USING MODELS TO SEE

Alan K. Mackworth
University of Sussex*

Abstract

Scene analysis programs offer the hope of providing a more adequate account of human competence in interpreting line drawings as polyhedra than do the current psychological theories. This thesis has several aspects. The aspect concentrated on here is that those programs have explored a variety of methods of incorporating a priori knowledge of objects through the use of models. After outlining the range of models used and sketching some psychological theories, the various proposals are contrasted. This discussion leads to two new proposals for exploiting model information that involve elaborations of an existing program, POLY.

1. Introduction.

In one of its many roles, artificial intelligence is cast as the vanguard of an army of psychologists who seek a new paradigm for cognitive and perceptual processes. Despite several clarion calls to this effect (Minsky and Papert, 1972; Clowes, 1972; Sutherland, 1973) AI may well be a vanguard without an army. This paper attempts to show that a small part of the scouted territory is ripe for capture.

The interpretation of line drawings as polyhedral scenes has been the focus of most attempts to build AI vision systems. As it is a natural human task, several psychologists have also studied it. In sketching and contrasting various resultant theories, we will concentrate on how they represent the a priori knowledge of the objects that exist in the world. Of necessity, other essential themes such as non-model knowledge of the world (for example, support and the picture-formation process itself) or the use of picture cues to access the models are slighted.

Sections 2 and 3 of the paper sketch the use of models in several AI and human vision proposals. Section 4 briefly contrasts them using a few examples. Some of the weaknesses exposed lead to two proposals in section 5.

2. Models in Machine Vision.

Roberts (1965) used the three simple models of Fig.1. These can be expanded along each of their coordinate axes. Compound objects are created

---

*Now at Department of Computer Science,
University of British Columbia,
Vancouver 8, B.C. Canada.

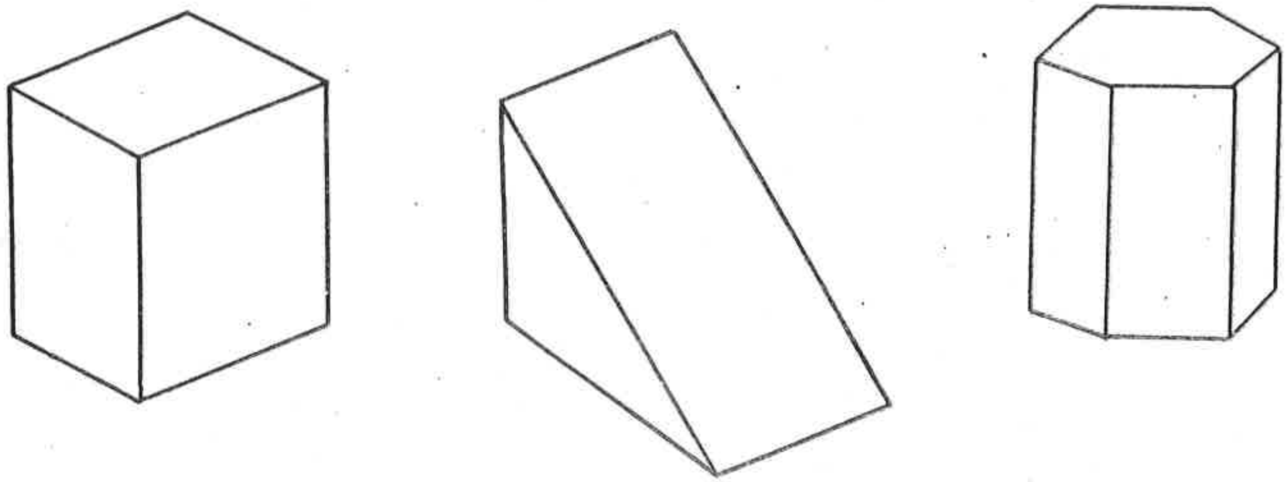


Figure 1. Roberts' simple object models

by abutting simple ones. Falk's (1972) recent state-of-the-art scene analysis system expected its visual world to be composed of instances of the nine polyhedral prototypes of specified dimensions shown in Fig.2.

