MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo 1630 March 12 1998

Recovery of Three-Dimensional Objects from Single Perspective Images

Thomas Marill

Abstract

Any three-dimensional wire-frame object constructed out of parallelograms can be recovered from a single perspective two-dimensional image. A procedure for performing the recovery is given.

1 Introduction.

We deal with the problem of recovering three-dimensional objects from single two-dimensional images. Up to now, a complete solution to this problem appears to have been found for only two special cases: rectangular parallelepipeds and twice bilaterally symmetric objects (see Section 5.1). Both of these solutions are based on orthographic projection.

In the present paper we deal with perspective projection. In Section 2, we show how, given the perspective image of a parallelogram in space, we can recover the parallelogram. (Our use of the term "parallelogram" is to be understood as including both rectangular and skewed parallelograms.) In Section 3 we present a technique for recovering members of a certain class of three-dimensional objects. Examples of this class are shown in Figure 1. The techniques discussed require knowing the focal length of the camera that generates the images, and in Section 4 we discuss how one might determine the focal length if it is not known. In Section 5 we discuss these results and related work.

2 Recovery of a parallelogram in space.

2.1 Introduction

In this section we outline a procedure for recovering a parallelogram in space from its perspective image. (The procedure may be considered a recipe for writing a program that performs the recovery.) The procedure yields a parallelogram which is known completely except for a scale factor. This scale factor is discussed in Section 2.6.

The parallelogram in space is represented by the coordinates of its vertices and a list showing which vertices are connected by lines. The coordinates of the vertices are given with respect to a fixed, left-handed coordinate system.

The image we deal with is a perspective projection onto the image plane $z = f$, where f is the focal length. The center of projection is at the origin. We assume that all four vertices are visible in the image.

We start by discussing the general case. We then treat three special cases that can arise. The mathematics involve nothing more than simple algebra and an elementary knowledge of vanishing points. For the most part, proofs (which are close to trivial in any case) have been omitted, and only the results are given.

