HOUGH CLUSTERING TECHNIQUE FOR SURFACE MATCHING

S.M. BHANDARKAR & MINSOO SUK

Dept of Electrical & Computer Engineering
Syracuse University
Syracuse, NY 13244-1240 (U.S.A.)

ABSTRACT

This paper presents a 3-D multiple object recognition technique using Hough clustering. The objects are modeled as polyhedra and the input image is a range image containing instances of the modeled objects. The features used for matching are surface discontinuity features namely dihedral junctions. Qualitative reasoning based on qualitative attributes assigned to scene features is shown to be an important aspect of the Hough clustering algorithm presented in this paper. The resulting algorithm is robust and well suited for multiple-object scenes with partial occlusion.

1. INTRODUCTION

This paper concerns the problem of 3-D object recognition via localization in a multiple-object scene with partial occlusion. The previous approaches to object recognition via localization based on the Hough (Pose) clustering approach[1] or the Interpretation Tree (I.T.) approach[2,3] concentrated mainly on a single-object scene. The I.T. approach has been generalised by Grimson[4] to deal with multiple-object scenes but with a loss in performance. Straightforward application of the Hough clustering technique to a multiple-object scene leads to generation of several spurious hypotheses which make accurate identification and localization difficult. In this paper we present a robust Hough clustering technique for 3-D object recognition in a multiple-object scene. Qualitative reasoning based on qualitative attributes assigned to scene features is shown to be an important aspect of this technique. Qualitative reasoning greatly reduces the number of spurious hypotheses generated and tested by providing a means for (i) intelligent use of geometric constraints and (ii) effective means of scheduling pose hypotheses for verification. The use of qualitative reasoning shows how both, efficiency and robustness can be incorporated in the Hough clustering algorithm.

2. HOUGH CLUSTERING ALGORITHM

The input images are range images of typical 3-D polyhedral objects such as cube, square pyramid and hexagonal cylinder. The objects are allowed to have six degrees of freedom. The range images contain multiple objects with partial occlusion. The process proceeds by extracting primitive geometric features from the scene in the form of dihedral junctions (junctions with a single vertex and two incident edges). Each match of a scene feature with a model feature yields a geometric transform which can be represented as a point in the six-dimensional Hough space. Subsequent clustering in Hough space generates hypotheses regarding the identity and pose of the object. The technique described in this paper proceeds iteratively by recognizing each object in turn, verifying the identity and pose of each object and recomputing the clusters in Hough space until all the objects in the scene have been identified and verified.

2.1 FEATURE EXTRACTION

Since the modeled objects were polyhedra, there were two edge types to be considered. :-
1) Step edges which indicate a discontinuity in depth.
2) Roof edges which indicate a discontinuity in surface normal but not in depth. Roof edges could be further classified as either convex or concave.

Step edges were first detected using 1 x 3 and 3 x 1 gradient operators in both the x and y directions respectively. The roof edges were detected using a 3 x 3 Laplacian operator or a 5 x 5 Laplacian operator with averaging. The response of the step edges to the Laplacian operator was selectively suppressed since they were already detected and localized. The roof edges were further classified as convex corresponding to the positive maxima in the output of the Laplacian operator or concave corresponding to the negative maxima. The edges were thinned using an asynchronous thining algorithm based on pixel connectivity.

Linear boundaries were extracted from edge points using the 2-D Hough transform . Although the lines were 3-dimensional, the line detection was done in 2-D Hough space using a two dimensional (r, θ) accumulator. The straightforward technique of choosing all cells in the Hough accumulator whose values exceed a certain threshold was not successful since the peaks in the Hough space cover several cells and they overlap resulting in several extracted lines for a single peak. Instead an iterative histogramming technique was employed. The Hough transform was recomputed as each line was extracted from the image using an edge tracking algorithm. As each line was extracted from the image using the edge tracking algorithm, the pixels belonging to that line were labeled. The response of these pixels in successive computations of the Hough transform were selectively suppressed. The output of the boundary of the boundary extraction process was a list of boundary tokens. Each boundary token was represented by a data structure which gave the boundary label, edge type (step, convex roof or concave roof) and the x, y and z coordinates of the two endpoints of the boundary.

