Object Recognition and Localization Using Optical Proximity Sensor System: Polyhedral Case

Sukhan Lee and Hern S. Hahn

Department of Electrical Engineering - Systems
Institute for Robotics and Intelligent Systems
University of Southern California
Los Angeles, California 90089-0273

Abstract

This paper presents an algorithm for the recognition and localization of 3D polyhedral objects based on an optical proximity sensor system. In particular, we consider 1) the representation of a polyhedral object, and 2) the determination of the optimal sensor trajectory for the next probing. The object representation presented is based on two levels of hierarchy: the description of a 3-D structure by an inter-surface relation description table (SDT) and the Surface Normal Vector (SNV) distribution graph, and the description of individual surfaces by inter-edge relation description tables (EDTs). The partially filled SDT and EDTs of the test object are matched against the SDT and EDTs of a model object to extract all the possible interpretations. In order to achieve the maximum discrimination among all the possible interpretations, the optimal sensor trajectory for the next probing is determined as follows: 1) Select the optimal beam orientation based on the SNV distribution graph of the Multiple Interpretation Image (MII), where the MII is formed with reference to the hand frame by localizing the test object based on individual interpretations, and 2) Determine the optimal probing plane by projecting the MII onto the projection plane perpendicular to the beam orientation and deriving the optimal path on the probing plane. Simulation results are shown.

1. Introduction

There has recently been a growing interest in developing a methodology for the recognition and localization of a 3-D object based on "active sensing" in which a sensor system collects data by actively scanning or probing over the object. The main goal of the above approach is, as pointed out by Grimson[8], the rapid determination of an object model and pose from sparse, noisy, and occluded sensory data by predicting an optimal sensor pose or trajectory that will force a unique interpretation of the object with as few data points as possible.

The accomplishment of the goal is based on the following procedure: 1) With the known features already obtained from the previous sensor readings, the system constructs an MII or a set of hypotheses representing all the possible interpretations on the object model and pose, 2) The system then determines an optimal sensor pose or trajectory for subsequent data collection, which provides a maximum distinguishability among all the possible interpretations. The research issues involved in implementing the above procedure include 1) an efficient matching between a test object and a model object with only partial information on the test object available for the system, 2) an efficient search of the Cartesian space for selecting the optimal sensor pose or trajectory that maximizes system performance, i.e. the minimization of data collection as well as sensor movement, 3) an object representation which best supports the matching and searching processes described above, and 4) a measure of distinguishability for a given sensor pose or trajectory under the consideration of measurement noise. The implementation details, however, should take into account the capability of a sensor system, specially with regard to the range of areas it can detect, and the type of features it can extract. For instance, the determination of an optimal trajectory or path is required for a point sensor, whereas the determination of an optimal sensor pose sufficies for an area sensor.

In dealing with scheduling of tactile sensor moves, Schneiter[7] presented a method to select an optimal path by testing the distinguishability of each path from the outside of the union boundary to the inside of the intersection boundary of an MII. Grimson[8] proposed determining of an optimal sensing position by testing the distinguishability of each partition of the sensing line (orthogonal to the sensing direction), generated from the projection of an MII onto the sensing line. Hutchinson et.al[6] partitioned the space into individual segment inside which a sensor detects the same features, and moved the view point to the segment providing most discrimination. These approaches, however, either do not consider an optimal continuous sensor trajectory by taking into account the intermediate sensor movement, or do not provide a general and rigorous definition of distinguishability, or have serious limitations in 3-D applications due to excessive computational complexity.

This paper presents the recognition and localization of a 3-D polyhedral object based on a proximity sensor system capable of measuring the distance and surface normal of a local object surface. In particular, this paper 1) determines an optimal continuous sensor trajectory by selecting optimal beam orientations from a surface normal vector distribution graph, and by determining an optimal probing plane by projecting an MII onto the projection plane perpendicular to a selected beam orientation, 2) establishes a hierarchical object representation to achieve efficiency in matching, and 3) defines a general form of discrimination power for the evaluation of a

CH2813-4/89/0000/0075/$01.00 1989 IEEE

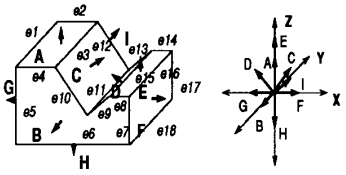


Figure 1: A polyhedral object and its Surface Normal (SNV) distribution graph

given trajectory.

