MODELING 3-D STRUCTURE

Thomas C. Henderson, INRIA, Le Chesnay, FRANCE
Amar Mitiche, U. of Texas, Austin, Texas, USA

Abstract

Recognition of 3-D objects and the determination of their orientation in space are two major problems of robot vision systems. Moreover, in an industrial environment, these tasks should be performed quickly and accurately. A simple representation of 3-D objects is given which makes possible a technique for recognition and orientation determination of 3-D objects in laser range images. This technique is an extension of the 2-D Hough shape transform to handle 3-D surfaces ; the technique is applied directly to a set of 3-D points extracted from a range image.

1. Introduction

The representation of 3-D objects has received much attention, and a plethora of models have been proposed (see [1]). Most of these models permit a hierarchical organization of primitive solids (or volumes) and are based on constructive solid geometry, boundary, or sweep representations. The generality of such models leads to complex object description and detection schemes and this reduces their effectiveness. We present here a 3-D representation based on the Hough transform ; this representation is a simple and efficient description of the surface of the object and does not include structural information.

The classical Hough transformation is used to detect curves by mapping all the feature points of an image into a parameter space (see Iannino and Shapiro [2] for an introduction to the Hough transform and its applications). The parameter space depends on the class of curves to be detected, and in the case of straight lines is characterized by either the slope-intercept plane (see Hough [3] or Rosenfeld [4]) or by the angle of the normal to the line and the minimum distance from the line to the origin (Duda and Hart [5]). In practice, the parameter space is quantitized, and an accumulator is associated with each point in the parameter space. An accumulator is incremented for every detected point whose associated curve in parameter space crosses that accumulator.

The 2-D shape Hough transform as described by Davis and Yam [6] and Sloan and Ballard [7] is a generalization of the Hough transform. The 2-D shape transform is applied to edge images produced from 2-D intensity images. Efficient detection algorithms can be devised for arbitrary shapes by using the edge responses and taking advantage of the gradient at the edge pixel to reduce the ambiguity in the parameter space.

The current method is applied directly to laser ranging images, i.e., given an image, $I(i,j)$, then $(i,j,I(i,j))$ is the (x,y,z) location on the surface of an object (or the background). In our laser ranging system, the distance to the background is known, and the non-background points are extracted from the range image and kept as a list. The points in this list are called the detected points. The surface of a 3-D object is likewise modeled as a list of points, and the detection procedure is to match the set of model points with the detected points. We show how the Hough transform can be efficiently used to perform this matching even without the knowledge of the surface normal at each detected point. Both the 2-D and the 3-D applications of the Hough technique can be used to find partial matches.

Section 2 describes the representation of 3-D objects and gives an algorithm for position invariant matching. Section 3 shows how the method can be used for orientation invariant matching. Finally, Section 4 discusses data compression methods and limitations.

2. Position Invariant 3-D Hough Transform

The representation used is basically a generalization of that of Merlin and Farber [8]. Given a set of points $P = \{x_i, y_i, z_i)\}$, $i=1,n$ representing a 3-D object, choose some reference point, $P_0 = (x_0, y_0, z_0)$, e.g., the centroid of the object. The object representation, $\mathcal{O}(P,P_0)$, is given as $\mathcal{O} = \{(dx_i, dy_i, dz_i)\}$, where $dx_i = x_0 - x_i$, $dy_i = y_0 - y_i$, and $dz_i = z_0 - z_i$. $\mathcal{O}$ is then a characterization of P as a displacement from each point of P to the reference point P_0.

Given a set of detected points, $D = \{(x_i, y_i, z_i)\}$, $i=1,m$, use a 3-D array, H, to accumulate counts for possible locations of P_0 in space. Namely :

$\forall\ (x_i, y_i, z_i) \in D$, $i=1,m$

$\forall\ (dx_i, dy_i, dz_i) \in \mathcal{O}$, $j=1,n$

Increment $H(x_i+dx_i, y_i+dy_i, z_i+dz_i)$ by 1.

Then the location in H having the maximum value corresponds to the translated position of the reference point, P_0, of the object, $\mathcal{O}$.

The algorithm produces a uninque maximum for any translation of P, and the maximum value is equal to the number of object points in D. This is true since the algorithm is simply an efficient way of computing the (3-D) convolution of the object template with the detected surface points. It must be noted that if all the points in P are not in D, then the maximum will be less than n, and if there are several copies of the object, then the maximum may not be unique ; however, the reference point is always guaranteed to be among the maxima. The ratio $Hmax/|D|$ can be used to judge the likelihood that the maximum location does indeed correspond to P_0.

