MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A Counterexample to the Theory that Vision Recovers Three-Dimensional Scenes

Thomas Marill

Abstract

The problem of three-dimensional vision is generally formulated as the problem of recovering the three-dimensional scene that caused the image. Here we present a certain line-drawing and show that it has the following property: the three-dimensional object we see when we look at this line-drawing does not have the line-drawing as its image. It would therefore be impossible for the seen object to be the cause of the image. Such an occurrence constitutes a counterexample to the theory that vision recovers the scene that caused the image.

1. Introduction: The concept of recovery.

The problem of three-dimensional vision is generally formulated as the problem of recovering the three-dimensional scene that caused the image. Thus Charniak & McDermott [1] tell us that "Unlike many problems in AI, the vision problem may be stated with reasonable precision: given a two-dimensional image, infer the objects that produced it."

The same idea is sometimes expressed in terms of "inverse optics." For example, Bulthoff & Mallot [2] write that "The problem of deriving a description of a three-dimensional scene from its two-dimensional image on the retina is the inverse of classical optics, wherein one has to find the two-dimensional image (brightness distribution) of a three-dimensional object."

This formulation is not usually considered to be a theory about vision, but rather a self-evident truth not open to question. In the present paper, however, we present an example that violates this formulation.

2. The line-drawing.

Let us start out by looking at Fig. 1. What we see is a wire-frame cube. Fig. 1 is a kind of prototype of cube-pictures; it is what many people will draw when asked to draw a cube.

We now make the viewing conditions more precise. We place the eye (more precisely, the optical center of the eye) at the origin of a left-handed coordinate system with the line of sight along the positive z-axis. And we place Fig. 1 in the image plane $z = f$ so that point D falls at (0,0) in the image plane and so that the top of the figure is parallel to the x-axis. Finally, for the sake of specificity and to have a comfortable viewing distance, we set $f = 24$ inches.

We can then say that when we look at Fig. 1 under these viewing conditions, we see a three-dimensional cube having a face parallel to the image plane, with the edges of this face aligned with the x- and y-axes, and having one corner on the z-axis. Let's refer to such a cube as being in *standard position*. In short, when we look at Fig. 1 under our viewing conditions we see a cube in standard position.

It should also be pointed out that once we have specified Fig. 1 and the viewing conditions, we have uniquely defined the retinal image. Let us call this retinal image R. We can therefore also say that when retinal image R obtains, what is seen is a cube in standard position.

3. The mystery.

The surprising aspect of all this is that there is no cube in standard position which perspectively projects into Fig. 1. The three-dimensional object we see when we look at Fig. 1 is not an object that has Fig. 1 as its image. Or, expressed in terms of the retinal image, what we see when retinal image R obtains is something that could not have caused retinal image R.

How can we be sure that there cannot be a cube in standard position that projects to Fig. 1? The following theorem helps.

Theorem. Let the center of projection be at the origin of a left-handed coordinate system and the image plane be $z = f$. Let L be a line in space parallel to the vector $\mathbf{v} = A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$ $(C \neq 0)$. Then the perspective image of L is a half-line in the image plane terminating at the "vanishing point" $(Af/C, Bf/C)$.

Returning our three-dimensional cube, let us label the vertices D', E', ... , K' (corresponding to the image points D, E, ..., K). Assume that the cube is in standard position. Then the line-segments D'H', E'I', F'J', and G'K' are all parallel to the vector $\mathbf{v} = 0\mathbf{i} + 0\mathbf{j} + 1\mathbf{k} = \mathbf{k}$. By the theorem, the images (DH, EI, FJ, and GK) of these line-segments all lie on lines having the same vanishing point, namely (0,0), in the image plane. That is, these image line-segments all lie on lines that intersect at the point D. But clearly, this is not the case. Hence Fig. 1 cannot be the image of a cube in standard position.

To reinforce this point, consider line FJ. If Fig. 1 is the image of a cube in standard position, FJ must lie on a line through point D. But the line FJ does not point toward D; in fact, it points in the wrong direction. Hence not only is Fig. 1 not the image of a cube in standard position, but it is not even close to such an image.

It may be of interest to consider the hypothesis that Fig. 1 is the image of a cube which is not in standard position. For that hypothesis to be true, one of two things would have to hold:

Either (a) Fig. 1 is the image of a cube none of whose faces are parallel to the image plane and aligned with the x- and y-axes;

Or (b) Fig. 1 is the image of a cube none of whose vertices are on the z-axis.

As regards condition (a), a little reflection shows that if the face of a cube is not parallel to the image plane, its image cannot be a square. (Because if a face is not parallel to the image plane, at least three of its edges must not be parallel to

the image plane; therefore at least two of its parallel (and non-adjacent) edges must not be parallel to the image plane. But then, by the above theorem, the image of at least two of its non-adjacent edges must lie on interesecting lines and therefore cannot form the edges of a square.) In Fig. 1, however, DEFG is the image of a face and it is a square. Hence (a) is false.

As regards condition (b) the only way a point in space can have its image on the z-axis is for it to be on the z-axis. If none of the corners of the cube were on the z-axis, no point in the image could be at (0,0) in the image plane. But point D is at (0,0). Hence (b) is false.

Thus we see that Fig. 1 cannot be the image of a cube in non-standard position. Hence Fig. 1 is not the image of a cube at all.

4. Polyhedral scenes.

We now wish to relate our discussion to the theory of line-drawings of polyhedral scenes.