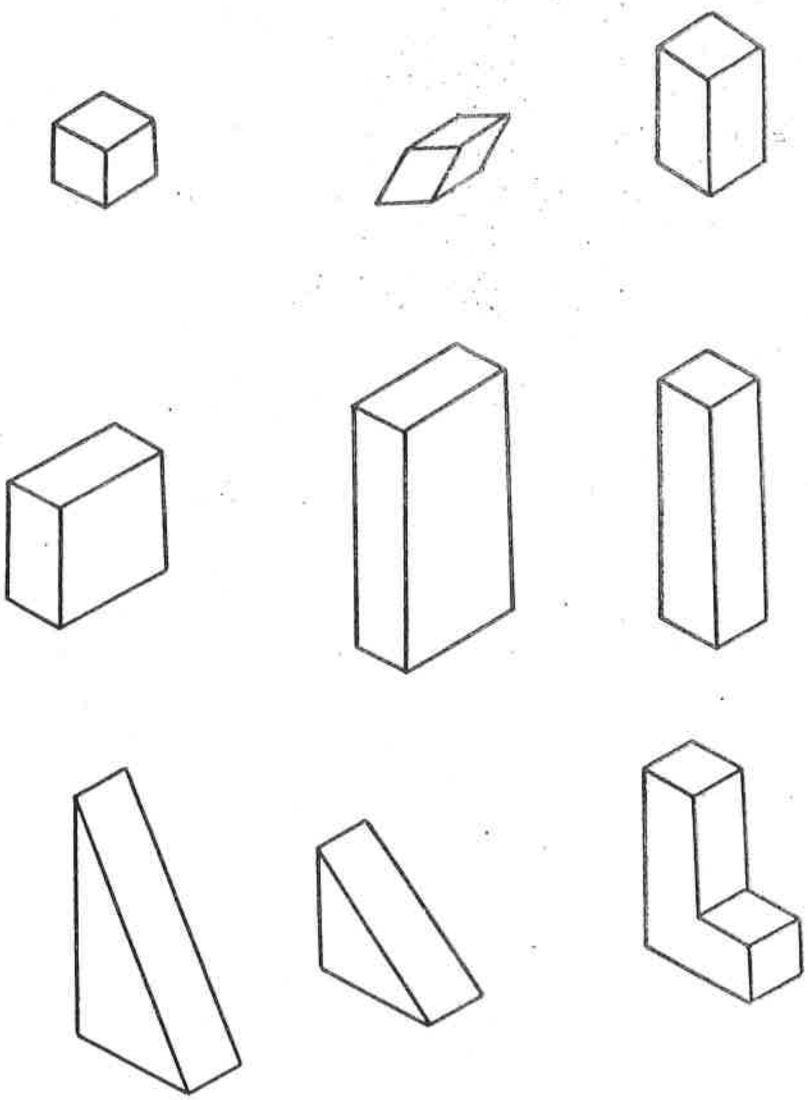


Figure 2. Falk's object prototypes

The size-specificity of the prototypes was exploited by the object recognition phase of the program in its use of the actual heights of the blocks and the lengths of their base edges.

At the other end of the size and shape specificity spectra for models are the edge-labelling procedures. These originated in Guzman's SEE (1968) which produces surface groupings corresponding to objects. Huffman (1971) and Clowes (1971) developed a procedure that relies on four prototype corners: the trihedral corners in which the object occupies 1, 3, 5 or 7 'octants'; the corners have no further shape-specificity. The corner models are accessed by the shape of the picture junctions. For each of four picture junction classes (L, FORK, ARROW, T), there is a list of possible corner/viewpoint configurations. These lists are used to label the edges depicted as convex or concave. The convex category is subdivided into three according to the viewpoint: either both surfaces depicted at the edge belong to it or the surface on the right, which does, is partially occluding the one on the left which doesn't or vice versa.

It has been shown (Mackworth, 1974) that SEE implicitly uses a single prototype corner: the one in which the object occupies only one 'octant'; whereas, Waltz (1972) has expanded the range of corner prototypes far beyond those of Huffman-Clowes.

The model information embedded in POLY (Mackworth, 1973) is minimal, confined as it is to a requirement that surfaces be planar and edges be occluding or connect (non-occluding); however, there is a marked preference for connect edges. With this apparatus somewhat augmented, POLY interacts with a representation called the gradient configuration (originally suggested by Huffman (1971))to produce a labelled interpretation. (The gradient of an edge is a vector in a 2D gradient space whose direction is that of the corresponding picture line. Its length is the tangent of the angle between the edge and the picture plane. The gradient of a surface is in the direction of steepest descent in the surface away from the picture plane; the magnitude of the gradient is the tangent of the dihedral angle between the surface and the picture plane.) The final gradient configuration needs only the origin and scale of the gradient space defined before it represents the absolute orientations of the object surfaces. (POLY assumes orthographic projection; see (Mackworth, 1974) for the perspective case.)