2. Object Representation

A hierarchical representation of a 3-D polyhedral object is presented, where the upper level describes the inter-surface relation of a polyhedral object, while the lower level describes individual surfaces by defining the inter-edge relation of each polygonal surface. The representation also includes a surface normal vector distribution graph. The representation scheme presented here is "complete" in terms of reconstructability and "efficient" in terms of matching.

2.1 Representation of Inter-Surface Relation

The inter-surface relation of a polyhedral object, as shown in Fig. 1(a), is organized into the Inter-Surface Relation Description Table(SDT), as shown in Table 1. Each diagonal entry of the SDT contains the address of the Inter-Edge Relation Description Table(EDT) of the corresponding surface. Each non-diagonal entry contains a 3-tuple: (Feature1, Feature2, Edge). *Feature*1 specifies one of the following four types of the inter-surface relation: adjacent and convex(I), adjacent and concave(N), non-adjacent and non-parallel(E), and non-adjacent and parallel(D). *Feature*2 defines the *angle* between the SNVs of a pair of non-parallel surfaces; or the *distance* between a pair of parallel surfaces. *Edge* shows the label of the edge formed by a pair of adjacent surfaces; an entry whose corresponding surfaces are not adjacent remains blank. SDTs of a model object and a test object will be called MSDT and TSDT, respectively.

The Surface Normal Vector(SNV) distribution graph of an object is also included in the object representation to emphasize the structural characteristics of the object and to achieve computational efficiency in determining optimal probing plane (refer to section 4). Besides, the SNV distribution graph is required for supporting a unique and unambiguous representation, since the SDT representation by itself is not sufficient for the unique reconstruction of the object due to the mirror image problem[1]. The SNV distribution graph of an object is obtained by translating all the unit SNVs of the surfaces of the object to the origin of the object frame, while making the vector addition of any overlapping SNVs, as illustrated by Fig. 1(b).

2.2 Representation of Inter-Edge Relation

A diagonal entry of the EDT represents the length of the corresponding edge. Each of the non-diagonal entries of the upper right half of the EDT is specified by a 4-tuple: (Feature1, Feature2, Vertex, Type). *Feature*1 specifies one of four available types of the inter-edge relation: I, N, E, and D. *Feature*2 defines the angle or the distance between the edge orientation vectors of a pair of edges. *Vertex* represents the coordinate of the vertex generated by the corresponding pair of edges w.r.t. the object frame. *Type* indicates whether the vertex is Real or Imaginary, where an imaginary vertex is generated by a pair of non-adjacent and non-parallel edges. The EDTs of a model object and a test object will be represented respectively by MEDT and TEDT. The EDT for surface B of the object shown in Fig. 1(a) is illustrated in Table 2.

2.3 Construction of TSDT from Sensor Data

A proximity sensor system acquires the following sequence of data by a probing operation (refer to section 4) on a test object:

[1]The SDT of a non-symmetric polyhedral object can generate two different objects, each of which is a mirror image of another. This non-uniqueness problem occurring in the object reconstruction is called the *mirror image problem*.

Table 1: SDT of the polyhedral object given in Fig. 1

Surface	A_M	B_M	C_M	D_M	E_M	F_M	G_M	H_M	I_M
A_M	SDT_A	(I,90,e_4)	(I,135,e_3)	(E,34,)	(D,1.6,)	(E,90,)	(I,90,e_1)	(D,4.3,)	(I,90,e_2)
B_M		SDT_B	(I,90,e_{10})	(I,90,e_9)	(I,90,e_8)	(I,90,e_7)	(I,90,e_6)	(I,90,e_5)	(D,3.9,)
C_M			SDT_C	(N,281,)	(E,45,)	(E,45,)	(E,45,)	(E,45,)	(I,90,e_{12})
D_M				SDT_D	(I,124,e_{15})	(E,34,)	(E,34,)	(E,56,)	(I,90,e_{13})
E_M					SDT_E	(I,90,e_{16})	(E,90,)	(D,2.7,)	(I,90,e_{14})
F_M						SDT_F	(D,10,)	(I,90,e_{18})	(I,90,e_{17})
G_M							SDT_G	(I,90,e_{21})	(I,90,e_{20})
H_M								SDT_H	(I,90,e_{19})
I_M									SDT_I

