ICOT Technical Report: TR-396

TR-396

Cognitive Model for Depth Perception from a Single Line Drawing

by

N. Oka

June, 1988

ICOT

Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan

(03) 456-3191~5
Telex ICOT J32964

Institute for New Generation Computer Technology

Cognitive Model for Depth Perception from a Single Line Drawing

Natsuki Oka

Institute for New Generation Computer Technology
4-28, Mita 1-chome, Minato-ku, Tokyo 108, JAPAN
csnet: oka%icot.jp@relay.cs.net
uucp: {enea,inria,kddlab,mit-eddie,ukc}!icot!oka

Abstract

Although the depth of an object is perceived quasi-quantitatively from a single two-dimensional line drawing, previously published interpretation methods of a line drawing do not wholly explain this phenomenon. For example, labeling methods interpret a line drawing only qualitatively, model-based methods are applicable only for predetermined objects, and regularity-based methods are valid only for regularly-shaped objects.

An inclusive cognitive model for depth perception from a single line drawing is proposed. Unlike most other approaches, it can explain the quasi-quantitative interpretation of irregularly-shaped and unfamiliar objects. It interprets a drawing as the one of many geometrically possible objects which is most likely to produce the drawing. The likelihood of producing the drawing depends on both the likelihood of an object (object likelihood) and the likelihood of a view of the object (view likelihood). The new concept, view likelihood, enables the interpretation of irregularly-shaped objects. The likelihood of the projected area is proposed as an approximation to view likelihood, and it enables the interpretation of unfamiliar objects.

This paper also examines line drawings which are drawn to communicate three-dimensional shapes. In this situation, the probability that an interpretation based on the view likelihood is correct increases.

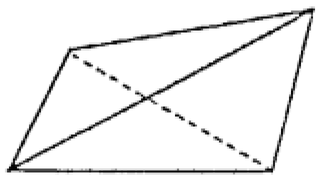


Figure 1

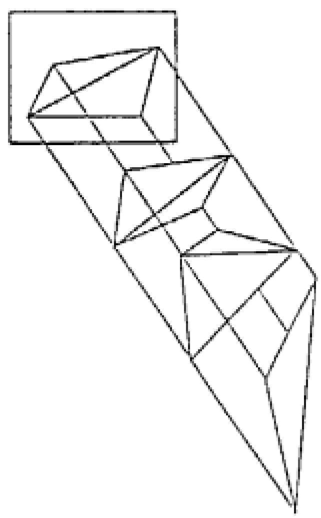


Figure 2

1 Introduction

This paper deals with line drawings obtained by orthogonal projection. Geometrically, the three-dimensional shape of an object cannot be decided quantitatively from a single two-dimensional line drawing such as Fig. 1. There is an infinite number of objects corresponding to a line drawing. Some examples of them are shown in Fig. 2.

In spite of a diversity of corresponding three-dimensional objects, man usually perceives one three-dimensional shape quasi-quantitatively from a single line drawing. The aim of this paper is to propose a cognitive model for perception of this kind.

First, section 2 examines previously published interpretation methods of a line drawing and points out their limitations. Section 3 describes an inclusive cognitive model for depth perception from a line drawing. Unlike most other approaches, it can explain the quasi-quantitative interpretation of irregularly-shaped and unfamiliar objects. Section 4 considers the interpretation of a line drawing which is drawn to communicate the three-dimensional shape.

2 Related Work

The method of labeling a line drawing was presented by Huffman [3] and Clowes [1], and extended by Waltz [9] and others. As shown in Fig. 3, the method qualitatively, but not quantitatively, interprets lines in an image as convex edges (+), concave edges (−), or occluding edges (→). Mackworth [6] proposed an interpretation method which uses the gradient space, but it does not produce quantitative interpretation, either. However, man perceives not only convexity or concavity, but also a quasi-quantitative three-dimensional shape from a single drawing. This paper proposes a cognitive model for quasi-quantitative depth perception.

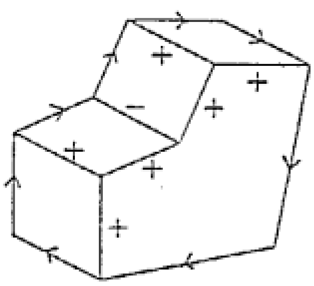


Figure 3

Roberts' system [8] interpreted a line drawing quantitatively on two assumptions:

1. The objects in the picture can be described by transformations of known models;
2. All objects are supported by other objects or by the ground plane.