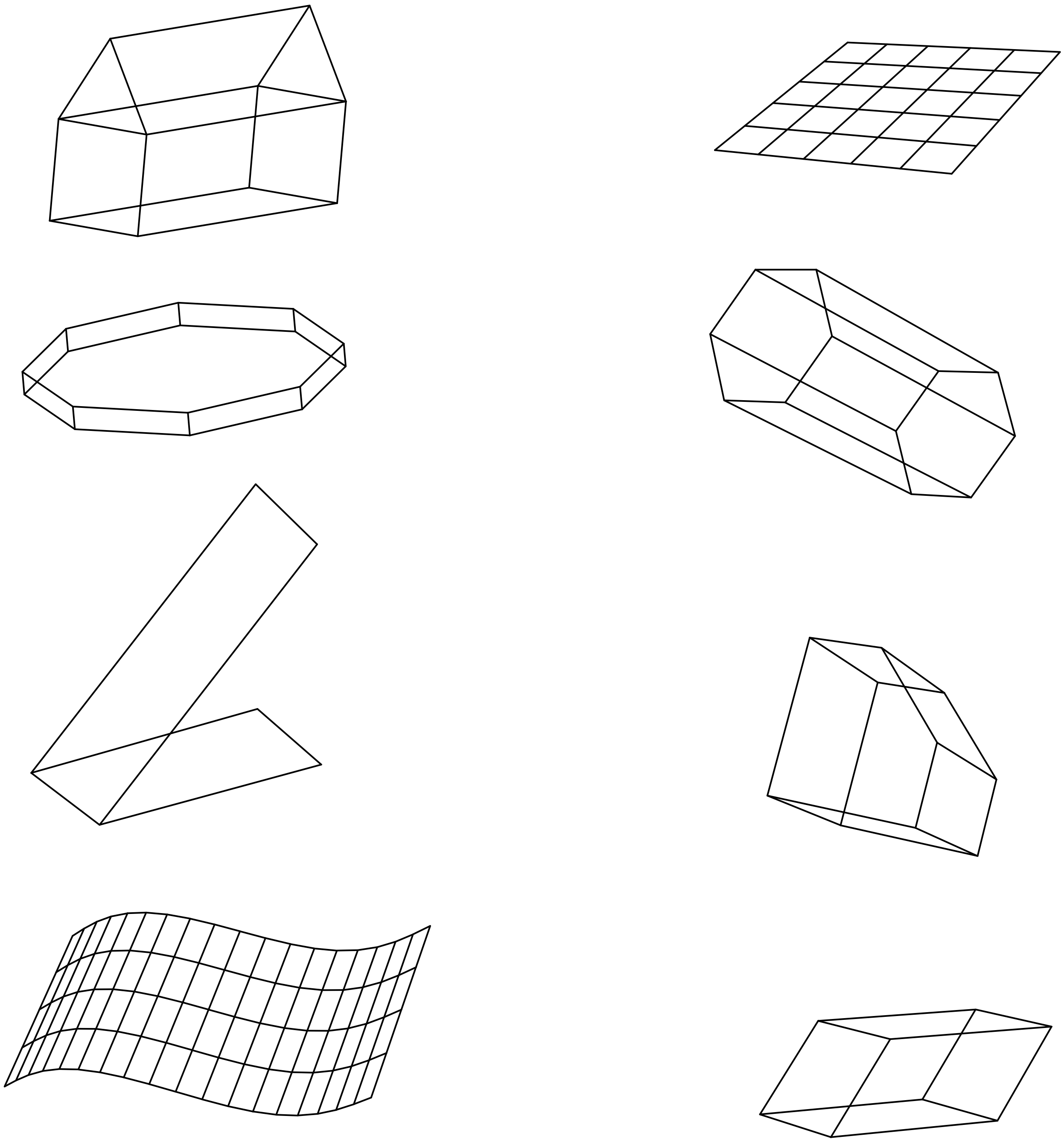

Figure 1: Examples of objects that can be recovered by the present technique. See Section 3.

2.2 Case 1: The general case.

Here we consider the case in which the perspective image consists of an arbitrary quadrilateral with no parallel lines. The recovery procedure for this case can be summarized as follows.

(Step 1) Find the two vanishing points in the image plane.

(Step 2) Find the vanishing line in the image plane and the surface normal vector of the plane in space containing the parallelogram. This vector defines a family of parallel planes.

(Step 3) From the family of planes found in Step 2, select the plane that passes through an arbitrarily selected point $P_0(x_0, y_0, z_0)$.

(Step 4) Find the location of each parallelogram vertex by intersecting the appropriate ray with the plane found in Step 3.

This procedure recovers the parallelogram up to a scale factor. That is, the orientation in space, the angles, and the ratio of the length of adjacent sides will be the same as in the parallelogram that produced the image. The scale of the recovered parallelogram depends on the arbitrary point P_0 of Step 3.

Step 1. Finding the vanishing points.

As is well-known, parallel lines in space have a common vanishing point in the image plane. This point is located at the intersection of the images of the lines.

In the present case, the image consists of a quadrilateral with opposite sides representing parallel lines in space. We find the two vanishing points by locating the intersections of the opposite sides of the quadrilateral (see Figure 2).

Step 2. Finding the vanishing line and the surface normal vector.

Corresponding to every plane in space, there is a "vanishing line" in the image plane. The vanishing line corresponding to a plane has two important properties: first, it uniquely determines the spatial orientation of the plane; second, any line that lies in the plane has its vanishing point on this vanishing line.

In view of the second property, it is easy to find the vanishing line of the plane of the parallelogram. This line is simply the line through the two vanishing points found in Step 1.

In view of the first property, having found the vanishing line, we can immediately determine the orientation of the plane of the parallelogram. The formulas required for doing this are as follows.

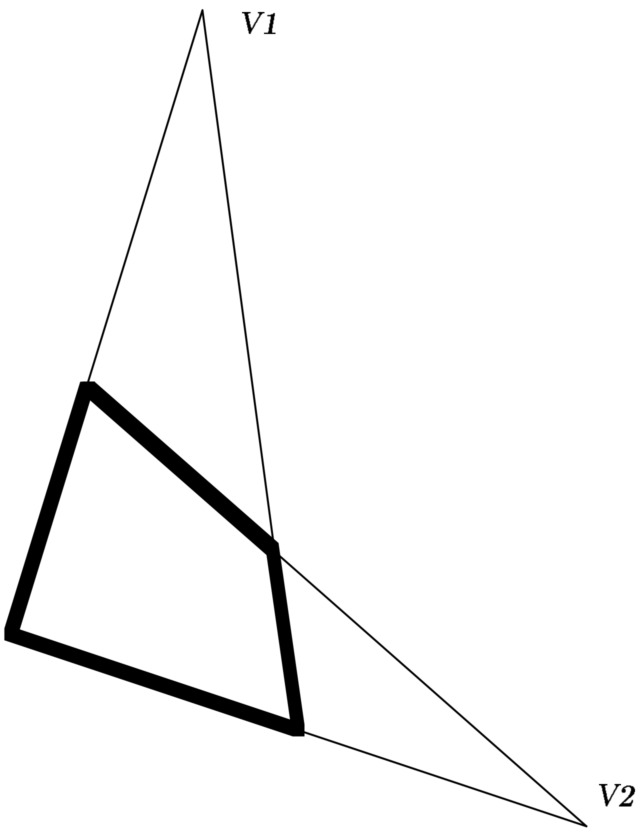

Figure 2: Finding the two vanishing points from the perspective image of a parallelogram.