Table 2: EDT of surface B of the polyhedral object given in Fig. 1

SDT_B	e_4	e_3	e_6	e_7	e_8	e_9	e_{10}
e_4	3.5	(I,90,(0,0,4.3),R)	(D,4.3,,)	(E,90,(10,0,4.3),I)	(D,1.6,,)	(E,34,(10.85,0,4.3),I)	(I,135,(3.5,0,4.3),R)
e_3		4.3	(I,90,(0,0,0),R)	(D,10,,)	(E,90,(6,0,2.7),I)	(E,56,(0.0,-2.22),I)	(E,45,(0,0,7.8),I)
e_6			10	(I,90,(10,0,0),R)	(D,2.7,,)	(E,56,(3.85,0,0),I)	(E,45,(7.8,0,0),I)
e_7				2.7	(I,90,(10,0,2.7),R)	(E,34,(10,0,3.78),I)	(E,45,(10,0,-2.2),I)
e_8					1.8	(I,124,(8.2,0,2.7),R)	(E,45,(5.1,0,2.7),I)
e_9						$\sqrt{5.44}$	(N,101,(6.2,0,1.6),R)
e_{10}							3.82

$$[(V_1, r_1)(V_2, r_2)...]\ [(V_j, r_j)(V_{j+1}, r_{j+1})...]...[(V_k, r_k)...(V_n, r_n)] \quad (1)$$

where V_j and r_j are the SNV and the distance measured with reference to the sensor frame at the jth data collection, and the brackets are used to cluster those tuples having similar SNV values. The raw data sequence (1) obtained by a probing operation is transformed into the following sequence of object surfaces:

$$[(N_1, R_1)(N_2, R_2)...(N_k, R_k)...(N_m, R_m)] \quad (2)$$

where N_k represents the unit surface normal vector of the kth surface w.r.t. the world frame, and R_k represents the distance of the kth surface from the origin of the world frame.

In transforming (1) into (2), the SNVs measured around or on an edge are identified and deleted from the sequence, since the SNV and distance in (1) are the average over a small area of the object surface.

With a pair of SNVs, $\{N_i, N_j\}$, in (2), corresponding to a pair of surfaces $\{S_i, S_j\}$ of a test object, the *types* of individual entries of the TSDT can be derived as follows: 1) Adjacency: S_i and S_j are adjacent when N_i and N_j are neighboring in (2), and the distance sequence of the two brackets corresponding to N_i and N_j is continuous, 2) Parallelism: S_i and S_j are parallel when N_i is equal to N_j, and R_i is different from R_j. (If $\{N_i, R_i\}$ and $\{N_j, R_j\}$ are not neighboring in (2) and yet they have the same value, then they can be considered either as parallel surfaces or as the same surface.) 3) Convexity: S_i and S_j form a convex shape when N_i and N_j are neighboring in (2), and the vector from N_i to N_j, $N_j - N_i$, and the directional vector, $\mathbf{p}$, of a probing have a positive inner product: $(N_j - N_i) \cdot \mathbf{p} > 0$.

The *values* of the features are determined as follows: the *angle* between a pair of non-parallel SNVs is determined by $cos^{-1}(N_i \cdot N_j)$, and the *distance* between a pair of parallel SNVs is determined by $\mid R_i - R_j \mid$.

The inter-edge relation features are determined between pairs of edges, where individual edges are generated by pairs of adjacent surfaces, based on the same procedure used for the construction of the TSDT.

3. Object Localization

The match of the test object and the model object is tested based on the TSDT and the MSDT as well as the TEDTs and MEDTS, with the partial information obtained for the test object. In case a successful match occurs, the surface and edge labels of the model object are assigned to the corresponding surfaces and edges of the test object. Such an assignment generates a possible "Interpretation" for the test object.

Since there may exists many interpretations of the test object, further probings are needed to prune out infeasible interpretations until a unique interpretation remains. This prompts the necessity of selecting an optimal probing which provides maximum discrimination among all the possible interpretations. The selection of an optimal probing is based on a Multiple Interpretation Image defined as follows:

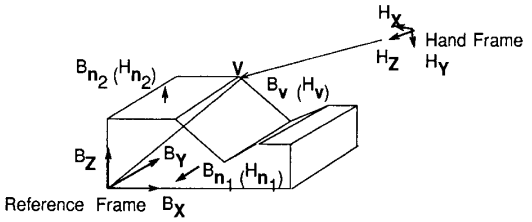


Figure 2: Defining the present hand position referring to reference vertices

> **Definition:** *Multiple Interpretation Image(MII)* An MII for a test object is the superposition of the images of the model object corresponding to individual interpretations. An MII is represented in Cartesian space with reference to the world frame. An image of the model object can be obtained by using rotational transformation between the object frame and the sensor frame, which is obtained from localizing the test object based on a particular interpretation.