A model is a representation of each possible type of object in a three-dimensional coordinate system.

Many models are needed for interpretating various kinds of objects, and as the number of models increases, the possibility that many models match a picture increases; therefore, the appropriate one must be selected from the matched objects, because man usually perceives one of them but does not perceive the others. Roberts did not handle this problem. The main interest in this paper lies in which of an infinite number of geometrically matchable objects is perceived.

The PICAX system [5] handles drawings in which three projected axes are specified. It quantitatively interprets drawings on the assumption that a line parallel to one of the projected axes represents a line really parallel to one of the real axes. If it cannot decide the three-dimensional coordinate of some of the vertices, it asks the user for construction lines for the unknown vertices. A construction line is parallel to the projected z axis. For example, the user must designate a construction line for the top vertex of a pyramid (Fig. 4).

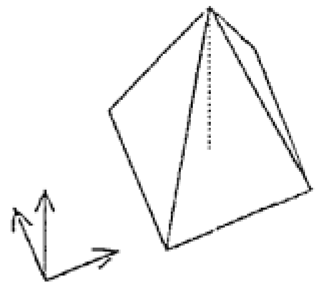


Figure 4

Hirotani et al.'s system [2] also handles the drawings in which three projected axes are specified. It quantitatively interprets drawings on two assumptions. One is that a line parallel to one of the projected axes represents a line really parallel to one of the real axes. The other is that objects are set on the xy plane (Fig. 5). However, the interpretation of a pyramid or an oblique prism cannot be determined from only these assumptions.

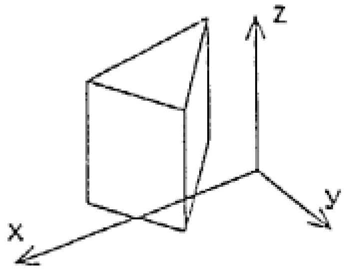


Figure 5

Kanade [4] proposed two regularity-based heuristics for the quantitative interpretation of line drawings: the parallel-line heuristic and the skewed-symmetry heuristic.

- The parallel-line heuristic is

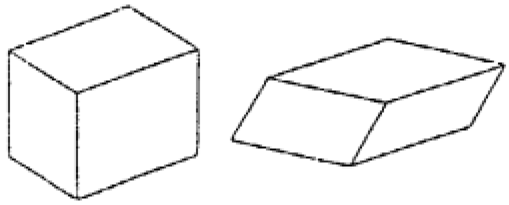
  "if two lines are parallel in the picture, they depict parallel lines in the scene".

- The skewed-symmetry heuristic is

  "skewed symmetry depicts a real symmetry viewed from some (unknown) view direction".

The concept of skewed symmetry is a class of two-dimensional shapes in which the symmetry is found along lines not necessarily perpendicular to the symmetry axis.

With these heuristics, regularly-shaped objects can be interpreted quantitatively, but irregularly-shaped objects cannot. For example, a drawing such as Fig. 6(a) can be interpreted quantitatively as a rectangular prism, but a drawing such as Fig. 6(b) cannot be interpreted quantitatively. The only definite statement that can be made about Fig. 6(b) is that it cannot be a right-angled block.



(a) (b)

Figure 6

As shown above, previously published interpretation methods of a line drawing do not wholly explain the characteristics of man's depth perception. It is common for man to perceive a three-dimensional shape without additional information such as a construction line even when the object includes no symmetrical shape, such as Fig. 1 or Fig. 6(b). The new cognitive model proposed in this paper explains the perception of this kind inclusively.

3 The Likelihood Model

Well-known objects and regularly-shaped objects can be interpreted utilizing strong constraints such as the object being similar to a predetermined object or the object including symmetrical shapes. However, man perceives the depth of an object even when there is no strong constraint such as these; therefore, we assume that man also utilizes some other weak constraints for depth perception.

The likelihood model proposed here inclusively explains the depth perception of various kinds of objects as follows:

1. Well-known objects,
2. Regularly-shaped objects,
3. Other kinds of objects.

The model represents strong constraints and weak constraints in a single framework, that is, the likelihood of an object producing the drawing. The basic idea of the model is to interpret a drawing according to the likelihood. That is, man normally interprets a drawing as the object which is most likely to produce the drawing, and it is difficult to perceive an object with a low likelihood although it can be perceived intentionally. This type of idea is known as maximum likelihood estimation in statistics.

