SPATIAL OBJECT PERCEPTION FROM AN IMAGE *

Radu HORAUD **

Laboratoire d'Automatique de Grenoble
B.P. 46, 38402 Saint-Martin-d'Heres
France

ABSTRACT

In this paper we address the problem of finding the spatial position and orientation of an object from a single image. It is assumed that the image formation process and an object model are known in advance. Sets of image lines are backprojected and constraints on their spatial interpretations are derived. A search space is then constructed where each node represents a space feature with a model assignment. Next, a hypothesize-and-test recognition strategy is used to select a solution, that is to determine six degrees of freedom of a part from a set of features. Finally we discuss the efficiency and the reliability of the method.

1. INTRODUCTION

Among various aspects of perception for which Computer Vision is trying to find a computational theory, an intriguing one is the mechanism by which two-dimensional (2D) shapes are sometimes perceived as three-dimensional (3D) objects. There are at least two reasons for investigating this subject. First, there is evidence that people perform spatial reasoning whenever they deal with images. If they are asked to match two images of the same object, they rotate the object mindly in the 3D space even if the transform occurs in the image plane. This implies that the interpretation space is different from the image plane. And second, we are interested in devising a technique for interpreting images of known things : when is it possible to retrieve the six degrees of freedom of an object from a single view assuming that the object and the camera geometry are known in advance ? Which are the theoretical and practical limitations of such a method if it were implemented as a computer algorithm ? An interesting application could be the recognition of man made parts in an industrial environment.

In this paper we suggest on possible approach limited to objects bounded by planar faces. First we extract linear edges from an intensity image and these edges are combined to form angles and junctions which are assumed to be projections of 3D object vertices. These sets of image features are backprojected using an inverse perspective camera model and some constraints on their spatial position and orientation are derived. A search space is built where each node represents a 3D feature with a model assignment. A hypothesize-and-test recognition strategy implemented as a depth-first tree search is used to find a solution, that is three rotations and three translations for each part. The object model together with physical constraints are used as heuristics for reducing the complexity of the search space.

The work reported herein is supported by "Laboratoire d'Electronique et de Technologie de l'Informatique", Grenoble, France.

**The author is now with LIF1A, B.P. 68, 38042 Saint-Martin d'Heres, France.

Previous approaches for interpreting line drawings have generally not been designed to deal with the geometry of perspective. They have usually used orthography and the gradient space,[l], [2]. Others have simplified the problem by using the "support hypothesis" which reduces the problem to three degrees of freedom, [3] , [4]. Our approach is more general and it can include this hypothesis as a physical constraint. More recently the "gaussian mapping" has been introduced as a tool for interpreting perspective views, [5]. A method for finding vanishing points is described and one for retrieving the spatial orientation of planes by backprojecting angles and curvature is suggested. We extend these results to junctions which we believe are more useful than faces. A two-stage, model-based recognition procedure is described in [6]. The planning stage computes all possible appearances of an object in terms of sets of simultaneously visible features. The recognition stage consists in a predict-observe-backproject sequence. This approach is different from ours since it doesn't explore the constraints available with sets of features.

2. BACKPROJECTION OF IMAGE FEATURES

This paragraph utilizes the perspective camera model for interpreting image linear features. Let us recall briefly this model, [5]. A space point with camera coordinates (x,y,z) projects onto the image at (x.f/z, y.f/z, f) where f is the focal length (see Figure 1). The camera frame has its origin at the focal center and the image is parallel to the x-y plane at distance f from the center along the z-axis. A unit vector can be expressed as a point on a unit sphere centered at the origin, the gaussian sphere. A point on this sphere has two angles as coordinates, the azimuth (a) and the elevation (3). Hence, the orientation of a space plane or the direction of any image or space line can be represented as a point on this sphere. Let's now associate an interpretation plane with an image line. This

plane is defined by an image line and the focal center and it contains all the spatial interpretations of the image line. The possible directions of these spatial lines lie on a great circle, the intersection of the interpretation plane with the gaussian sphere. If we denote by P the vector normal to the interpretation plane and by L the space line direction vector, the equation of the great circle is :

$$L.P = 0 \qquad (1)$$

Similarly we can develop constraints for the spatial interpretations of image angles and junctions. The motivation for choosing these features is that they are the projections of object vertices. Let L_1, and L_2, be two image lines forming an angle. Their spatial interpretations are denoted L_1 and L_2 and their interpretation planes are denoted P_1 and Pp. L_1 and L_2 are constrained to be coplanar : they belong to a space plane S and form a space angle w. We are seeking the orientation of S when w is known. The following equations stand :