The construction of an MII requires the localization of the test object based on individual interpretations. The sufficient conditions for unambiguous localization of a test object are as follows: 1) the SNVs and distances of three independent surfaces are known, 2) the SNVs and the distances of two non-parallel surfaces and one (actual/imaginary) vertex position are known, or 3) the SNV and distance of one surface and two (actual/imaginary) vertex positions are known.

Assuming the sufficient conditions are satisfied, the transformation ${}^{H}R_B$ from the object frame to the sensor frame and the position vector ${}^{H}P$ can be derived (refer to Fig. 2), where ${}^{B(H)}V$ is the position vector of vertex V w.r.t. the object (hand) frame and ${}^{B(H)}n_i$ is the surface normal vector of surface i w.r.t. the object (hand) frame:

$$ {}^{H}V = {}^{H}P + {}^{H}R_B\ {}^{B}V \quad (3)$$

$$ {}^{H}n_1 = {}^{H}R_B\ {}^{B}n_1 \quad (4)$$

$$ {}^{H}n_2 = {}^{H}R_B\ {}^{B}n_2 \quad (5)$$

${}^{H}R_B$ is obtained by

$$ {}^{H}R_B = {}^{H}N\ {}^{B}N^{-1} \quad (6)$$

where

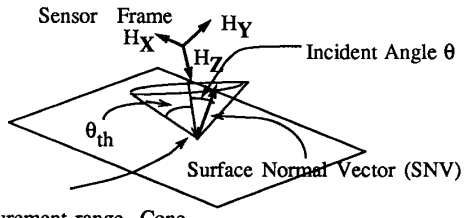
$$ {}^{H}N = [\ {}^{H}n_1,\ {}^{H}n_2,\ {}^{H}n_1 \times {}^{H}n_2] \quad (7)$$

$$ {}^{B}N = [\ {}^{B}n_1,\ {}^{B}n_2,\ {}^{B}n_1 \times {}^{B}n_2] \quad (8)$$

Once ${}^{H}R_B$ is obtained, ${}^{H}P$ can be computed from Eq. (3).

4. Optimal Probing

An additional probing of the test object is required in the cases where previous probings have failed to provide sufficient information for the unique recognition and localization of the test object. We define a probing as follows:



SNV Measurement range Cone

Figure 3: SNV Measurement Range Cone for a given Beam Orientation

Definition: *Probing*
A probing is defined here as a series of data (depth and orientation) collection operations on the test object by a proximity sensor system, carried out by moving the sensor frame along a trajectory satisfying the following constraints: 1) The orientation of the light beam with reference to the world frame is kept constant during a probing, 2) The trajectory of the origin of the sensor frame lies on a linear plane called a probing plane, and 3) The depth, defined as the distance from the origin of the sensor frame to a probing point on an object surface, is kept within the measurement range.

More precisely, a probing plane π, in which the probing trajectory lies, is defined by a beam orientation vector ${}^{W}l$ and a probing direction vector ${}^{W}v$: ${}^{W}n_{\pi} = {}^{W}l \times {}^{W}v$, where ${}^{W}n_{\pi}$ represents the surface normal vector of π and the superscript w indicates that the vectors are represented w.r.t. the world frame. Then, we can consider the determination of the optimal probing plane π^* as the problem of finding the optimal combination of a light beam orientation ${}^{W}l^*$ and a probing direction ${}^{W}v^*$.

4.1 Optimal Light Beam Orientation

Given a light beam orientation, the sensor system can measure only those surfaces whose SNVs reside inside the cone, as illustrated in Fig. 3. The cone is referred to as "a SNV range cone".

We first select a list of candidates for the optimal beam orientation, based on the "SNV Distribution Graph (SNVDG) of a Multiple Interpretation Image (MII)". The optimal cone orientation will be finalized later out of the selected candidates by considering the optimal combination of a beam orientation and a probing direction. The generation of a list of candidates for the optimal cone orientation is based on "the SNV clustering algorithm" composed of two phases: 1) the generation of tentative candidates, and 2) the refinement of the tentative candidates towards the local maxima:

SNV Clustering Algorithm

Phase 1: Generation of the list of Tentative Candidates for Optimal Cone Orientation

1. Select a SNV S_k in the SNVDG of the MII, and pose the cone apex of the SNV range cone at the origin of the SNVDG of an MII by making the cone axis coincide with the selected SNV S_k.
2. Generate a SNV cluster $C(S_k)$ by collecting all the SNVs inside the cone, and evaluate the cluster in terms of its discrimination power $M_p[C(S_k)]$. (Refer to Eq. (9))
3. Repeat 1) and 2) for all the individual SNVs, S_k, in the SNV Distribution Graph of an MII.
4. Order the SNV clusters based on their discrimination power; Select those SNV clusters whose discrimination power M_p is greater than the predefined threshold M_{pTH}, and put the SNVs corresponding to the selected SNV clusters in the list of tentative candidates for optimal cone orientation.

