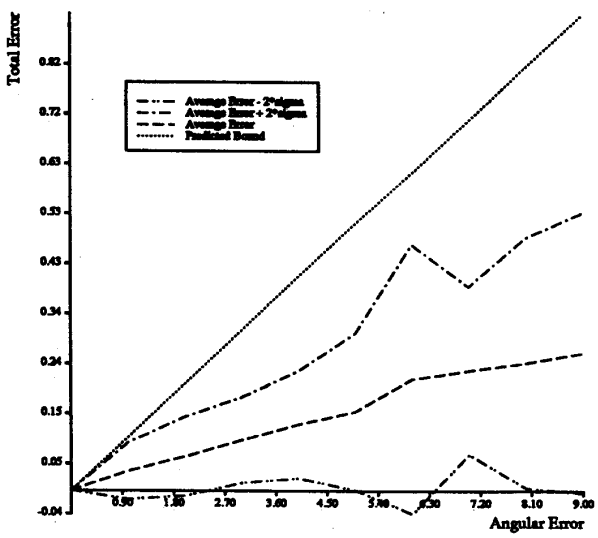

Fig. 24. Total error versus angular error for Model 2 ($8.50 \leq \Omega_\theta < 9.50$).

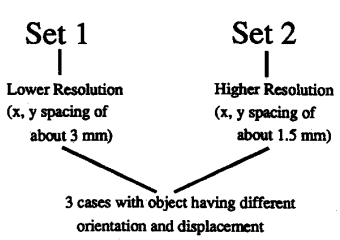

Fig. 25. Experimental scheme.

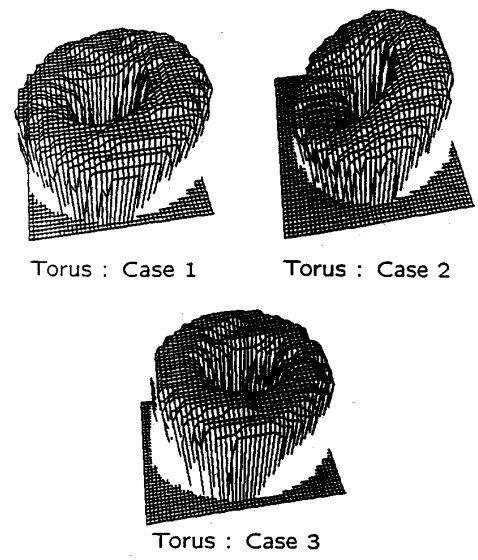

Fig. 26. Surface maps for the torus at three poses (Set 1).

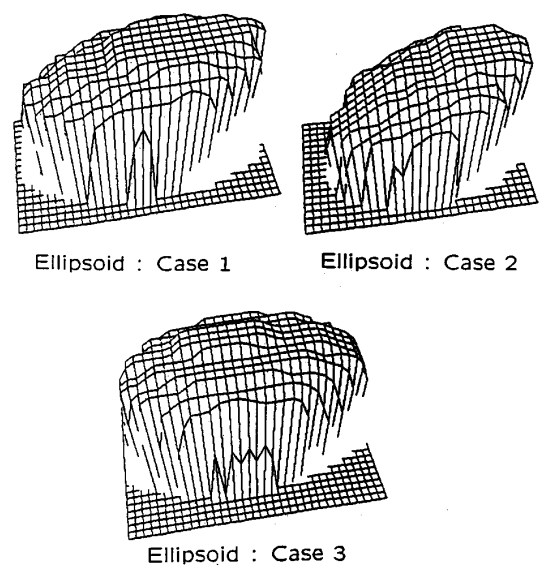

Fig. 27. Surface maps for the ellipsoid at three poses (Set 1).

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. The method of determining the translation parameters is simple because it is based on the least squares formulation. An important subsequence of the complex encoding is the ability of the CEGI to disambiguate objects having similar EGI's. Indeed, this property is one of the primary reasons for choosing this representation.