The likelihood is defined as follows:

$$\begin{aligned} & Lt(D; O, V) \\ = \quad & Lg(D; O, V) \times Lo(O) \times Lv(V|O), \end{aligned}$$

where the left side of ; is an observed term, and the right side of ; is the estimated terms. | denotes conditional probability. Lt denotes the total likelihood which is the likelihood of object O with view V on given drawing D. The simple expression, likelihood, means the total likelihood. Lt is the product of Lg, Lo, and Lv, which are explained below. Lg denotes the geometrical likelihood which is the likelihood of geometrical correspondence, that is, the projective relation between given drawing D and object O with view V. Lo denotes the object likelihood which is the likelihood of object O's existence. Lv denotes the view likelihood which is the likelihood of view V of object O on condition that object O is selected. Putting it another way, Lg selects the geometrically possible objects, and one of them is perceived according to the strong constraint, Lo, and the weak constraint, Lv.

For example, suppose that drawing D can be interpreted geometrically as object $O1$ with view $V1$ or as object $O2$ with view $V2$ such that

$$Lg(D; O1, V1) = Lg(D; O2, V2).$$

If

$$Lo(O1) = Lo(O2),$$
$$Lv(V1 \mid O1) > Lv(V2 \mid O2)$$

hold, then

$$Lt(D; O1, V1) > Lt(D; O2, V2)$$

holds. This means that object $O1$ is normally perceived from the drawing.

It seems that humans learn how to reconstruct three-dimensional shapes from two-dimensional images from many examples of image-solid pairs. The likelihood model seems to be a natural result of this type of learning. That is, the object likelihood and the view likelihood depend on personal visual experience and expectation of a scene. Consequently, according to the model, a drawing tends to be perceived as a well-known object with a well-known view.

Against probabilistic models like this, there may be a criticism that man can interpret a drawing of an object which he has never seen. However, it does not deny the validity of the model, because even when he has never seen the object, he has certainly seen similar objects, and human judgment about geometrical matching is not very strict. In the model, rough matching can be represented using the geometrical likelihood, and another representation of rough matching is the likelihood of the projected area as an approximation to view likelihood, which is described in a later section.

3.1 Geometrical Likelihood (Lg)

Since human judgment about geometrical matching is not very strict, the value of the geometrical likelihood is at a maximum at strictly possible objects and decreases continuously around the maximum.

Although the likelihood model can handle ambiguities in geometrical matching of human vision by setting the geometrical likelihood as described above, it is not the main aim of this paper to treat such ambiguities. Therefore, for simplicity, this paper regards geometrical likelihood as a two-valued function such that

$$Lg(D;O,V) = \begin{cases} k(const.) & \text{if object } O \text{ with view } v \text{ exactly matches drawing } D \\ 0 & \text{otherwise.} \end{cases}$$

The paper does not refer to the mechanism for judging geometrical matching. The problem treated in this paper is how to choose one of the infinite kinds of geometrically possible objects.

3.2 Object Likelihood (Lo)

Object likelihood is the likelihood of an object's existence. This likelihood seems to be basically similar among humans, but varies, reflecting the interpreter's individual visual experience and expectations. That is, it depends on what he has seen and what he expects to see. This variation naturally corresponds to variations in interpretation among humans and variations in context.

In previous work, drawings have been interpreted assuming that objects are similar to the known model [8], or that objects have a symmetrical shape [4], or that objects have parallel lines [4]. These assumptions correspond to restrictions by the simplified object likelihood function. That is, in previous work, the object likelihood is a two-valued function such that

$$Lo(O) = \begin{cases} k(const.) & \text{if } O \text{ is an assumed object} \\ 0 & \text{otherwise.} \end{cases}$$

By these methods, complete interpretation is possible if and only if exactly one object is geometrically possible among assumed objects. Hence, the number of drawings which can be interpreted by these methods is strictly limited.

For the interpretation of general drawings, a many-valued object likelihood function is necessary as well as the view likelihood function. Examples of this case are shown in a later section.

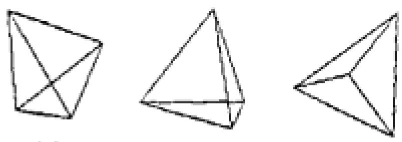
3.3 View Likelihood (Lv)

View likelihood is the likelihood of a view of an object on condition that the object is selected. If one is familiar with a view of an object, the view of that object has a high likelihood. For example, we are used to seeing cars on roads from the height of our eyes, so those views have a high likelihood. A similar bias arises for the ordinary objects on a desk. The assumption that objects are supported [8], and [2] can be regarded as one particular case of bias in view likelihood.