Phase 2: Refinement of Tentative Candidates for the Final Candidate List

1. Select a SNV S_k from the tentative candidate list if S_k is not included in the list of tentative candidates at the end of Phase 1, and set $S_k^{old} = S_k$. If all the SNVs in the candidate list are already selected, goto step 6.
2. For the SNV cluster $C(S_k)$ corresponding to a SNV S_k, generate a weighted cluster center $S_k^{weighted}$ by calculating the weighted average of all the SNVs in the cluster, where the weight of a SNV is proportional to the discrimination power of the corresponding SNV cluster:
$S_k^{weighted} = \sum_{S_i \in C(S_k^{old})} \sigma_i S_i, \sigma_i = M_p[C(S_i)] / \sum_{S_i \in C(S_k)} M_p[C(S_i)]$,
where $M_p[C(S_i)]$ represents the discrimination power of $C(S_i)$.
3. Move the cone axis from S_k^{old} to S_k^{new} incrementally toward the weighted cluster center $S_k^{weighted}$:
$S_k^{new} = S_k^{old} + \triangle S_k$, $\triangle S_k = \alpha(S_k^{weighted} - S_k^{old})$, where α is the scaling constant.
4. Generate a new cluster $C(S_k^{new})$ corresponding to S_k^{new} and evaluate its discrimination power $M_p[C(S_k^{new})]$.
5. If $M_p[C(S_k^{new})] \geq M_p[C(S_k^{old})]$, then set $S_k^{old} = S_k^{new}$ and goto step 2. If $M_p[C(S_k^{new})] < M_p[C(S_k^{old})]$, then set $S_k^{refined} = S_k^{old}$ and put $S_k^{refined}$ in the list of refined candidates. Go to step 1.
6. Finally, select the predetermined number of candidates from the combined list of refined candidates and the tentative candidates (from the phase 1), in a decreasing order of discrimination power.

Note that the phase 2 of the SNV clustering algorithm can be omitted, in case minimal computational complexity is a premium, without major degradation of system performance.

The measure of discrimination power, evaluating the performance of individual SNV clusters, is a function of 1) the number of measurable SNVs in the cluster, and 2) the distribution of SNVs included in the clusters over the individual interpretations. Formally, the measure of discrimination power $M_p[C(S_i)]$ of a SNV cluster $C(S_i)$ is defined as follows:

$$M_p[C(S_i)] = \alpha U_n(C(S_i)) + \beta[m(A(C(S_i)))/(\gamma + \sigma(A(C(S_i))))] \quad (9)$$

where the first term $U_n(C(S_i))$ represents the normalized utility of $C(S_i)$, and the second term represents the number of SNVs included in $C(S_i)$ in terms of mean $m(A(C(S_i)))$ and variance $\sigma(A(C(S_i)))$ of the accumulated utility histogram over the interpretations. α, β, and σ are the weighting coefficients which are heuristically determined. Refer to the appendix for the definitions of individual terms.

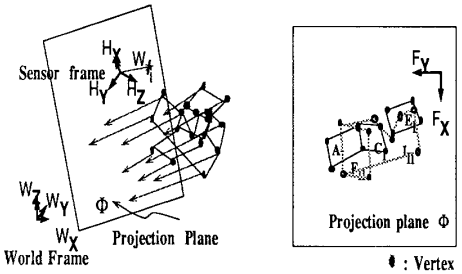


Figure 4: Generation of a surface image on the projection plane

4.2 Determination of Optimal Probing Plane

For each candidate for the optimal beam orientation ${}^W l_i$ obtained previously, an optimal probing direction ${}^W v_i$ can be selected such that a candidate for the optimal probing plane π, whose surface normal vector ${}^W n_\pi$ is determined by ${}^W n_\pi = {}^W l_i \times {}^W v_i$, can be generated. The determination of ${}^W v_i$, given ${}^W l_i$, is based on the following procedure:

1. From the MII for the test object, select all the vertices of those surfaces that belong to the candidate SNV cluster $C(S_i)$ and form a vertex distribution graph, as shown in Fig. 4(a). Each vertex is labeled with a particular interpretation(s) and a particular surface(s).