Although the basic line segments forming the boundaries of the objects were detected at this stage, further post-processing was necessary to fill the gaps in the boundary segments and form junction points. The output of the post-processing stage was a list of junction tokens. Furthermore if a boundary segment terminated at a **T** type junction then the boundary segment was further classified as *occluded*. Since **T** type junctions are highly viewpoint and scene dependent, they were not considered for matching. Each scene junction is assigned a degree which equals the number of edges which belong to the junction. Scene junctions of degree greater than 2 are decomposed into dihedral junctions.

The scene coordinate system attached to the display monitor was such that the positive direction of the z-axis pointed into the screen. This implied that all visible faces had outward surface normals with a negative z-component. Keeping this in mind, edges of a dihedral junction were so ordered that the vector cross product of the unit vector in the direction of the first edge with the unit vector in the direction of the second edge was in the direction of the outward surface normal. Face visibility was an important qualitative property used to constrain the matching process in Section 2.3.

2.2 FEATURE MATCHING

The dihedral junctions extracted from the range image were matched against dihedral junctions in the object model. Fig. 1 shows a candidate scene junction to be matched against a model junction. The match between the candidate scene and model junctions was subjected to a series of tests based on local geometric constraints guided by qualitative attributes assigned to scene boundaries. The result of the series of tests was a heuristic match quality assigned to the match or a rejection of the match. The series of tests applied to the candidate scene and model junction pair are as follows :-

Angle Constraint :-

$$M_\theta = K_1 \times (\theta_{\max} - |\theta_m - \theta_s|) \qquad \ldots(2.2.1)$$

where $\theta_{\max}$ is the maximum allowed deviation in angle and K_1 is a constant. θ_m and θ_s are the angles enclosed by the scene and model junctions respectively (Fig. 1). If M_θ is < 0 then the match is rejected else the match quality is incremented by M_θ .

Length Constraint :-

If the edge is <u>not occluded</u> then:

$$M_l = K_2 \times (L_{\max} - |L_m - L_s|) \qquad \ldots(2.2.2)$$

where $L_{\max}$ is the maximum allowed deviation in length and K_2 is a constant. L_m and L_s are the lengths of the model and scene edges respectively. If $M_l < 0$ then the match is rejected else it is incremented by M_l.

If the edge is <u>occluded</u> and if $L_s < L_m$ then :

$$M_l = K_3 \times (L_m - |L_m - L_s|) \qquad \ldots(2.2.3)$$

If $L_s < L_m$ the match is incremented by M_l else the match is rejected.

2.3 VIEWPOINT DETERMINATION

For a successful match between a scene feature and a model feature the viewpoint parameters were computed as described in the remainder of this section. The description is given in homogeneous coordinate systems and transformations. The coordinates (x, y, z) refer to the model coordinate system and the (u, v, w) to the scene coordinate system. The operations $\otimes$ and $\bullet$ denote the vector cross product and the vector scalar product respectively.

With reference to Fig. 1. let $\mathbf{m}_1$ be the unit vector in the direction **BA** and let $\mathbf{m}_2$ be the unit vector in the direction **BC.** Similarly let s_1 be the unit vector in the direction **ED** and s_2 be the unit vector in the direction **EF**. The homogeneous coordinates of B in model coordinate system is given by the column vector $[x_0, y_0, z_0, 1]^T$ and the homogeneous coordinates of E in the scene coordinate system are given by the column vector $[u_0, v_0, w_0, 1]^T$. The goal is to find a transformation **T** such that :

$$\mathbf{T}\ [x_0, y_0, z_0, 1]^T = [u_0, v_0, w_0, 1]^T \qquad \ldots(2.3.1)$$

There is an inherent ambiguity in the matching of the junctions as shown in Fig. 1. in the sense that whether $\mathbf{m}_1$ should match s_1 and $\mathbf{m}_2$ should match s_2 or vice versa. The directions of the outward normals $\mathbf{n_s}$ and $\mathbf{n_m}$ to the faces bound by the corresponding scene and model junctions, were used to resolve the ambiguity. In Fig. 1. since $\mathbf{n_s} = \mathbf{m}_1 \otimes \mathbf{m}_2$ and $\mathbf{n_m} = s_1 \otimes s_2$, $\mathbf{m}_1$ should match s_1 and $\mathbf{m}_2$ should match s_2 .