3. Rotation Invariant 3-D Hough Transform

Given a set of detected points in $D = \{(x_i, y_i, z_i)\}$, $i=1,m$, and an object representation $\mathcal{O}$ as described in Section 2, use $\mathcal{O}$ to define a set of radii, $R = \{r_i\}$, $i=1,k$ where the r_i's represent all the distinct lengths of vectors in $\mathcal{O}$. With every $r \in R$ associate a list, Sr, of offset vectors which describes the surface of the digital sphere of radius r. Then the rotation invariant 3-D Hough transform is computed by :

$\forall p = (x,y,z) \in D$

$\forall r \in R$

$\forall s = (dx,dy,dz) \in Sr$

Increment $H(x+dx, y+dy, z+dz)$ by 1.

The reference point, P_0, for the object representation $\mathcal{O}$ is found the same way as for the translation transform. However, there is now no guarantee of a unique maximum. Even if the maximum location does correspond to P_0 for a rotated version of the object, the orientation of the object remains unknown.

As suggested by Davis in the 2-D case, two reference points, P_0 and P_0', can be chosen and used to produce two distinct object representations, $\mathcal{O}$ and $\mathcal{O}'$. In this way, the vector $P_0 - P_0'$ has a direction and gives the orientation of the object. In 3-D, three reference points must be used.

4. Discussion

A direct model of a 3-D object in a 3-D array, i.e., the characteristic function in 3-D space is essentially empty and for direct implementation of the convolution would require a 3-D accumulator array which could easily exhaust the memory of a machine. Therefore, it is convenient (and necessary) to compress the size of the representation. We have developed an alternative approach which drastically reduces the set of accumulators. This is done by choosing two detected points and keeping accumulators only for the points of intersection of the various spheres centered at the two points. This can be further constrained by choosing k more points and checking that each hypothesized reference location lies on the surface of some sphere for each of the k points ; we currently use 2 such extra points. Note that the accumulators are kept as a list, and the quantization of the parameter space can be to any precision desired and can also vary from place to place.

A model of the object shown in Figure 1 was constructed. The object description contained 8334 surface points. Different views of the object were located under various transformations. For example, one view containing 914 points was correctly located with on the order of 700 points contributing to the accumulator (over 100 points more than for any other accumulator) at the transformed reference point. Obviously, the thresholds chosen for sampling the surface of the spheres will influence the number of points contributing to the maximum, and this threshold will be dependent on the sampling rate on the surface of the object.

Another way to reduce the size of the representation is to map each face of the 3-D object into a 4-D transform space and model these points considered as an object. Planar faces can be found, for example, using the technique described by Duda et al. [9]. However, they assume that intensity information is also available, and this aids in finding planar regions ; even so, finding planar regions is a non-trivial task. Once the set of faces are found, associate each face, fi, with the 4-D point (a_i,b_i, c_i,d_i) whose coordinates define the plane containing fi. The number of faces is usually small, and the corresponding 4-D points can be kept as a list instead of in a 4-D array. Obviously, the disadvantage is to locate the faces of the object ; moreover, the object may be curved and not have any planar faces.