2. Define the projection plane Φ_i whose surface normal is the selected beam orientation ${}^W l_i$ and which includes the origin of the sensor frame from the last probing, i.e. $\Phi_i = \{{}^W X_i \mid {}^W l_i \cdot ({}^W X_i - {}^W S_i) = 0\}$. ${}^W S_i$ is the position vector of the origin of the sensor frame w.r.t. the world frame.

3. Project all the vertices located in the vertex distribution graph onto the projection plane Φ_i, as illustrated in Fig. 4(a).

4. Select the probing direction, ${}^W v_i$, on the projection plane Φ_i by finding a straight line on Φ_i which can provide the maximum discrimination power.

The details of the step 4 of the above procedure are listed in the following (refer to Fig. 5):

1. All the vertices projected on the projection plane Φ_i are once again projected onto the projection circle Ψ_i

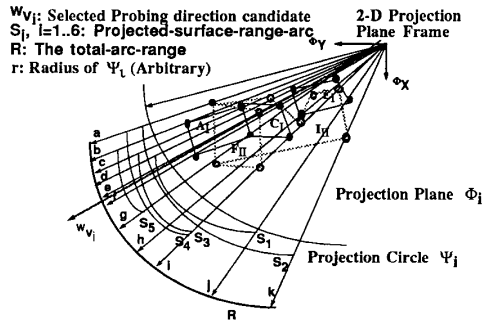


Figure 5: Determination of the probing direction ${}^W v_i$

by the projection rays emitted from the center of Ψ_i. The center of Ψ_i is the origin of the sensor frame and the radius r is arbitrarily determined. Then, the outer two vertices representing the boundaries of each surface on Ψ_i is connected forming an arc called a projected-surface-range-arc or simply a range-arc. The union of all the range-arcs on Ψ_i defines the total-arc-range.

2. The total-arc-range is partitioned into a number of arc segments with the distinctive boundary points of individual range-arcs. An arc-segment represents the intersection of a number of range-arcs.

3. For each arc-segment a_k, collect all the surfaces corresponding to the range-arcs which include an arc-segment a_k, and form an arc-segment cluster $C(a_k)$. Evaluate $C(a_k)$ using $M_p[C(a_k)]$ defined in Eq. (9).

4. Repeat step 3 for all the arc-segments, $a_k, k = 1..m$, and select the one with the highest measure of discrimination power as the optimal arc-segment a^*. Then, ${}^W v_i$ is determined by the direction passing through the center of the optimal arc-segment a^*.

The above procedure associates the SNV cluster $C(S_i)$ with the pair $({}^W l_i, {}^W v_i)$ and associates $M_p[C(a_k^*)]$ to the measure of discrimination power $M_p[C(S_i)]$ for each candidate for the optimal beam orientation. Finally, the optimal probing plane is determined by selecting the optimal $({}^W l^*, {}^W v^*)$, where $({}^W l^*, {}^W v^*) = \{({}^W l_i, {}^W v_i) \mid \max_i M_p[C(S_i)]\}$.

4.3 Generation of Probing Trajectory

The trajectory of the origin of the sensor frame, called the probing trajectory, is now defined on the optimal probing plane determined above. Here, we let the automatic surface tracking mechanism of a proximity sensor system[10] automatically determine the probing trajectory on the optimal probing plane, where the tracking mechanism maintains constant depth of a sensor from the probing point, while avoiding collisions. Along the probing trajectory, the system detects the presence of a certain surface at the assumed position specified by the MII and prunes out infeasible interpretations.

For example, in Fig. 6, six surfaces of the test object

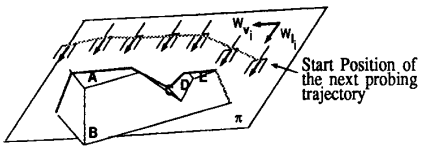


Figure 6: Determined optimal probing trajectory

intersect with the probing plane in the following order: *B-E-D-C-A-G*. Among those six surfaces, the SNVs of surface *B*, *D*, and *G* are not detectable by the sensor system since they are located outside the SNV range cone generated by ${}^{W}l^{*}$. The first surface detected by the probing is surface *E* which is associated with interpretation I. Thus, the surfaces associated with interpretation I remain in the MII of the test object, while other surfaces associated with interpretation II are eliminated from the MII. Once infeasible interpretations are pruned out in this manner, the MII is updated accordingly such that the next surface expected to be in contact during the probing can be correctly identified.