Some other researches can also be regarded as related to view likelihood. The parallel-line heuristic [5], [2], and [4] is based on the low likelihood of the special view at which nonparallel lines happen to be seen in parallel. Kanade [4] states that the least slanted planes are the most reasonable selection among the surface orientations which satisfy the skewed-symmetry heuristic if no other constraints are available. This can be regarded as another specific example of utilizing view likelihood. That is, view likelihood has been considered only in particular cases or only for a plane.

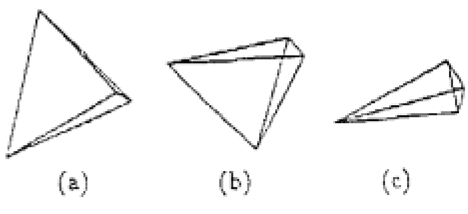
We consider general cases which have no particular bias such as those described above. If an object looks the same from many views, those views of the object have a high likelihood, and human judgment about geometrical matching is not very strict; therefore, if an object looks similar from many views, those views of the object have a high likelihood, too.

For example, each view of an isotropic object has a high likelihood, because, roughly speaking, it looks similar from many views (Fig. 7), and the change of figure when the view shifts is slow. The extreme case of isotropic objects is a sphere. The view likelihoods of flat objects are lower than those of isotropic objects, because flat objects look different from different views (Fig. 8). Among views of a flat object, the views nearly perpendicular to the flat extent of the object (Fig. 8(a)) have a higher likelihood than the views nearly parallel to the flat extent (Fig. 8(c)), because around the former views, the change of figure when the view shifts is slow, but around the latter views, it is quick. The extreme case of flat objects is a plane. Similarly, the view likelihoods of long and narrow objects (Fig. 9) are lower than those of isotropic objects, and among views of a long and narrow object, the views nearly perpendicular to the long axis (Fig. 9(a)) have a higher likelihood than the views nearly parallel to the axis (Fig. 9(c)). The extreme case of long and narrow objects is a line.



(a) (b) (c)

Figure 7



(a) (b) (c)

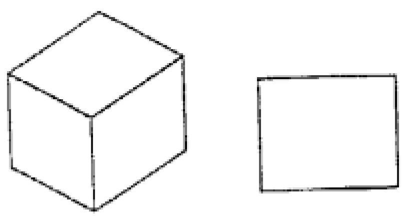
Figure 8

(a) (b) (c)

Figure 9

Here, we introduce an approximation to view likelihood so that it can be determined by calculation. It is the likelihood of a projected area, that is, the possibility that the projected area from an arbitrary view is nearly equal to the view's projected area. Although the projected area is insufficient to represent a view in detail and the view likelihood can also be affected by context or individual experience, the projected area likelihood is a good approximation because it reflects the total figure and can be calculated easily. The approximation can also be regarded as representing rough matching in human visual perception.

Moreover, we consider the relation between the projected area likelihood and the projected area itself. The larger the projected area, the larger its likelihood, and the projected area likelihood is at a maximum at the view angle at the maximum projected area, because around the view of maximum projected area, the projected area changes slowly as the view shifts. This can be understood intuitively if you think of a line or a plane as the extreme case of an object. (See for example, Fig. 8 and Fig. 9.) Consequently, we can say that the view likelihood and the projected area of the view are positively correlated. Here, we also point out that in a rectangular parallelepiped, oblique views such as Fig. 10(a) have larger projected areas than front views such as Fig. 10(b).



(a) (b)

Figure 10

Here, we propose a new heuristic which relates to the view likelihood. It is called a projected-area heuristic. This heuristic can be applied to any kind of object and is especially useful for unfamiliar objects which include no symmetrical shape. As shown in a later section, the probability that an interpretation based on the view likelihood is correct increases for communication-oriented drawings. Thus, the probability that an interpretation based on this heuristic is correct also increases in that situation.

- The projected-area heuristic is to

  "interpret a drawing assuming that it is projected from the angle from which the projected area becomes maximum".

Two examples of the application of this heuristic are shown below.