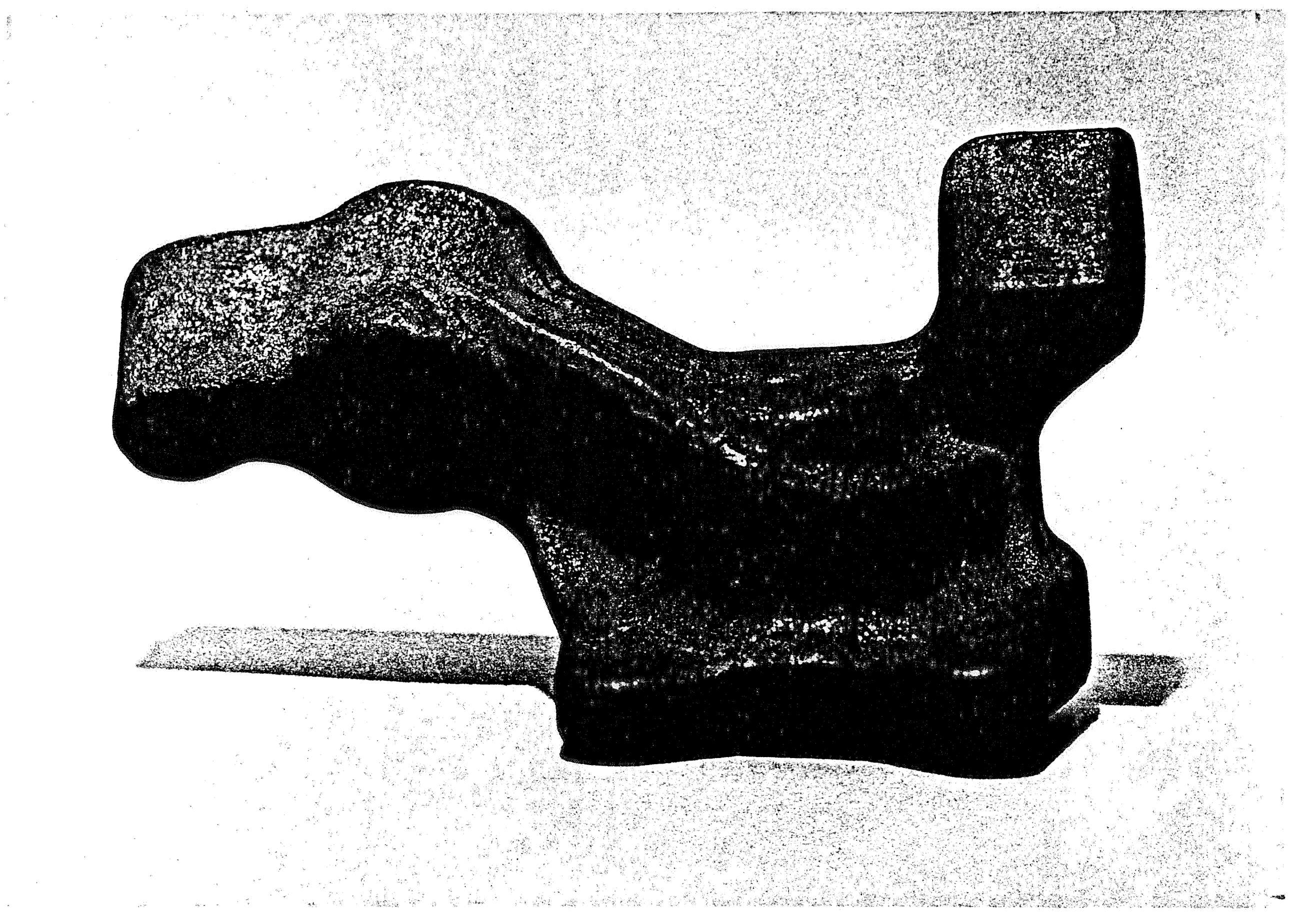

Figure 1. Workpiece (part of a Renault).

In summary, a fast technique for the recognition of 3-D objects in laser range images and for determination of their orientation in space has been demonstrated. Examples have been presented, and methods for reducing the memory requirements proposed.

References

1. Bajcsy, R., Workshop on the Representation of Three-Dimensional Objects, May 1-2,

Univ. of Penn., 1979.

2. Iannino, A. and S.D. Shapiro, "A Survey of the Hough Transform and Its Extensions for Curve Detection", Proc. of Patt. Rec. and Image Proc. Conf., Chicago, 1978, pp.32-38.
3. Hough, P.V.C., Method and Means for Recognizing Complex Patterns, U.S. Patent 3069654, 1962.
4. Rosenfeld, A. and A. Kak, Digital Picture Processing, Academic Press, N.Y., 1976, p.379.
5. Duda, R.O. and P.E. Hart, "Use of the Hough Transformation to Detect Lines and Curves in Pictures", Comm. ACM, 15, Jan. 1972, pp.11-15.
6. Davis, L.S. and S. Yam, "A Generalized Hough-like Transformation for Shape Recognition", TR-134, Univ. of Texas, Feb. 1980.
7. Sloan, K.R. and D. Ballard, "Experience with the Generalized Hough Transform", Image Understanding Workshop, April 1980, pp.150-156.
8. Merlin, P.M. and D.J. Farber, "A Parallel Mechanism for Detecting Curves in Pictures", IEEE Trans. Comp. C24, 1975, pp.96-98.
9. Duda, R.O., D. Nitzan and P. Barrett, "Use of Range and Reflectance Data to Find Planar Surface Regions", IEEE Trans. Patt. Anal. Mach. Intell., Vol. PAMI-1, n° 3, July 1979, pp.259-271.