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*. On the other hand, the polyhedral object whose translation parameters need to be determined should be relatively complex in terms of the

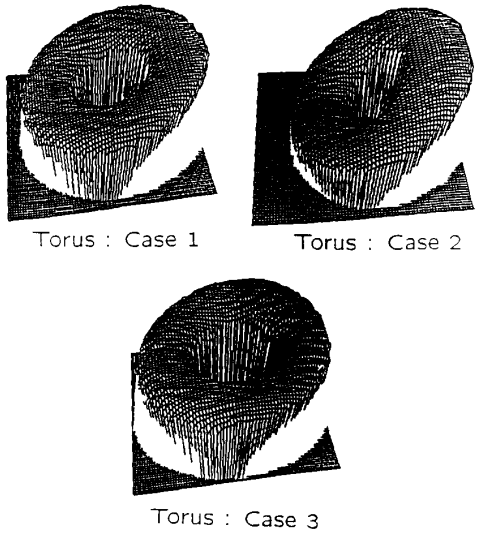

Fig. 28. Surface maps for the torus at three poses (Set 2).

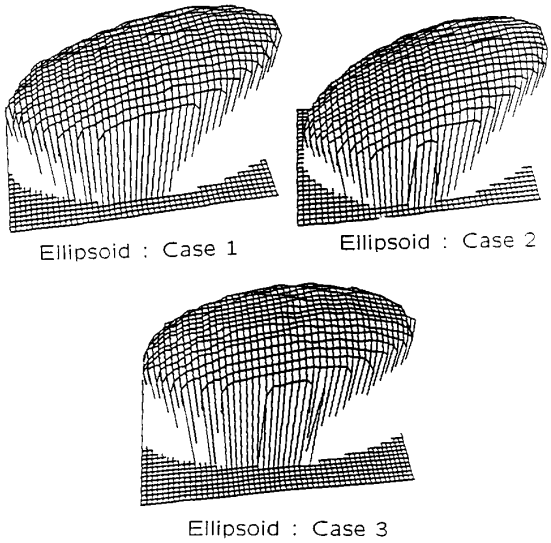

Fig. 29. Surface maps for the ellipsoid at three poses (Set 2).

TABLE IV
EXPERIMENTAL RESULTS FOR TORUS

Case #	Lower Resolution				Higher Resolution			
	ϵ_{d_x}	ϵ_{d_y}	ϵ_{d_z}	$\epsilon_{d_{xyz}}$	ϵ_{d_x}	ϵ_{d_y}	ϵ_{d_z}	$\epsilon_{d_{xyz}}$
1	1.7	5.9	5.0	8.5	1.4	2.9	4.0	5.4
2	2.6	4.1	7.3	9.5	4.2	2.5	5.6	7.7
3	3.1	4.2	3.6	6.6	3.2	1.4	2.9	5.5

number of faces and distribution of surface normals. The more the number of faces and discrete normal directions, the more accurate the resulting estimated displacement is likely to be. Similarly, this method is likelier to yield better results for a smooth object with a more even Gaussian distribution and a lower degree of concavity.

TABLE V
EXPERIMENTAL RESULTS FOR ELLIPSOID

Case #	Lower Resolution				Higher Resolution			
	ϵ_{d_x}	ϵ_{d_y}	ϵ_{d_z}	$\epsilon_{d_{xyz}}$	ϵ_{d_x}	ϵ_{d_y}	ϵ_{d_z}	$\epsilon_{d_{xyz}}$
1	0.9	1.4	3.2	3.7	0.9	3.1	1.6	3.6
2	0.4	2.1	4.9	5.4	0.9	4.1	0.2	4.2
3	1.3	2.7	2.1	3.9	1.6	0.5	1.9	2.6

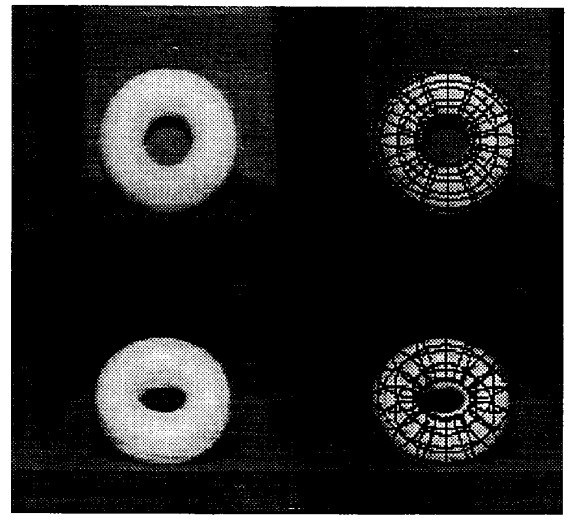

Fig. 30. Top, left: Intensity image of torus (Set 2, Case 1); top, right: superimposed model (Set 2, Case 1); bottom, left: intensity image of torus (Set 2, Case 2); bottom, right: superimposed model (Set 2, Case 2).