We write the equation of the vanishing line in the form

(1) $Ax + By + Cf = 0$

where f is the focal length.

The surface normal vector of the plane of the parallelogram is then given by

(2) $\mathbf{N} = A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$

Thus, the surface normal can be read off directly from the equation of the vanishing line.

Step 3. Select a particular plane.

The surface normal vector found in the preceding step defines the orientation in space of a family of parallel planes. From this family, we specify a particular plane by choosing a point $P_0(x_0, y_0, z_0)$ on the desired plane. This point can be chosen arbitrarily.

Step 4. Determine the position in space of the vertices.

Given the plane selected in Step 3, our last job is to determine the location in that plane of the four vertices of the parallelogram. For each vertex V in the image, we define a ray passing through V and through the origin. The intersection of this ray with the plane just selected gives us the position of the vertex of the parallelogram in space (as a function of P_0).

The equation of the three-dimensional ray through the origin and the image point (u, v, f) can be written parametrically:

(3) $x = tu, \quad y = tv, \quad z = tf$

We want the point where this ray intersects a plane that goes through $P_0(x_0, y_0, z_0)$ and has surface normal given by (2). This intersection point is given by (3), with

$$t = \frac{Ax_0 + By_0 + Cz_0}{Au + Bv + Cf}$$

This determines the position of the first vertex in the parallelogram being recovered. The other three vertices are determined the same way. The location of these vertices is expressed as a function of P_0, which affects the overall scale (see Section 2.6).

2.3 Case 2: Two non-vertical parallel lines in the image.

A special case arises when two opposite sides of the quadrilateral that forms the image are parallel. Under these conditions, the two parallel lines do not have a (finite) vanishing point, and the procedure of Section 2.2 must be modified slightly. We assume, in Case 2, that the parallel lines are not vertical.

Given such a configuration, we can determine a single vanishing point by intersecting the two non-parallel image lines. The vanishing line is then simply a line that goes through this vanishing point and is parallel to the two parallel lines in the image. If the vanishing point is (x_1, y_1), and the slope of the parallel lines is s, the vanishing line is equation (1) with

$A = s,\ B = -1,\ C = y_1 - sx_1.$

Once the vanishing line has been determined, the procedure in Section 2.2 is followed.

2.4 Case 3: The case of two vertical parallel lines in the image.

If there are two vertical parallel lines in the image, we can determine a single vanishing point as in the preceding case. The vanishing line is vertical and goes through this vanishing point (x_1, y_1). In this case, the vanishing line is equation (1) with

$A = 1,\ B = 0,\ C = -x_1.$

Once the vanishing line has been determined, the procedure in Section 2.1 is followed.

2.5 Case 4: Two sets of parallel lines in the image.

If the image contains two sets of parallel lines, it is itself a parallelogram. Since this case arises only when the object that gives rise to the image is in a plane parallel to the image plane, the present case is trivial: one simply takes the coordinates of the image vertices as the coordinates of the object vertices and scales them appropriately so that the plane passes through $P_0(x_0, y_0, z_0)$.

If the image vertices are (x_i, y_i), we can write the object coordinates as

$((z_0/f)\, x_i,\ (zo/f)\, y_i,\ z_0\,),\quad i = 1...4$

2.6 The scale factor.

The recovered parallelogram lies in a plane through the arbitrary point P_0. By changing P_0, we obtain solutions in other planes; but all such planes are parallel and have the correct orientation. The result of changing the value of P_0 is then simply to change the scale of the recovered parallelogram: the angles, the ratio of adjacent line-lengths, and the orientation remain unchanged. Thus we say that we have recovered the parallelogram "up to a scale factor."

3 The recovery of a class of three-dimensional objects.

We now deal with a procedure for recovering a class of three-dimensional objects from single perspective images of these objects. The objects in this class are constructed out of parallelograms. Put more precisely, we say that the objects in this class consist of wire-frames such that (a) all quadrilaterals are parallelograms and (b) every vertex falls on some quadrilateral. Pictures of such objects are seen in Figure 1.