$$L_1 = S \wedge P_1, \; L_2 = S \wedge P_2 \qquad (2)$$

$$\cos w = L_1 . L_2 / ||L_1|| . ||L_2|| \qquad (3)$$

Since w is imposed, equation (3) provides a constraint for the possible orientations of the space plane S. Consider now a junction formed by three image lines, 1_1, 1_2 and 1_3 whose spatial interpretation is a right vertex with edges L_1,L_2 and L3. (There is no loss of generality in considering a right vertex ; this merely simplifies the exposition), l_1 and 1_2 can be combined just as above to form an angle constraint. Notice that L3 is parallel to the vector normal to the plane formed by L_1 and L_2. L_3 is constrained to lie on the great circle corresponding to the spatial interpretations of I3. Therefore, the only possible orientations of S are the intersection of this great circle (eq.(l)) with the angle constraint (eq.(3)). Figure 4 shows the solutions for the image junction indicated by an arrow on Figure 3. Only half of the gaussian sphere is projected and shown on Figure 4 with a (horizontal) varying from IT/2 to 3n/2 and 0 (vertical) varying from *-n/2* to n/2. The two solutions correspond to two orientations of S, one for a concave vertex and the other for a convex one. Without additional information it is impossible to decide which solution to select. This is a simplified version of the Necker's cube illusion. In [I], Kanade developed an analytical solution in the case of orthographic projection but his method requires the measurement of the skewed symmetry of all the faces forming a junction.

3. IMAGE TO OBJECT CORRESPONDENCE

The ultimate goal of a recognition procedure is to assign an object model to a set of image features and to find the spatial parameters of each object. These parameters will be embedded in a 4x4 homogeneous matrix that maps an object from model coordinates to camera coordinates. Let us show now how such a transform may be computed.

We describe first a simple scheme for modelling objects within the context of visual recognition. For a more complete discussion, see[7J. Such a model contains lists of those features and combinations of features that are the most likely to be detected in an image. The features are also ranked according to the contribution they can make for recognition. The model of an object bounded by planar faces provides a list of all faces with pointers from each face to its bounding edges and similarly each edge points back onto the two faces forming it. Another list contains all the vertices and each vertex points onto its three edges. Let V be one vertex and L^, L_2 and L3 its edges. There is a vertex centered coordinate system whose axes are L^, L_2 and the normal to the face bounded by these two edges. The relation between this frame and an object centered coordinate system is completely defined by the geometry of the object and it can be expressed by a 4x4 homogeneous transform matrix, A_m. This transform embeds three rotations and three translations that allow to overlap one frame onto the other.

Suppose now that we know the object assignment of an image junction. That is, there is a unique correspondence between the junction's lines and the edges of the vertex. Since the backprojection of the junction constraints the orientations of the face S formed by L_1 and L_2 to just one direction, we can use equations (2) to determine the vectors L_1 and L_2. This will determine the rotation part of a matrix A_c that maps the vertex centered frame into the camera centered frame. The position of the junction in the image determines two translations. In conclusion, under a junction-to-vertex assignment five degrees of freedom are determined. Depth can be computed by triangulation if there is another junction or angle to which a vertex can be assigned. From A_m and A_c one can compute the object-to-camera transform, A :

$$A = A_c \times A_m^{-1} \qquad (4)$$

The actual correspondence between the model and an image feature set is performed by a hypothesize-and-test procedure. A search space is built where each node represents a junction-to-vertex assignment. The goal is to find the largest set of nodes that are mutually compatible, i.e., they uniquely define the six degrees of freedom of the part. An object orientation and location is hypothesized from one node (excluding the depth for which initial lower and upper bounds are given). From this assignment a set of visible vertices is computed and for each such vertex its image projection is determined. This could be a junction, if two or three faces are visible or an angle if only one face is visible. For each prediction, the best image feature match is selected. Notice, however that the low level segmentation process is not perfect and the data are noisy. For these reasons some lines may be missing. If the verification step fails in finding a predicted junction or angle it checks for partial descriptions of these items in the line list. For each assignment a score is computed by calculating the percentage of object features

predicted visible that actually overlap image features. If this score is high enough, the location of the image features as well as their spatial orientation constraints are used for refining the object locational parameters and for estimating tighter bounds for the depth. If the score is too low, the algorithm backtracks to the last choice point.