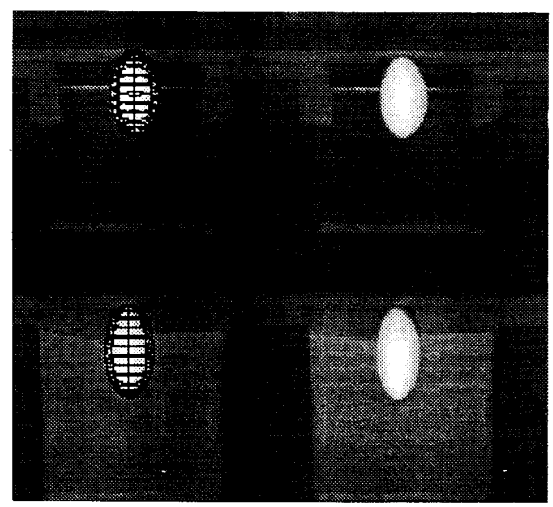

Fig. 31. Top, left: Intensity image of ellipsoid (Set 2, Case 1); top, right: superimposed model (Set 2, Case 1); bottom, left: intensity image of ellipsoid (Set 2, Case 2); bottom, right: superimposed model (Set 2, Case 2).

Based on error analysis, the accuracy of the derived translation parameters is sensitive to the angular error of the surface normals, the magnitude of the actual translation parameters, and the distribution of the surface normals. An index that can be used as an indication of the *bound* on the translation

parameters has been derived and is called the view suitability index. It has been shown through simulations that the mean of the translation errors is linear with normal angular error.

ACKNOWLEDGMENT

Many thanks to M. Hebert, M. Wheeler, and K. Porsche for proofreading draft copies of this paper. The authors would also like to express their appreciation to the referees. Their constructive comments have helped to significantly improve this paper.

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. government.

REFERENCES

[1] P. Balakumar, J. C. Robert, R. Hoffman, K. Ikeuchi, and T. Kanade, *VANTAGE: A Frame-Based Geometric Modeling System-Programmer/User's Manual*. Pittsburgh, PA: CMU, Dec. 1988.

[2] P. J. Besl and R. C. Jain, "Three-dimensional object recognition," *Artificial Intell.*, vol. 17, no. 1, pp. 75–145, Mar. 1985.

[3] B. Bhanu, "Representation and shape matching of 3-D objects," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-6, no. 3, pp. 340–350, 1984.

[4] R. M. Bolle and B. D. Copper, "On optimally combining pieces of information, with application to estimating 3-D complex-object position from range data," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-8, no. 5, pp. 619–638, 1986.

[5] R. C. Bolles, P. Horaud, and M. J. Hannah, "3DPO: A three-dimensional part orientation system," in *Proc. First Int. Symp. Robotic Res.*, 1984, pp. 413–424.

[6] P. Brou, "Using the Gaussian image to find the orientation of objects," *Int. J. Robotics Res.*, vol. 3, no. 4, pp. 89–125, Winter 1984.

[7] T. J. Fan, G. Medioni, and R. Nevatia, "Matching 3-D objects using surface descriptions," in *Proc. Int. Conf. Robotics Automat.*, Apr. 1988, pp. 1400–1406.

[8] O. D. Faugeras, "New steps toward a flexible 3-D vision system for robotics," in *Proc. Int. Conf. Patt. Recogn.*., July 1984, pp. 796–805.

[9] O. D. Faugeras and M. Hebert, "A 3-D recognition and positioning algorithm using geometrical matching between primitive surfaces," in *Proc. Int. Joint Conf. Artificial Intell.*., Aug. 1983, pp. 996–1002.

[10] W. E. L. Grimson and T. Lozano-Perez, "Recognition and localization of overlapping parts from sparse data," in *Three-Dimensional Machine Vision* (T. Kanade, Ed.). Boston: Kluwer, 1987, pp. 451–510.

[11] B. K. P. Horn, "Extended Gaussian images," *Proc. IEEE*, vol. 72, no. 12, pp. 1671–1686, Dec. 1984.