Fig. 11 shows a general triangular pyramid. The projected-area heuristic states that this is a view of the maximum projected area. From the assumption of a local maximum, it can be shown that

$z(A) = z(C)$,
$z(B) = z(D)$

where the projection direction is parallel to the z axis, and $z(A)$ represents the z coordinates of apex A.

From the assumption of the total maximum, it can be shown that

$| z(A) - z(B) | \leq limit$

where *limit* is a value which can be calculated from the drawing.

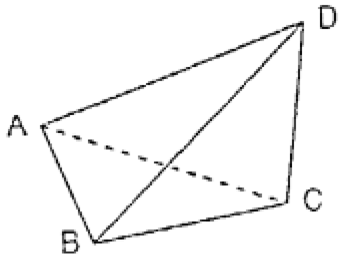


Figure 11

Suppose that Fig. 12 shows a parallelepiped. According to the projected-area heuristic, it can similarly be shown that

$z(A) = z(C) = z(H)$,
$z(B) = z(E) = z(G)$,
$| z(A) - z(B) | \leq limit$.

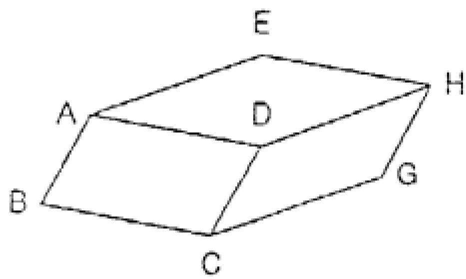


Figure 12

This paper has shown what can be determined by the projected-area heuristic. There remains some ambiguity in interpretation. A full interpretation can be obtained using both object likelihood and view likelihood, as shown in the following section.

3.4 Total Function of the Likelihood Model

The previous sections described the function of object likelihood and view likelihood separately. This section explains the total function of the likelihood model.

First, the relation between the object likelihood and the view likelihood is clarified. If there is one object with an outstandingly high object likelihood among geometrically possible objects corresponding to a drawing, that drawing is interpreted as that object, regardless of the view likelihood, because the variation of the view likelihood is generally not very large. Otherwise, the view likelihood is important. The product of object likelihood and view likelihood determines the interpretation.

The former case corresponds to the interpretation of well-known or regularly-shaped objects, and previous work mostly handles the former case. Example 1 is an example of the former case. Examples 2 and 3 are examples of the latter case. These two examples cannot be interpreted by conventional methods, because they do not consider the view likelihood.

Example 1: Rectangular parallelepiped

Geometrically, the object shown in Fig. 13 can be various hexahedrons. However, the object likelihoods of rectangular parallelepipeds are generally much higher than those of other hexahedrons, and there is one geometrically possible rectangular parallelepiped. Thus, the view likelihood does not affect the interpretation and the rectangular parallelepiped is perceived.

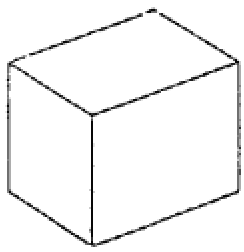


Figure 13

Example 2: Skewed parallelepiped

Geometrically, the object shown in Fig. 14 can be various hexahedrons other than rectangular parallelepipeds. The object likelihoods of skewed parallelepipeds are much higher than those of other possible hexahedrons. However, there are many possible skewed parallelepipeds, each of which has only a slightly different object likelihood. The view likelihood does not vary very much, either. Thus, both the object likelihood and the view likelihood affect the maximum total likelihood, and the skewed parallelepiped with the maximum total likelihood is perceived.

In general, the relations among the view likelihood of various objects with various views are as follows:

$$\begin{aligned} & Lv(\text{any view}|\text{ isotropic object}) \\ > \ & Lv(\text{view with large projected area}| \\ & \quad \text{nonisotropic object}) \\ > \ & Lv(\text{view with small projected area}| \\ & \quad \text{nonisotropic object}). \end{aligned}$$

Therefore, if we assume that the change in the object likelihood can be ignored, the most isotropic object of those supported by the projected-area heuristic, that is, the object with the maximum depth, tends to be perceived.

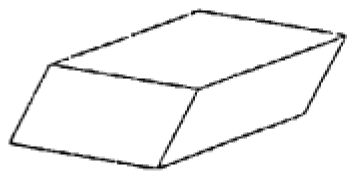


Figure 14

Example 3: Triangular pyramid

In Fig. 15, the object likelihood of possible objects does not vary greatly. Thus, in the same way as Example 2, both the object likelihood and the view likelihood affect the maximum total likelihood, and the pyramid with the maximum total likelihood is perceived. If we assume that the change in the object likelihood can be ignored, the most isotropic object of those supported by the projected-area heuristic, that is, the object with the maximum depth, tends to be perceived.