4. EXPERIMENTAL RESULTS

Figure '? shows a digitized picture which we have used for verifying the effectiveness of the method. The picture is taken with a TV camera through a 25 mmlens. The orientation of the camera relatively to the table top is not known. The object model comprises 24 right vertices (16 are convex and 8 are concave) but two are sufficient to uniquely identify the object. The image segmentation process comprises edge detection (zero-crossings of the convolution of the image with the difference-of-gaussian operator), edge linking (formation of edge chains) and approximation of these chains with straight lines (piecewise polygonal approximation using a split-arid-merge control structure). Short lines are interpreted as noisy data and are thrown out. Within a chain an angle is formed by two adjacent lines. For each angle we seek a third line which, if combined with the angle's lines could form a junction. Figure 3 shows the image junctions extracted by this segmentation process. Similarly there are angle and line lists. Figure 4 shows the orientation constraint for the junction indicated by an arrow. The final recognition result is shown on Figure 5 which displays wireframe projections of the object model with partial hidden line elimination. In [8] we have repeated this experiment with a 90 mm lens (where the perspective distorsion is low) and we have obtained similar results.

5. DISCUSSION

We have discussed a method for matching 3D object models with intensity images. To increase the efficiency of the method, i.e., to reduce the explosion of the search space, we have derived three-space constraints from image features using the mathematics of perspective and knowledge about the object to be located. The method is limited to a class of objects containing vertices formed by intersections of planar faces. These vertices form, by projection angles and junctions. The backprojection of junctions provides a powerful constraint that is valid, unlike the backprojection of polygonal shapes (as is done in [5]), even in the absence of strong perspective distorsion. However, the method will fail in finding an object if no junction has been detected in the image for this object.

Although this technique looks attractive, it's generalisation is not straightforward. In order to deal with a wide range of realistic situations such as missing and imperfect data, complex objects and various ligthing conditions, this method should be combined with other techniques (stereo, motion, shading) and with other sources of information (range and tactile data).

In the future we plan to increase the set of features to include such things as ellipses and to derive three-space constraints from an extended catalogue of feature clusters such as combinations of ellipses and lines. We also plan to implement a program that will automatically derive perception-oriented object descriptions from a CAD-like database.

REFERENCES

[l] Kanade, T. "Recovery of the Three-Dimensional Shape of an Ob jet from a Single View", Artificial Intelligence, vol. 17, n° 1-3, August 1981, pp. 409-460.

[2] Brady, M. and Yuille, A., "An Extremum Principle for Shape from Contour", I.E.E.E. Trans, on Patt. An. and Mach. Int., vol. PAM1-6, n° 3, May 1984, pp. 288-301.

[3] Chakravarty, J., "The Use of Characteristic Views as a basis for Recognition of Three-Dimensional Objects", PhD Dissertation, Image Processing Laboratory, Rensselaer Polytechnic Institue, Troy, New-York, October 198?.

[4] Stockman, G. and Esteva, J.C., "Use of Geometrical Constraints and Clustering to Determine 3D Object Pose", Technical Report TR84-00? Dept. Computer Science, Michigan State University, East Lausing, Michigan 48824, 1984.

[5] Barnard, S. "Interpreting Perspective Images" Artificial Intelligence, vol. 21-1983, pp. 43b-462.

16 | Goad, C. "Special Purpose Automatic Programming for 3D Model Based Vision", Proceedings Image Understanding Workshop, Arlington, Virginia, June 1983, pp. 94-104.

[7 J Bolles, R.C., Horaud, R., Hannah, M.J. "3DP0: A Three-Dimensional Part Orientation System", Proceedings 8th IJCAI, Karlsruhe, Germany, August 1983.

|8I Horaud R., "From Images to Spatial Perception", Proceedings Cognitiva, Paris, France, June 1985.