The transformation **T** was determined in a stepwise manner as outlined below:-

(1) Points B and E are translated to their respective origins. Let TRANS(-B) and TRANS(-E) denote the respective homogeneous transformation. This ensures that both junctions have their vertices translated to the origin.

(2) The vectors $\mathbf{m}_1$ and $\mathbf{m}_2$ are rotated about an axis $\mathbf{k} = (\ k_x, k_y, k_z\ )$ (where **k** is the unit vector in the direction of the axis) by a scalar magnitude of rotation θ . The corresponding homogeneous transformation is denoted by ROT(**k** , θ). ROT(**k**, θ) aligns $\mathbf{m}_1$ with s_1 and $\mathbf{m}_2$ with s_2 . The values of **k** and θ are computed as follows :

$$\mathbf{k} = (\mathbf{m}_1 - \mathbf{s}_1) \ \otimes \ (\mathbf{m}_2 - \mathbf{s}_2) \qquad \ldots(2.3.2)$$

$$\cos\theta = 1 - \frac{[1 - (\mathbf{m}_1 \bullet \mathbf{s}_1)]}{[1 - (\mathbf{k} \bullet \mathbf{m}_1)(\mathbf{k} \bullet \mathbf{s}_1)]} \qquad \ldots(2.3.3)$$

$$\sin\theta = \frac{[(\mathbf{k} \otimes \mathbf{s}_1) \bullet \mathbf{m}_1]}{[1 - (\mathbf{k} \bullet \mathbf{m}_1)(\mathbf{k} \bullet \mathbf{s}_1)]} \qquad \ldots(2.3.4)$$

3) The final transformation can be thus written as :

$$\text{ROT}(\ \mathbf{k}\ ,\ \theta\ )\ \text{TRANS}(-\text{B})\ [x_0, y_0, z_0, 1]^T \ = \ \text{TRANS}(-\text{E})\ [u_0, v_0, w_0, 1]^T \qquad \ldots(2.3.5)$$

From (2.3.1) and (2.3.5)

$$\mathbf{T} = \text{TRANS}^{-1}(-\text{E})\ \text{ROT}(\ \mathbf{k}, \theta\ )\ \text{TRANS}(-\text{B}) \qquad \ldots(2.3.6)$$

The transformation T from the model coordinate system to the scene coordinate system could be thus be written as:

$$\mathbf{T} = \text{ROT}(\mathbf{k}, \theta\ )\ \text{TRANS}(\ t_x,\ t_y,\ t_z\ ) \ =$$

$$\begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \ldots(2.3.7)$$

where

$$\begin{aligned}
r_{11} &= k_x^2\ (\ 1 - \cos\theta\ ) \ + \ \cos\theta \\
r_{12} &= k_x\ k_y\ (\ 1 - \cos\theta\ ) \ - \ k_z\ \sin\theta \\
r_{13} &= k_x\ k_z\ (\ 1 - \cos\theta\ ) \ + \ k_y\ \sin\theta \\
r_{21} &= k_x\ k_y\ (\ 1 - \cos\theta\ ) \ + \ k_z\ \sin\theta \\
r_{22} &= k_y^2\ (\ 1 - \cos\theta\ ) \ + \ \cos\theta \\
r_{23} &= k_y\ k_z\ (\ 1 - \cos\theta\ ) \ - \ k_x\ \sin\theta \\
r_{31} &= k_x\ k_z\ (\ 1 - \cos\theta\ ) \ - \ k_y\ \sin\theta \\
r_{32} &= k_y\ k_z\ (\ 1 - \cos\theta\ ) \ + \ k_x\ \sin\theta \\
r_{33} &= k_z^2\ (\ 1 - \cos\theta\ ) \ + \ \cos\theta \qquad \ldots(2.3.8)
\end{aligned}$$

and

$$\begin{aligned}
t_x &= u_0 - r_{11}x_0 - r_{12}y_0 - r_{13}z_0 \\
t_y &= v_0 - r_{21}x_0 - r_{22}y_0 - r_{23}z_0 \\
t_z &= w_0 - r_{31}x_0 - r_{32}y_0 - r_{33}z_0 \qquad \ldots(2.3.9)
\end{aligned}$$