[12] B. K. P. Horn and K. Ikeuchi, "The mechanical manipulation of randomly oriented parts," *Sci. Amer.*, vol. 251, no. 2, pp. 100–113, Aug. 1984.

[13] K. Ikeuchi, "Recognition of 3-D objects using the extended Gaussian image," in *Proc. Seventh IJCAI*, 1981, pp. 595–600.

[14] ——, "Determining attitude of object from needle map using extended Gaussian image," Tech. Rep. AI Memo 714, MIT, 1983.

[15] S. B. Kang and K. Ikeuchi, "3-D object pose determination using complex EGI," Tech. Rep. CMU-RI-TR-90-18, Carnegie Mellon Univ., Oct. 1990.

[16] J. J. Little, "Recovering shape and determining attitude from extended Gaussian image," Tech. Rep. TR 85-2, Univ. of British Columbia, Apr. 1985.

[17] D. G. Lowe, *Perceptual Organization and Visual Recognition*. Boston: Kluwer, 1985.

[18] H. Minkowski, "Allgemeine Lehrsatze uber die konvexe Polyeder," *Nachr. GHes. Wiss. Gottingen*, 1897.

[19] V. S. Nalwa, "Representing oriented piecewise C^2 surfaces," in *Proc. 2nd Int. Conf. Comput. Vision*, Dec. 1988, pp. 40–51.

[20] R. Nevatia and T. O. Binford, "Description and recognition of curved objects," *Artificial Intell.*, no. 8, pp. 77–98, 1977.

[21] J. Ponce and D. J. Kriegman, "On recognizing and positioning curved 3-D objects from image contours," in *Proc. Image Understanding Workshop*, May 1989, pp. 461–470.

[22] J. W. Roach, J. S. Wright, and V. Ramesh, "Spherical dual images: A 3-D representation method for solid objects that combines dual space and Gaussian spheres," in *Proc. CVPR*, June 1986, pp. 236–241.

[23] K. Sato, H. Yamamoto, and S. Inokuchi, "Range imaging system utilizing nematic liquid crystal mask," in *Proc. Int. Conf. Comput. Vision*, 1987, pp. 657–661.

[24] T. M. Silberberg, D. Harwood, and L. S. Davis, "Three dimensional object recognition using oriented model points," in *Techniques for 3-D Machine Perception* (A. Rosenfeld, Ed.). New York: Elsevier, 1986, pp. 271–320.

[25] F. Stein and G. Medioni, "Structural indexing: Efficient 3-D recognition," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 4, no. 2, pp. 125–145, 1992.

Sing Bing Kang was born in Kuala Lumpur, Malaysia. He received the B.Eng and M.Eng. degrees, both in electrical engineering, from the National University of Singapore in 1987 and 1990, respectively. He is currently working toward the robotics Ph.D. degree at Carnegie Mellon University. His doctoral work is in recognizing grasping tasks from observation.

Recently, his paper on the CEGI (coauthored with K. Ikeuchi) won the IEEE Computer Society Outstanding Paper Award at the 1991 IEEE International Conference on Computer Vision and Pattern Recognition.

Katsushi Ikeuchi (M'89) received the B.Eng. degree in mechanical engineering from Kyoto University, Kyoto, Japan, in 1973 and the Ph.D. degrees in information engineering from the University of Tokyo, Tokyo, Japan, in 1978.

After working at the AI Lab at the Massachusetts Institute of Technology and ETL of MITI in Japan, he joined the School of Computer Science, Carnegie Mellon University (CMU), Pittsburgh, PA, in 1986. Currently, he is a Principal Research Scientist with the Computer Science Division and the Robotics Institute at CMU. He developed the "smoothness constraint," which forces neighboring points to have similar surface orientations, by which he iteratively recovered shape from shading and shape from texture. He pioneered the use of specular reflections to recover surface orientations. Instead of discarding specular reflections, he actively used them for recovering shape and reflectance. He developed the extended Gaussian image (EGI) to represent 3-D objects and later built a bin-picking system based on EGI. Currently, he has been developing vision algorithm compilers, which automatically convert object and sensor models into vision programs.

Dr. Ikeuchi has received several awards, including the D. Marr Best Paper Award in 1990 and the IEEE Outstanding Paper Award in 1991. In 1992, his paper was selected as one of the most influential papers that appeared in the journal *Artificial Intelligence*.