5. Simulation

The recognition and localization algorithm presented has been tested on various polyhedral objects by simulation. For instance, by assuming a specific initial probing trajectory on the test object shown in Fig. 1(a), four possible interpretations are obtained and the MII for the test object is constructed as shown in Fig. 7. Based on the interpretations, 5 measurement cones are selected and the projected images of the surfaces generated by the selected cones are shown in Fig. 8. From the projected image in Fig. 8, the corresponding probing directions are derived using the projection circle shown in Fig. 9. The center of the 15th arc-segment (78.32°) of the first candidate for the optimal light orientation is selected as the probing direction.

In the simulation, the system has been evaluated from two view points: 1) the number of probings required for a complete recognition and localization, and 2) the time complexity of the algorithm. If the goal is to localize a known (model) object and p interpretations for the test object are possible from the previous k probings, then the maximum number N of probings for complete localization is:

$$N = k + \max_{i=1}^{q} [\ \text{card}\ \{M_{l_i}\}\ ]$$

where M_{l_i} is a disjoint subset defined in the Appendix, and $card\{M_{l_i}\}$ is the number of interpretations that the disjoint minimal set M_{l_i} covers; $\sum_{i=1}^{q} card\{M_{l_i}\} = p$. If we assume that k previous probings detected n_k surfaces among n surfaces, then the number of probings for complete recognition is the number of measurement cones that disjointly covers $(n - n_k)$ SNVs of the model object.

The time complexity of the recognition algorithm mainly depends on the time complexities of the matching process and the optimal probing process. The total time complexity of the recognition algorithm can be expressed as $O((m^2+p^2+q)\,n^2)$, where m is the number of detected surfaces, and q is the number of candidates for the optimal light orientation.

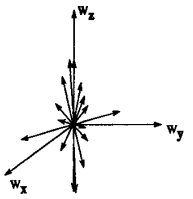
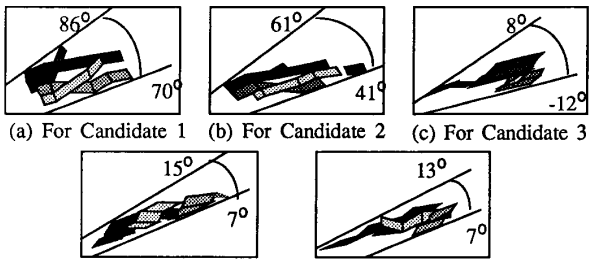


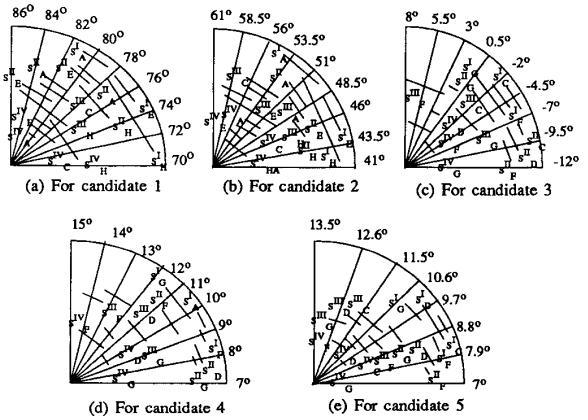
Figure 7: SNV Distribution graph of the Multiple Interpretation Image for the test object



(a) For Candidate 1 (b) For Candidate 2 (c) For Candidate 3

(d) For Candidate 4 (e) For Candidate 5

Figure 8: Surface Images projected on the projection plane



(a) For candidate 1 (b) For candidate 2 (c) For candidate 3

(d) For candidate 4 (e) For candidate 5

Figure 9: Arc-Segments projected on the projection line

6. Conclusions

The recognition and localization of a 3-D object based on "active sensing" provides several advantages over fixed-view sensing: 1) it can reduce a large amount of data processing time by using a simple sensor detecting simple features, 2) it can actively search for critical features having high discrimination power based on an optimal probing strategy, 3) it can sequentially prune out infeasible interpretations among all the

possible interpretations, and 4) it can easily handle an occluded or, even, invisible object. The price to be paid for the above advantages is the necessity of determining an optimal probing, as well as sequential sensor movement.

This paper presented an algorithm which determines an optimal probing, an optimal continuous sensor trajectory of a proximity sensor system, for recognizing and localizing a 3-D object as well as a hierarchical object representation with a SNV distribution graph for efficient matching. This paper also introduced the measure of discrimination power based on utility concept to be used as a general form of distinguishability. An extension of this work is underway to include the recognition and localization of a 3-D object having regular curved surfaces such as cylinders, spheres, and cones.

Acknowledgement

This work was supported in part by Jet Propulsion Laboratory, California Institute of Technology, under Contract 956501.