Thus the transform **T** is uniquely specified by the 6-tuple (t_x, t_y, t_z, k_x, k_z, θ)

2.4 HOUGH CLUSTERING

There are two principal approaches to the implementation of the generalized k-dimensional Hough transform. (i) Using a k-dimensional accumulator array and (ii) Clustering k-dimensional feature vectors in Hough space using disparity matrices. The latter approach was chosen in this experiment for the following reasons :

(i) Implementation a k-dimensional accumulator array for k = 6 was not practical in terms of memory requirement.

(ii) Choosing an optimum quantization of the accumulator array was not an easy problem. Fine quantization would lead to the scattering of the peaks among several cells whereas coarse quantization caused a loss in accuracy.

In this experiment a disparity matrix was described for each object model. An element D(i, j) of the disparity matrix represents a match between the scene feature i and the model feature j. D(i, j) equals the geometric transform (t_x, t_y, t_z, k_x, k_z, θ) if the match is successful and NIL otherwise. Since we require that each cluster in the Hough space correspond to the occurrence of a single object from a single viewpoint, the following constraints were imposed during the clustering process :-

(i) In a given cluster no scene feature should match more than one model feature. Conflicting matches of a scene feature to more than one model feature are assigned to different clusters.

(ii) In a given cluster no model feature should match more than one scene feature. Conflicting matches of a model feature to more than one scene feature are assigned to different clusters.

In terms of disparity matrices, given two elements D(i, j) and D(m, n) of a single disparity matrix, if either i = m or j = n, the two elements represent conflicting matches in terms of the two constraints described above. It would have been difficult to impose these constraints using an accumulator array.

In order to initiate the clustering process, the initial seeds were chosen as follows :

(i) A location (i, j) in the disparity matrix where the match quality was a maximum was chosen as a cluster seed.

(ii) Locations (non-NIL) in row i and j were chosen as cluster seeds since they are in conflict with the cluster seed chosen in (i).

The clustering process was based on the k-means clustering algorithm where the initial cluster seeds were chosen as the initial values for the k-means. The clustering was done in the 6-dimensional (t_x, t_y, t_z, k_x, k_z, θ) space. At the end of the clustering process, each cluster mean represents a geometric transform or a pose hypothesis. Each pose hypothesis was also assigned a match quality which was the aggregate of the individual match quality measures of the cluster members.

2.5 POSE VERIFICATION

The output of the clustering process is a set of pose hypotheses to be further verified. Instead of verification by direct depth comparison as proposed by earlier researchers, verification by feature comparison was resorted to. Verification by direct depth comparison was found to be highly unreliable since slight errors in pose computation resulted in large errors in depth comparison.

The verification process could be described in a stepwise manner as described below :-

1) The pose hypotheses are ordered by an ordering function $O = K_1 (d_{max} - t_z) + M$ where d_{max} is the depth of the background, M is the quality of the of the hypothesis and K_1 is a constant. The ordering function ensures that the hypothesis with the least number of occluded features and the least depth (corresponding to the topmost object) is selected first for verification. The qualitative attribute based on occlusion assigned to scene features made such an ordering function possible.

2) For the selected pose hypothesis the corresponding object model is projected onto the scene. A three-dimensional window is defined around the projection. The window serves as a crude filter. If the number of unlabeled scene features in the window is less than a predefined threshold the hypothesis is rejected.