The recovery procedure is as follows:

1. Pick any quadrilateral in the image and arbitrarily select a three-dimensional point P_0. Apply the procedure of Section 2 to this quadrilateral, thereby recovering the corresponding parallelogram of the object.

2. Pick any quadrilateral in the image for which a parallelogram has not yet been recovered, provided that this quadrilateral has a vertex in common with one that has been recovered. This vertex has already been assigned three-dimensional coordinates. Let these coordinates become the value of P_0 for the current step. Recover the parallelogram for the current quadrilateral.

3. Until every vertex in the image is assigned to a vertex in the recovered object, repeat step 2.

Given a perspective image of an object in our class, this three-step procedure recovers the object up to a scale factor. That is, its three-dimensional structure and its orientation in space will be the same as in the original object.

4 The focal length.

We have assumed in equation (1) the that focal length f is known. If the focal length is not known, it can be determined from the perspective image of a rectangle.

Let (x_1, y_1) and (x_2, y_2) be the two vanishing points obtained from the perspective image of a rectangle in space. (For example, in Figure 2 assume that the image is the projection of a rectangle.)

The focal length of the camera that generated the image is then given by

(4) $f = \sqrt{-x_1x_2 - y_1y_2}$

5 Discussion.

5.1 Recovery from a single image.

The problem of recovery has been discussed in the literature under the headings of "qualitative" and "quantitative" recovery. The goal of qualitative recovery is to assign certain line-labels to the lines in a drawing (Huffman 1971, Clowes 1971); these labels help us to distinguish possible from impossible scenes. The goal of quantitative recovery is to determine, in numerical terms, the shape and position of a three-dimensional object from its image. In the present paper our interest lies solely in quantitative recovery.

The problem of quantitative recovery from line-drawings has been extensively studied (Kanade 1981, Sugihara 1986), but it would appear that up to now a completely satisfactory approach to recovery from a single image has been achieved in only two cases, both of which assume orthographic projection.

The first is the case of a rectangular parallelepiped. It has been shown (Horn 1972, Kanatani 1986) that one can determine the dimensions and pose of a rectangular parallelepiped from its orthographic image, up to a translation in distance.

More recently, Poggio and Vetter (1992) presented a solution for the case of the orthographic image of a "twice bilaterally symmetric object" (an object with two planes of symmetry). Given the image of such an object, the structure can be determined up to a reflection in the image plane.

5.2 Recovery requires assumptions.

In order for a system to make use of recovery procedures such as the ones discussed, the system must be willing to make certain assumptions.

Thus, the procedure for recovering a rectangular parallelepiped starts with an orthographic image (a line-drawing) of a "brick". In order for the procedure to work, it is necessary for the system to assume (or know) that the image *is* the orthographic image of a brick. Without this assumption, the image could be the image of an infinite number of three-dimensional objects.

Similarly, for the technique of Poggio and Vetter to work, the system must assume (or know) that the image is the orthographic projection of a twice bilaterally symmetric object. Without this, the object could be any one of an infinite number of three-dimensional objects.

The same is true for the technique of the present paper. In order for the technique to work for an image of an object in the class of objects defined in Section 3, the system must assume or know that the image is the image of an object in this class. With this assumption, the system can determine the three-dimensional structure and orientation of the object. Without this assumption, the image could represent any of an infinite number of objects.

5.3 Gradient space.

The idea of the gradient space was introduced by Mackworth (1973) as a technique for representing the orientation of planes and lines space. If the surface normal of a plane is given by

$\mathbf{N} = A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$

then the gradient can be written as

$(A/C)\mathbf{i} + (B/C)\mathbf{j}$,

or more simply as $(A/C,\ B/C)$. This gradient can be considered to be a point in a plane, the "gradient space."

If we have two planes in three-dimensional space, we will have two points in gradient space. Mackworth shows that the line joining these two points has the property that, if overlaid on the image plane, it is perpendicular to the orthographic image of the line formed by the intersection of the two planes. This is the basic gradient-space configuration.

The gradient-space technique was used by Mackworth as an aid in determining whether a line-drawing was the image of a polyhedral scene.

5.4 Vanishing space.

In the present work we have implicitly used a certain form of representation to which one might give the name "vanishing space". Vanishing space is applicable to perspective projection, and is in some ways analogous to gradient space, which is applicable to orthography (see also Shafer, Kanade, and Kender 1983).