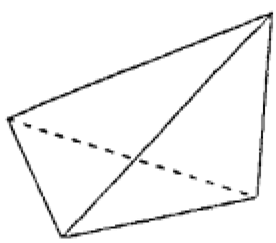


Figure 15

4 Line Drawings for Communication

The interpretation which depends on the view likelihood does not always agree with the scene which produces the drawing. However, this is also the case for human vision. The model proposed in this paper should be evaluated with accuracy in simulating human perception, not with accuracy in recovering the original scene.

If drawings are limited to those which are drawn to communicate the three-dimensional shape of an object, the accuracy of interpretation by the likelihood model increases greatly. This is also true in human vision. This is because drawings are made to communicate the three-dimensional shape as accurately as possible taking into account usual human interpretation. That is, a drawing for communication is usually made as follows:

1. It is drawn from a familiar view, that is, a view with a high likelihood. We do not choose the view from which the object can be perceived accidentally as a regularly-shaped object.
2. We choose the view which can communicate a large amount of information, that is, the view with a large projected area and small depth.

It is interesting that views with a high likelihood correspond to views which communicate a great deal of information. This is easy to understand if the projected area is regarded as a measure of the amount of information. That is, views with a high likelihood have a large projected area, and therefore have a large amount of information.

5 Conclusion

This paper proposed a new cognitive model which gives a unified explanation of the interpretation of various objects based on object likelihood and view likelihood. Consideration of the view likelihood enables interpretation of objects which have no symmetrical shape. The projected area likelihood was proposed as a good approximation to view likelihood. It was shown that the probability that an interpretation based on view likelihood is correct increases if drawings are limited to those which are drawn for communication.

This likelihood model can be applied to interpretation for input of CAD systems or animation systems. The present condition is that users of those systems must designate the full information of shapes accurately even when the accurate shape is unnecessary for their purpose, because there is no way to communicate an outline. The likelihood model is suitable for approximate recovery of shapes from line drawings.

It is planned to carry out psychological experiments to determine the likelihood functions and to implement the model from the results of the experiments. Another direction of future research is to generate the likelihood model automatically by learning from examples.

Acknowledgments

I would like to thank Tsutomu Matsuura in Gunma University, Katsuhiko Noda and Fumiyuki Kato in Matsushita Electric Industrial Co., and Koichi Furukawa and Hirohisa Seki in ICOT for helpful discussions and comments.

References

[1] Clowes, M. B., On Seeing Things, *Artificial Intelligence 2:* 79-116, 1971.

[2] Hirotani, T., Fukui, Y., and Ohira, T., A Study on Interactive Computer Input System, in *Proc. 3rd Symposium on Computer Use, Architecture Society of Japan,* 343-348, 1981 (in Japanese).

[3] Huffman, D. A., Impossible Objects as Nonsense Sentences, in R. Meltzer and D. Michie (eds.), *Machine Intelligence 6,* pp. 295-323, Elsevier, 1971.

[4] Kanade, T., Recovery of the Three-Dimensional Shape of an Object from a Single View, *Artificial Intelligence 17:* 409-460, 1981.

[5] Liardet, M., Holmes, C., and Rosenthal, D., Input to CAD Systems: Two Practical Examples, in Latombe (ed.), *Artificial Intelligence and Pattern Recognition in Computer Aided Design,* pp. 403-427, North-Holland, 1978.

[6] Mackworth, A. K., Interpreting Pictures of Polyhedral Scenes, *Artificial Intelligence 4:* 121-137, 1973.

[7] Oka, N., Quantitative Interpretation of Line Drawings Based on Default Understanding, in *Preprint of 32nd National Conference of Information Processing Society of Japan,* 1433-1434, 1986 (in Japanese).

[8] Roberts, L. G., Machine Perception of Three-Dimensional Solids, in J. Tippett (ed.), *Optical and Electro-Optical Information Processing,* pp. 159-197, MIT Press, 1965.

[9] Waltz, D., Generating Semantic Descriptions from Drawings of Scenes with Shadows, in P. Winston (ed.), *The Psychology of Computer Vision,* pp. 19-92, McGraw-Hill, 1975.