3) For hypotheses that pass stage 2) a more detailed comparison based on feature matching is carried out. Each projected model feature is matched to a scene feature within the window. Based on the proximity of junction points and the difference in angle, boundary length and boundary orientation, a match quality is defined for each match. The equations for the computation of the match quality are similar to the equations (2.3.1) and (2.3.2). An optimal global match quality is computed for the individual matches by treating the problem as an assignment problem for which polynomial-time algorithms such as the *Hungarian Marriage* algorithm are known to exist. A variant of the *Hungarian Marriage* algorithm was used in the experiment. If the global match quality exceeded a threshold, the hypothesis is accepted else rejected.

4) For a hypothesis which is accepted, the corresponding scene features are labeled as belonging to that particular object model. These scene features are removed from further consideration.

5) The remaining features are reclustered and steps 1) to 4) are carried out until all the scene features are labeled.

2.6 ROLE OF QUALITATIVE REASONING

Qualitative reasoning had an important role to play in the Hough Clustering process. Qualitative attributes assigned to scene features proved useful in the following situations :-
1) The qualitative property based on occlusion was used to selectively adjust the stringency of the geometric constraints. For unoccluded scene boundaries, a tighter geometric bound based on length equality was placed on the match (equation (2.2.2)) whereas for unoccluded scene boundaries a looser geometric bound based on length inequality was placed on the match (equation (2.2.3)). The appropriate selection of constants ensured that matches based on unoccluded features had a higher match quality assigned to them as compared to the matches based on occluded features.
2) The edge type (roof or step) was used to infer visibility of the face bound by the scene edges. Face visibility and the implied direction of the outward surface normal was used to resolve a potential ambiguity in the feature matching and viewpoint determination process as was seen in Section 2.3.
3) The qualitative attribute based on occlusion provided criteria for scheduling pose hypotheses for verification. Hypotheses with a fewer number of occluded features were

given higher priority over those with a greater number of occluded features. The scheduling function ensured that a fewer number of hypotheses were tested for verification. The conventional Hough clustering algorithm has no such provision.

The effect of qualitative reasoning on performance of the Hough clustering algorithm was experimentally verified. The experimental results are discussed in the following section.

3. EXPERIMENTAL RESULTS AND DISCUSSION

A comparative analysis of the Hough clustering technique both with and without qualitative reasoning was made by means of an experiment. The object models were CAD/CAM wireframe models of simple polyhedra such as cube, square pyramid and hexagonal cylinder. The range data was simulated using a z-buffer algorithm. Two candidate scenes were analyzed. The first set of experiments used the Hough clustering algorithm with qualitative reasoning whereas in the second set of experiments the qualitative attribute based on occlusion was neglected. Thus the geometric constraint based on length equality in equation (2.3.2) was replaced by the weaker constraint based on length inequality (2.3.3). The ordering function for scheduling hypotheses was suitably altered since the criteria based on the number of occluded features could no longer be used. The qualitative attribute based on surface visibility was also neglected leading to two potential matches for the scene and model junction pairs. The effect of the absence of qualitative reasoning was analyzed in terms of two measures (i) average number of hypotheses generated for each object labeled in the scene which is denoted by measure M1 and (ii) average number of hypotheses tested for each object labeled in the scene denoted by measure M2 . Figures 6 through 17 show the original range image, output of the edge extraction routines, output of the post-processing routines and the reconstructed scene after recognition and localization for the two candidate scenes that were considered in the experiment. Table I. summarizes the experimental results for the Hough clustering technique with qualitative reasoning and Table II. summarizes the the experimental results for the Hough clustering technique without qualitative reasoning.

TABLE I (With qualitative reasoning)

Scene#	#Obj	#hyp. gen	#hyp. tested	#obj. labeled	M1	M2
1	3	145	28	3	48	9
2	3	116	31	3	39	10

TABLE II (Without qualitative reasoning)

Scene#	#Obj	#hyp. gen	#hyp. tested	# obj. labeled	M1	M2
1	3	96	77	1	96	77
2	3	165	60	3	55	20

As is brought out by the experimental results, the Hough clustering algorithm gave far superior results when coupled with qualitative reasoning both in terms of measure M1 and measure M2 . In the case of Scene #1 the conventional Hough clustering algorithm was unable to generate reliable pose hypotheses for two of the three objects in the scene. This was primarily due to the excessively cluttered Hough space. In this case the algorithm was prematurely halted after all of the generated hypotheses were rejected on verification. The algorithm would have possibly run to completion had the verification criteria (the acceptance threshold) been relaxed but that would have implied a greater error in localization.