In vanishing space we deal with vanishing points and vanishing lines, both of which lie in the image plane. (There is no need of "overlay" as in gradient space.) The rules of vanishing space are as follows: The orientation of a line in space parallel of $P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ is uniquely represented by the vanishing point $(Pf/R,\ Qf/R)$ of its perspective image, where f is the focal length. (Note the resemblance to the representation of the orientation of a plane in gradient space.) The orientation of a plane in space whose surface normal vector is $A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$ is uniquely represented by its vanishing line

$$Ax + By + Cf = 0$$

where f, as before, is the focal length.

A fundamental relation between vanishing points and vanishing lines is the following: Given a plane in space that contains some lines, the vanishing points of these lines lie on the vanishing line of the plane. Analogously to the basic gradient-space configuration, we can define the basic vanishing-space configuration: the vanishing point of the image of a line L formed by the intersection of two planes lies at the intersection of the vanishing lines of the two planes; and the perspective image of L goes through this intersection.

It is these concepts that underlie much of the present work.

5.5 Skewed symmetry.

A different approach to recovering three-dimensional objects from single images was studied by Kanade (1981), who proposed novel techniques based on the concept of skewed symmetry. Kanade hypothesized that a skewed symmetry in the image is the orthographic projection of a real symmetry in the object, and he shows how the real symmetry can be recovered. However, solutions obtained in this way are not unique. In order to obtain unique solutions, Kanade proposed a heuristic principle: of the infinite number of real symmetries that project to a skewed symmetry, select the one that has the least slant. This heuristic, however, fails in certain cases.

5.6 Human vision system.

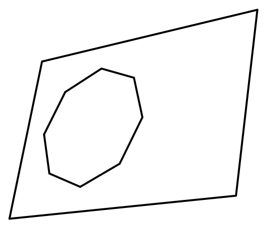

Figure 3: A rectangle “decorated” with an octagon.

We have discussed the recovery problem: given the image of an object, determine the structure and orientation of that object. This is a problem of geometry. There has also been discussion in the literature of the problem of human interpretation: given an image, find an interpretation that matches the one found by the human vision system (Marill 1991, 1995). This is a problem of psychology. These two problems are different and can generate different solutions. Indeed, some authors have argued that the two bear little relation to one another (Brooks 1991).

5.7 Future work.

We saw in Section 5.2 that assumptions were required for recovery, and we have discussed the ones required for the present technique. If we are willing to make additional assumptions, we can broaden the class of objects that can be recovered.

For example, we can take a parallelogram and “decorate” its interior with additional lines (Figure 3). The three-dimensional position of all vertices in the image can be recovered from the perspective image if we make the assumptions (a) that the outside quadrilateral represents a parallelogram and (b) that the vertices in the interior are coplanar with this parallelogram. The techniques we have given in the present paper are sufficient for obtaining a solution.

References.

R.A. Brooks (1991). Intelligence without reason. *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence* (IJCAI-91). Morgan Kaufmann Publishers, San Mateo, Calif.

M.B. Clowes (1971). On seeing things. *Artificial Intelligence*, 79-116, vol. 2.

D.A. Huffman (1971). Impossible objects as nonsense sentences, in *Machine Intelligence,* 295-322, vol. 6, B. Meltzer and D. Michie, Eds. Edinburgh University Press.

B.K.P. Horn (1972) VISMEM: A bag of robotics formulae. Working Paper 34, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

T. Kanade (1981) Recovery of the three-dimensional shape of an object from a single view. *Artificial Intelligence*, 17, 409-460.

K-I. Kanatani (1986). The constraints on images of rectangular polyhedra. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No. 4.

A.K. Mackworth (1973). Interpreting pictures of polyhedral scenes. *Artificial Intelligence,* 4, 121-137.

T. Marill (1991). Emulating the human interpretation of line-drawings as three-dimensional objects. *International Journal of Computer Vision* Vol. 6 No. 2, 147-161.

T. Marill (1995). The three-dimensional interpretation of a class of simple line-drawings. AI Memo 1555. Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

T. Poggio and T. Vetter (1992) Recognition and structure from one 2D model view: Observations on prototypes, object classes and symmetries. AI Memo 1347, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

S.A. Shafer, T. Kanade, and J. Kender (1983). Gradient space under orthography and perspective. *Computer Vision, Graphics, and Image Processing*, Vol. 24, no. 2, 182-199.

K. Sugihara (1986). Machine Interpretation of Line Drawings. Cambridge, Mass.: MIT Press.