References

[1] Aristides A. G. Requicha *Representation for Rigid Solids: Theory, Methods, and Systems.* ACM Computing Surveys, Vol. 12, No. 4, December 1980

[2] Utpal Roy and C.R.Liu *Feature-Based Representational Scheme of A Solid Modeller For Providing Dimensioning and Tollerancing information* Robotics and Computer-Integrated Manufacturing, Vol. 4, No 3/4, pp. 335-345, 1988

[3] S Joshi and T C Chang *Graph-based heuristics for recognition of machined features from a 3-D solid model* Computer-Aided Design. March 1988.

[4] Shun-En Xie and Thomas W. Calvert *CSG-EESI: A New Solid Representation Scheme and a Conversion Expert System* IEEE Transaction on PAMI, Vol. 10, No. 2, March 1988

[5] Peper C. Gaston and Tomas Lozano-Perez *Tactile Recognition and Localization Using Object Models: The Case of Polyhedra on a Plane.* IEEE Transaction on Pattern Recognition and Machine Intelligence, Vol. PAMI-6, No.3, May 1984

[6] S.A.Hutchinson, R.L.Cromwell,and A.C.Kak *Planning Sensing Strategies in a Robot Work Cell with Multi-Sensor Capabilities* 1988 IEEE Conference on Robotics and Automation. pp. 1068-1075

[7] John L. Schneiter *An Objective Tactile Sensing Strategy for Object Recognition and Localization.*pp.1262-1267, Proceedings of the 1986 IEEE Conference on Robotics and Automation

[8] W. Eric L. Grimson *Sensing Strategies for Disambiguating Among Multiple Objects in Known Poses* IEEE Journal of Robotics and Automation, Vol. RA-2, No.4, December 1986

[9] M. Dhome and T. Kasvand *Polyhedra Recognition by Hypothesis Accumulation* IEEE Transaction on PAMI, Vol. PAMI-9, No. 3, May 1987

[10] Sukhan Lee and Hern S. Hahn *Design of a Miniaturized Optical Proximity Sensor System: HexEYE* JPL Report 1988.

Appendix: The measure of discrimination power $M_p[C(S_i)]$ of a SNV cluster $C(S_i)$

To derive the utility of a SNV cluster, we define the utility of a SNV S_a as follows:

$$\begin{aligned} U(S_a) &\triangleq Prob(S_a)Prune(S_a) + Prob(\bar{S}_a)Prune(\bar{S}_a) \\ &\triangleq U_{\frac{1}{2}}(S_a) + U_{\frac{1}{2}}(\bar{S}_a) \end{aligned}$$

For a SNV cluster $C(S_i) = \{S_l, j = 1..p\}$, where S_l is associated with a set of interpretations M_l, the normalized utility, $U_n(C(S_i))$, of a SNV cluster $C(S_i)$ is defined as follows:

$$U_n(C(S_i)) \triangleq U(C(S_i))/(n-1) = \sum_{l=1}^{q} U_{\frac{1}{2}}(S_l')$$

n: the number of interpretations for an MII

where S_l' is derived from Disjoint Minimal Covering which is defined as follows:

Definition: Disjoint Minimal covering, DMC(N) of a set N: A collection of disjoint subsets, $\{M_l', l = 1..n\}$,$M_{l_1} \cap M_{l_2} = \phi, \forall M_{l_1}, M_{l_2} \in \{M_l', l = 1..n\}$ and $l_1 \neq l_2$, covering the set N, $N = \cup_{l=1}^{n} M_l'$, is called the disjoint minimal covering DMC(N) of N:

$$DMC(N) \;=\; \{M_l', l = 1..n\}$$

Definition: $DMC(N/C(S_j))$
A disjoint minimal covering of N, which is derived from a cluster of SNV $C(S_i) = \{S_j, j = 1..p\}$ and $S_j \sim M_j, j = 1..p$, is defined as $DMC(N/C(S_j))$.

Definition: Accumulated Utility Histogram over Interpretations:
For each interpretation i, we accumulate the individual utilities of SNVs associated with the interpretation i:
$A_l(C(S_i))=\sum U_n(S_k)$, $\forall\ S_k \in C(S_i)$, $S_k \sim M_k$, interpretation $l \in M_k$
$A(C(S_i)) = \{A_l(C(S_i)), l = 1..n\}$ forms the accumulated utility histogram over the interpretations.
An accumulated Utility histogram over the interpretations is characterized by its mean $m(A(C(S_i)))$ and its variance $\sigma^2(A(C(S_i)))$.