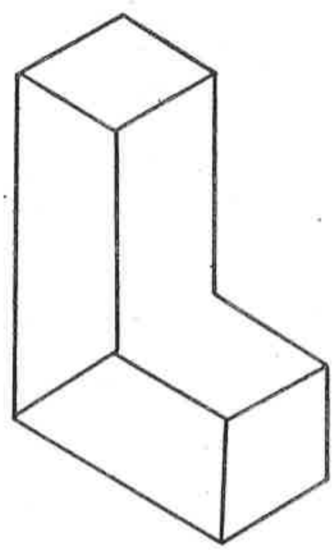
3. Some Psychological Theories.

Attempts to provide psychological theories of the interpretation of line drawings have not usually provided an algorithm by which interpretation may proceed though, presumably, the usual monocular depth cues are thought to be relevant. Rather, such theories seem to assume the existence of such an algorithm and concentrate on the tension set up between the 3D (scene) and 2D (picture) organisations. Kopfermann (1930) held that the impression of tridimensionality varies with the degree to which the scene organization is simpler than that of the picture. In extending that theory Hochberg and Brooks (1960) provided a quantified measure of simplicity as the sum of the number of lines, the number of angles and the number of angles differing in magnitude. Attneave and Frost (1969) presented a similar theory in which the competition between the scene and the picture is resolved by figural simplicity criteria.

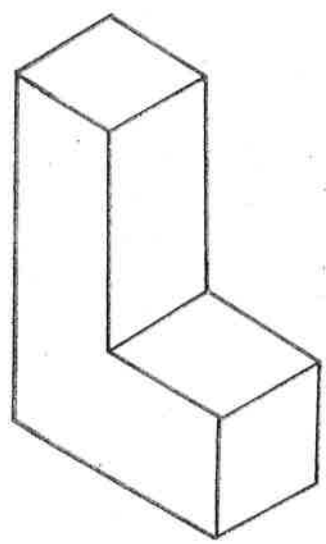
Finally Hochberg (1968) almost anticipated the Huffman-Clowes algorithm as he demonstrated, with an ingenious experiment, that junctions act as 'local depth cues'.

4. Some Examples.

The discussion of this section uses, as examples, the two pictures in Fig. 3, which the reader should look at without reading further. The usual



(a)



(b)

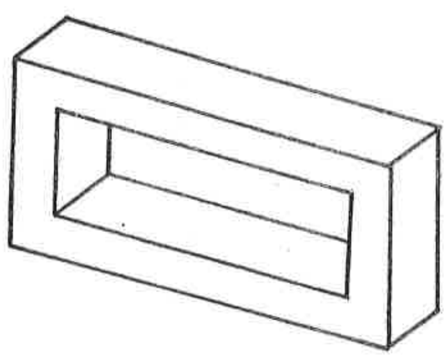
Figure 3. Two examples

impression given by Fig. 3(a) is that it remains obstinately flat on the page (provided one can suppress the tendency to see an edge that is not depicted and to ignore one that is thereby transforming the picture into Fig. 3(b)). Fig. 3(b) has a solid, three-dimensional appearance. This major difference which holds even though both are equally faithful depictions of polyhedra is a phenomenal fact requiring explanation.

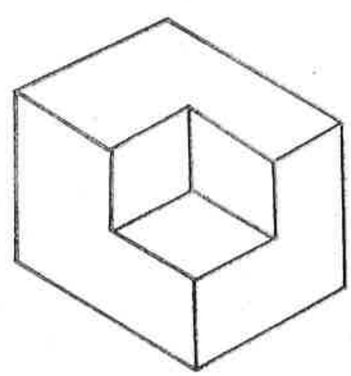
The Huffman-Clowes algorithms do not offer that explanation. The pictures are successfully labelled with equal ease. The corners are all trihedral and both interpretations require three hidden surfaces to complete the object.