4. CONCLUSIONS

In this paper we have presented a robust and efficient Hough clustering algorithm well suited for 3-D object recognition and localization in multiple-object scenses with partial occlusion. The conventional techniques for recognition via localization such as Hough clustering or pruning of the Interpretation Tree are based on propagation of geometric constraints through local matches of geometric features. These algorithms perform poorly in multiple-object scenes with partial occlusion. Our algorithm which uses qualitative reasoning along with Hough clustering shows how qualitative reasoning can provide a means for (i) intelligent and selective use of geometric constraints and (ii) scheduling pose hypotheses for verification. Our experimental results show how qualitative reasoning when used in conjunction with techniques such as Hough clustering which rely mainly on propagation of geometric constraints could lead to greater efficiency and robustness. An important fact to be noted is that since the features used were primitive, the qualitative attributes used were fairly simple, yet effective. These qualitative attributes did not require extensive preprocessing of the image data as would be needed for a higher semantic-level description of the scene. This paper shows that qualitative reasoning has an important role to play within the recognition via localization approach to object recognition.

5. REFERENCES

1. G. Stockman, Object Recognition and Localization via Pose Clustering, *Comput. Vision Graphics Image Processing* **40**, 1987, 361-387.

2. P.C. Gaston and T. Lozano-Perez, Tactile Recognition and Localization using Object Models : Case of Polyhedra on a Plane, *IEEE Trans. PAMI*, **6(3)**, May 1984, 257-266.

3. W.E.L. Grimson and T. Lozano-Perez, Model-based Recognition and Localization from Sparse Range or Tactle Data, *Int. Journal. Robotics Research*, **3(3)**, Fall 1984, 3-35.

4. W.E.L. Grimson and T. Lozano-Perez, Localizing Overlapping Parts by Searching the Interpretation Tree, *IEEE Trans. PAMI* **9(4)**, July 1987, 469-482.

5. W.E.L. Grimson, The Combinatorics of Local Constraints in Model-Based Recognition and Localization from Sparse Data, *Journal ACM*, **33(4)**, October 1986, 658-686.

6. G. Stockman et-al, *Computing a Pose Hypothesis from a Small Set of 3-D Object Features*, Dept. of Computer Science Tech. Report, MSU-ENGR-87-001, Michigan State University, East Lansing, MI 48824, Jan 1987.

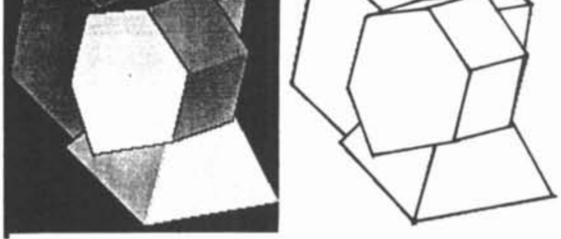

Fig 1. Matching Candidate Model and Scene Junctions

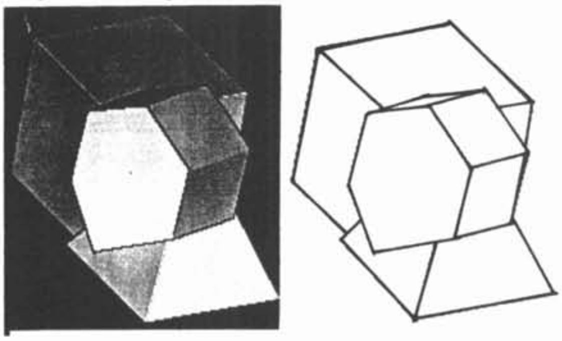

Fig 2. a) Original range Image b) reconstructed Image