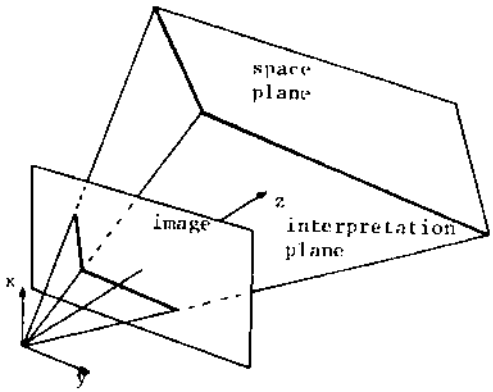
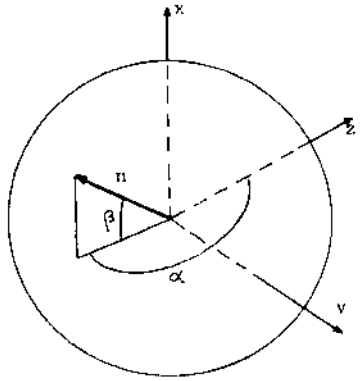


FIGURE 1: a) The geometry of perspective. An image angle backprojects onto a space plane (S).



b) A unit space vector may be represented as a point on the gaussian sphere.

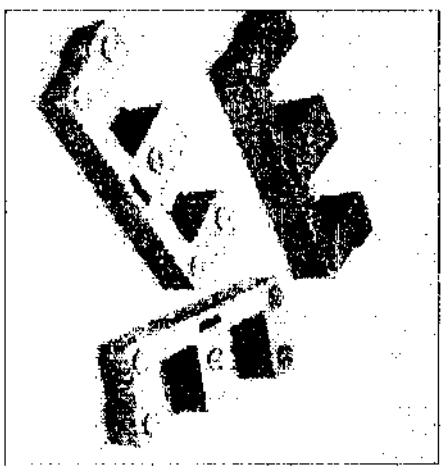


FIGURE 2: A 256x256 image with 64 grey levels taken through a 25 mm lens.

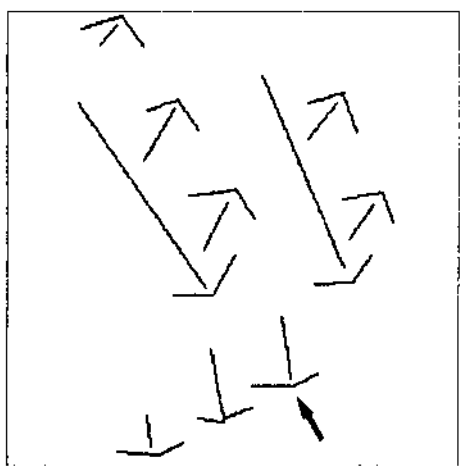


FIGURE 3: Image junctions detected by the data-driven segmentation process. The arrow indicates the junction processed on the next Figure.

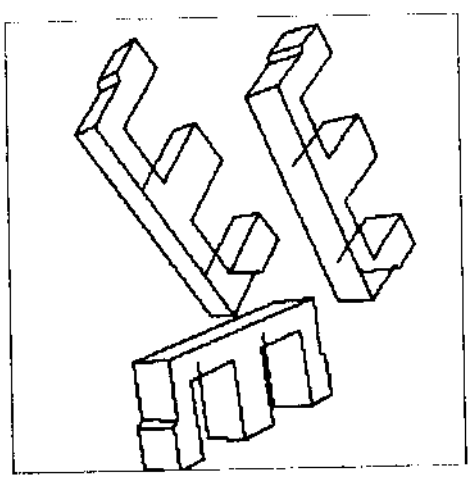


FIGURE 5: The result of recognition. The six locational parameters of each part are determined in camera coordinates.

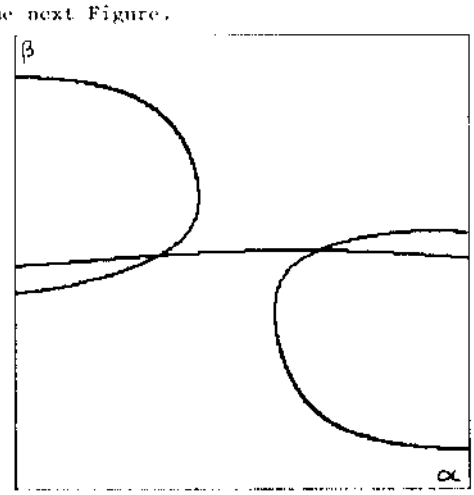


FIGURE 4: Spatial orientation constraints represented on the gaussian sphere. The two solutions correspond to a right vertex interpretation of an image junction.