Perhaps the scene interpretation of Fig. 3(b) derives from its familiarity as Falk's program suggests: the L-beam is one of its nine prototype objects. Look then at Fig. 5(a) which is surely unfamiliar to the reader (although it is derived from pictures used by Shepard and Metzler (1971)); is that object any less solid than the one depicted in Fig. 3(b)?

Both Fig. 3(b) and Fig. 5(a) are interpreted by Roberts' program as compound objects made from cuboids (two and four, respectively). However, that program cannot, contrary to expectation, similarly interpret the two objects in Fig. 4.



(a)



(b)

Figure 4. Two concave rectangular objects

The traditional depth cue theory of picture interpretation has nothing to say. In four of the pictures (Figs. 3(a), 3(b), 4(b), 5(a)) there are no traditional depth cues at all! Yet, in Fig. 5(a) for example, corner 1 is clearly nearer the observer than corner 2.

The Hochberg-Brooks criterion doesn't contradict the phenomenon. (By that criterion, the pictures have equal complexity while the scene in Fig. 3(b) is just marginally simpler than that in Fig. 3(a)). And yet, that doesn't take us very far. What mechanisms produced those scene interpretations in the first place? Hochberg (1968) has provided an excellent rebuttal of his earlier theory.

On the other hand, Hochberg's claim (1968) that the junction configurations act as 'local depth cues' is not very powerful. The only depth evidence given by edge labelling is provided by the occluding edges; that is the traditional depth cue of partial occlusion. Even if the edges of Fig. 5(a) are appropriately labelled convex, concave or occluding there is still no evidence that corner 1 is closer to the observer than corner 2; that is, an ordinary polyhedron with that appearance and those edge labels can be constructed for which that is not true.

5. Two Proposals.

Surely the solidity of Fig. 3(b) and Fig. 5(a) as contrasted with the 'flat' appearance of Fig. 3(a) can be explained as follows: in the former cases, the polyhedron interpretation can be seen as made up of surfaces of very familiar shape (in this case, rectangular) whereas in the latter that is not possible. This explanation suggests extensions to POLY that use the shapes of surfaces as models. Here two such extensions are presented. Rectangularity is often a major feature of the worlds we build for ourselves. The first proposal shows how a straightforward extension to POLY can exploit that feature. The second proposal is more of a substantial upheaval than an extension in that it suggests integrating the use of prototypes into the interpretation process.

5.1 Rectangularity.

Consider the rectangular object of Fig. 5(a). For this object POLY produces the gradient configuration that appears as a triangle in Fig. 5(b) There are only three possible values for the gradient so several surfaces