In this domain the concept of recovery can be easily be stated with precision. Thus Horn [3], in a section dealing with recovering three-dimensional structures, defines the problem as follows. "Given a line drawing of a polyhedron, we might be asked to determine its three-dimensional structure... . Each vertex lies on a ray defined by the corresponding junction in the line drawing. We do not know how far along this ray the vertex lies, however. Determining the arrangement of vertices in three dimensions, then, implies finding the distances to each of the vertices."

In our example, we see a polyhedral object when we look at a line drawing. But we have shown that the object does not have the drawing as its image. What is happening in our case, therefore, is quite simply that the vertices of the object "recovered" by the visual system do not lie on the rays at all. (If they did, the image would be identical to the drawing.)

5. Conclusions.

We have discussed the concept of recovery. In the vision literature this concept is usually considered to be axiomatic. However, we have given an example which, at a minimum, should cause us to consider the concept, not as an axiom, but as a theory subject to empirical verification or falsification. And if the example stands

up under further examination, it would have to be considered as evidence that the theory is false.

Acknowledgments.

I am grateful to Rodney Brooks and Boris Katz for their many helpful comments.

References.

1. E. Charniak & D. McDermott, *Introduction to Artificial Intelligence*, Reading, MA: Addison-Wesley (1985).

2. H. H. Bulthoff & H. A. Mallot, " Interaction of different modules in depth perception," Proc. First International Conf. on Computer Vision, 295-305, IEEE Computer Society Press (1987).

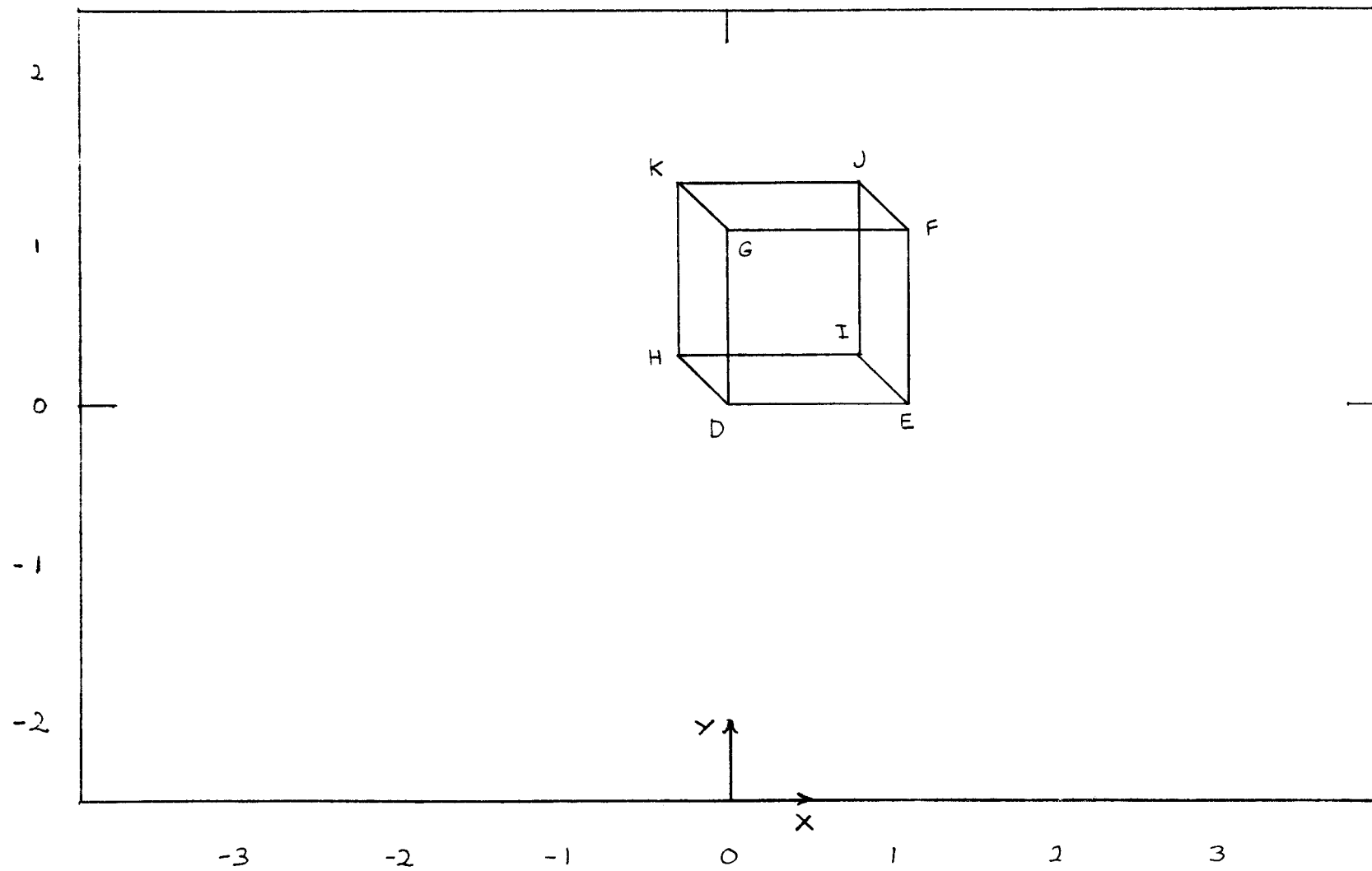3. B. K. P. Horn, *Robot Vision*, Cambridge, MA: MIT Press (1986).

FIGURE 1