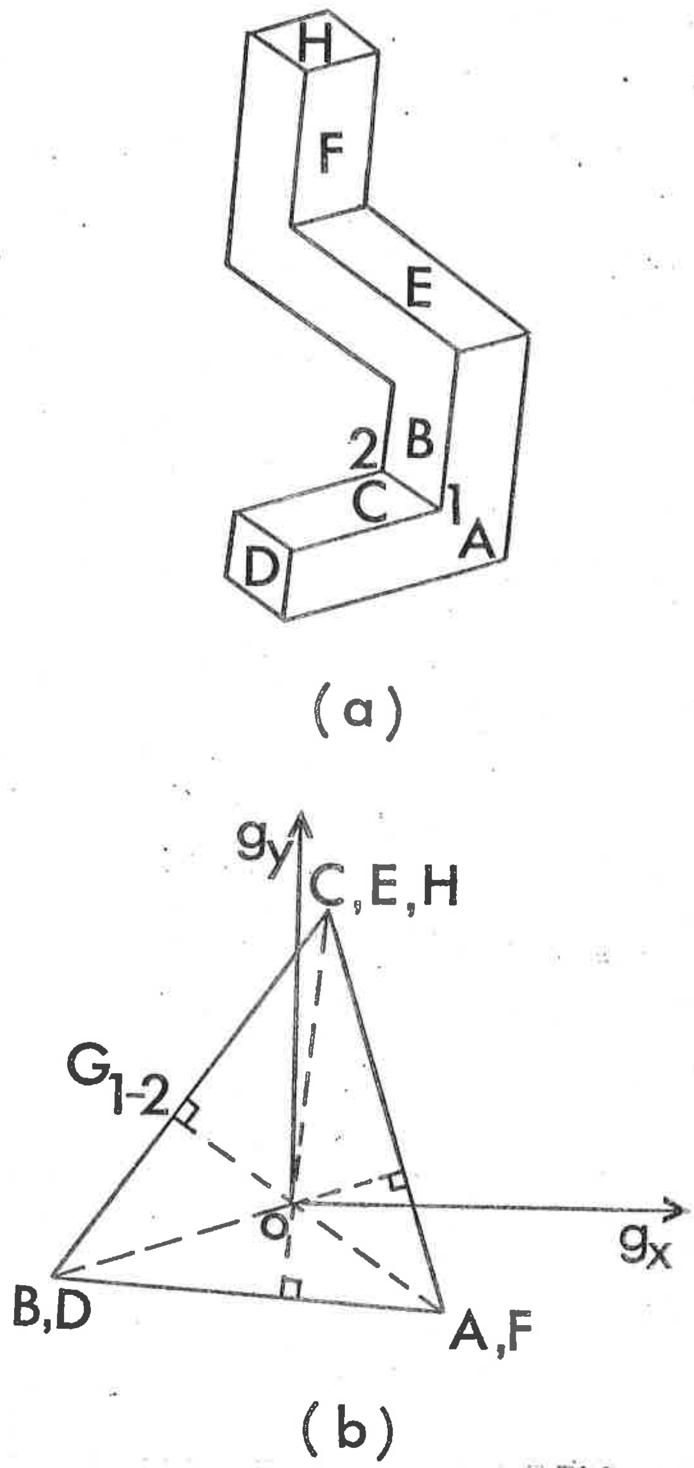


Figure 5. A rectangular object and its gradient space configuration

are superimposed at each position. This obscures the fact that the configuration is intricately connected:. each pair of surfaces meeting in a connect edge is joined by a line perpendicular to the picture line showing that edge. Neither the position of the origin nor the size of the triangle is yet specified but note that E and A are ordered in the gradient configuration just as they are across their common edge in the picture so that edge is convex whereas the relative positions of B and C are reversed in the picture and gradient spaces so that edge 1-2 is concave; however, as the actual values of the gradients are not determined, we still cannot say that corner 1 is closer than 2.

At any corner such as corner 1 in Fig. 5(a) there are three edges (which may not all be visible). Each pair of edges defines a surface at that corner. Each edge is normal to the surface defined by the other two. Since the direction of the gradient vector is the direction in the picture in which the normal appears to point, the direction of the gradient of each surface at the corner is given by the edge that does not belong to it. Thus gradient A must be in the direction of picture line 2-1. Since the vector difference between gradients B and C is required to be perpendicular to picture line 2-1, the origin must be on a perpendicular dropped from gradient A to the opposite side of the gradient triangle. Hence the origin must be at 0 shown in Fig. 5(b). The scale is immediately determined by the requirement that the product of the magnitudes of the gradient of A and the gradient of edge 1-2, G_{1-2}, must be unity. Now that the orientations of all the surfaces and edges are defined it is an obvious consequence that corner 2 is further from the picture plane than corner 1; that is shown by the fact that G_{1-2} points up to the left (not down to the right).

5.2 Using prototype surfaces.

The idea of using specific prototypes is attractive but as suggested in Section 4 complete polyhedral prototypes are, in a sense, too monolithic. In this section we show how the use of prototype surfaces can be integrated directly into the POLY interpretation process.

Consider Falk's list of nine prototype objects. They have in all fifty-four separate faces; yet those faces have only fourteen distinct polygonal shapes. The size-specificity of these shapes will be dropped for the sake of this argument although it could be retained. Dropping size-specificity (so that a 1 x 2 rectangle represents itself and the 2 x 4 rectangle etc.) leaves a total of twelve distinct surface shapes.

First, a geometrical fact must be stated (Mackworth, 1974). Suppose one is given the true shape of a surface in the form of a polygon (where the dimensions may be uniformly scaled up or down by a factor, k), the projected shape of that surface and three or more pairs of non-collinear points on the true and projected shapes that correspond. From this information it is easy to compute whether the true shape could produce the projected shape and, if it does, the value of k and the gradient of the surface.

For each picture region, by considering the topologically identical surfaces, a set of possible surfaces each with a corresponding k and gradient could be computed. If that set is empty then the region depicts a partially occluded surface.

This is now a labelling situation comparable to the corner labelling algorithms of Huffman, Clowes and Waltz. In those algorithms each junction has associated with it a set of possible corners; the aim of the interpretation is to discover a unique corner corresponding to each junction. Here, besides labelling each edge, the aim is to assign a unique surface to each region. Agreement between the interpretations of adjacent regions is necessary if the edge is taken to be connect. The agreement takes two distinct forms. First, the POLY coherence rules must be satisfied and second, model-based coherence rules must be used. Such model-based rules would, at the lowest level, be of the form: Are there two such surfaces meeting at an edge in the set of prototypes? If so, do those surfaces meet at this dihedral angle? Do they agree on the scale factor? Higher levels would also be required: Are there three such surfaces meeting at a corner?

Procedurally, this approach need not be implemented in a depth or breadth-first fashion. It is amenable to the two-stage Waltz search procedure which would first weed out the lists of possible surfaces (just as Waltz weeded the lists of possible corners) based on consideration of the mutual interpretation of each pair of adjacent regions and only then try to build complete coherent interpretations.

6. Conclusion.

World knowledge of the type incorporated as models in scene analysis programs is an essential feature of any psychological theory that attempts to explain human competence in interpreting line drawings as polyhedra. Furthermore, in those programs that knowledge is used in a procedural fashion; they demonstrate, at the very least, how a scene interpretation can be achieved.

The discussion of Section 4 has pointed out some of the ways in which the available range of models is deficient for purposes of psychological explanation. The two proposals of Section 5 are designed to provide mechanisms that reflect particular human competence in this task domain.

Acknowledgements

The author is grateful to Max Clowes, Roddie Cowie, Frank O'Gorman and Aaron Sloman for discussion and criticism.

The research reported is part of a project supported by the Science Research Council.

Bibliography

Attneave, F. and Frost, R. (1969) The determination of perceived tridimensional orientation by minimum criteria. Perception & Psychophysics 6, 391-396.

Clowes, M.B. (1971) On seeing things. Artificial Intelligence 2,1, 79-112.

Clowes, M.B. (1972) Artificial intelligence as psychology. AISB Bulletin, November, 1972.

Falk, G. (1972) Interpretation of imperfect data as a three dimensional scene. Artificial Intelligence 3, 2, 101-144.

Guzman, A. (1968) Decomposition of a visual scene into three-dimensional bodies. AFIPS Proc. Fall Joint Comp. Conf. Vol. 33, pp.291-304.

Hochberg, J. and Brooks, V. (1960) The psychophysics of form: reversible-perspective drawings of spatial objects. Amer. J. Psychol. 73, 337-354.

Hochberg, J. (1968) In the mind's eye. Contemporary Theory and Research in Visual Perception Haber, R.N. (Ed.), Holt, Rinehart and Winston, N.Y., pp. 309-331.

Huffman, D.A. (1971) Impossible objects as nonsense sentences. Machine Intelligence 6 Meltzer, B. and Michie, D. (Eds.), Edinburgh University Press, Edinburgh, pp.295-323.

Kopfermann, H. (1930) Psychologische Untersuchungen uber die Wirkung zweidimensionaler Darstellungenkorperlicher Gebilde. Psychol. Forsch., 13, 293-364.

Mackworth, A.K. (1973) Interpreting pictures of polyhedral scenes. Artificial Intelligence, 4, 2, 121-137.

Mackworth, A.K. (1974) On the interpretation of drawings as three-dimensional scenes. D.Phil. thesis, Laboratory of Experimental Psychology, University of Sussex.

Minsky, M. and Papert, S. (1972) Progress Report Memo. No. 252. Artificial Intelligence Lab., Mass. Inst. of Technol., Cambridge, Mass.

Roberts, L.G. (1965) Machine perception of three-dimensional objects. Optical and Electro-Optical Information Processing Tippett, et al (Eds.), MIT Press, Cambridge, Mass., pp. 159-197.

Bibliography

Shepard, R.N. and Metzler, J. (1971) Mental rotation of three-dimensional objects. Science 171, 701-703.

Sutherland, N.S. (1973) Some comments on the Lighthill report and on Artificial Intelligence. Artificial Intelligence: a paper symposium Science Research Council, London, pp. 22-31.

Waltz, D.L. (1972) Generating semantic descriptions from drawings of scenes with shadows. AI TR-271(Thesis), MIT, Cambridge, Mass.