The Projective Approach to Object Description

VISION FLASH 37

by

J. M. Hollerbach

Massachusetts Institute of Technology

Artificial Intelligence Laboratory

Robotics Section

December 15, 1972

Abstract

A methodology is presented for generating descriptions of objects from line drawings. Using projection of planes, objects in a scene can be parsed and described at the same time. The descriptions are hierarchical, and lend themselves well to approximation. Possible application to curved objects is discussed.

This paper reproduces a thesis proposal of the same title submitted to the EE Department for the M.S. degree.

Work reported herein was conducted at the Artificial Intelligence Laboratory, a Massachusetts Institute of Technology research program supported in part by the Advanced Research Projects Agency of the Department of Defense and monitored by the Office of Naval Research under Contract Number N00014-70-A-0362-0003.

Vision flashes are informal papers intended for internal use.

1. Introduction

For nearly any task one might posit for a machine vision system, there must occur some form of object identification in the visual scene. Central to the identification task are the perception of "objectness", and the subsequent description of the object by extracting appropriate features for identification. This imposes two problems on a description generator: to recognize what the features are, and to select those features that are necessary for identification.

The world is not composed of simple objects like cubes and wedges, but of complex objects that defy exact description. In fact exact descriptions are hardly ever desirable, and if available may serve only as a source of confusion. The keyword then is approximation. Approximation can come about by either ignoring certain perceived features, or by simplifying others. Previous works on object description pretty much ignored the need for approximation, and even in describing fairly simple objects they were not very successful.

The most significant early approach to the problem was that of Roberts (1). He obtained descriptions by projecting three-dimensional models into the scene to obtain exact fits with objects in the scene. If this were not possible with the models available, wedge and cube models were used to split the complex object into simpler pieces. At the end a complex object would be described as a conglomeration of wedges and cubes. This is a prime example of an approach that yields an exact, but unusable description.

The second approach worthy of note was Guzman's (3), which was later cast into a learning framework by Winston (4). All the topologically different ways that a particular object could be projected into 2-space served as models for that object. There are two objections to such an approach: first, the failure to include some notion of projection in the models necessitated multiple models for each object; and second, this scheme really loses for complex objects or for objects that are only slightly different from a model.

The best success in the past has not been with object description, but with recognition of "objectness", i.e., parsing a scene into bodies. Notable in this regard have been the works of Guzman (5), Huffman (6), and most recently Waltz (7). Even after parsing a scene into bodies, however, there remains the problem of describing each separated body.

The main goal of my proposed thesis work is to create a descriptive mechanism that generates useful descriptions of complex objects from line drawings. It's basis is the projection of planes. Using this technique, I hope to show that the objections raised to the previous works can be met, and that in fact the processes of separation, description and approximation, which have hitherto been studied independently, can be carried out simultaneously. Briefly, a plane can be projected to form a body if it satisfies certain conditions to be described in the next section. Separation is accomplished by finding such planes in a scene and projecting them into bodies. Identification comes about by equating an object with the projection of its most complex plane. When the object is too complex to be so described, then plane projection induces a natural decomposition of the object into such pieces, as was shown in Vision Flash 31 (2). Modelling an object as the projection of its most complex plane, incidentally, is projectively invariant. Finally, the projective approach is well suited for approximation, since it is possible to project through and smooth out such minor irregularities as protrusions and indentations.

The choice of lines as the basis from which to build descriptions is firmly rooted in past work. Lines are clearly the best indicator of shape for polyhedra, and I hope to show that they are passable shape indicators for curved objects as well.

2. Plane Projection

To indicate what is meant by plane projection, let me first define two terms. By plane will be meant a planar region, and a ray is defined to be a line from a vertex of a given plane that is not an edge of that plane. A plane projection is the process of moving a plane along its rays; for example, the block in fig. 1 can be described as the projection of the rectangle A along its rays r1, r2 and r3.

In a later section, the rays along which a plane is projected will be interpreted as inducing a preferred orientation or direction in the scene. Projectability therefore is a concept most readily applied to object planes with convex edges only. For, the presence of a concave edge in a plane means there will be rays oriented in a generally opposite direction to the projection, and there is consequent ambiguity in preferred direction.

The restriction to planes with convex edges means that a projectable plane may have only type 1 and type 3 vertices, as indicated in fig. 2A (see Huffman (6) for a discussion on vertex types). There are a small number of ways in which these vertices may be pairwise connected around a plane, dependent only on a common edge label. In order to enumerate these possibilities, the plane must be located with respect to the vertex lines, and for this purpose the labeling scheme of Waltz (7) shown in fig. 2B will be used. The result of this enumeration is essentially the finite state machine in fig. 3, which has a few additional features discussed below.

The concept of using a regular grammar to generate planes with type 1 and 3 vertices is originally due to Waltz. He has only published a grammar, however, for planes with type 1 vertices in (11).

A plane is accepted by the FSM if, when applied to the vertices in a clockwise direction, the FSM ends in the same state from which it began. It is clear all object planes with convex edges are accepted by the FSM, but it is not clear that the FSM recognizes only such planes. Nevertheless, this assumption is the basis of projectability, namely, any plane accepted by the FSM may be projected to form a body. An additional restraint has been imposed on the FSM to enter two consecutive states from the set (AO,A1+,A2+,F), the reason for which will be offered later.

If there is a ray along which the projection of an edge is visible, we would like as a condition that this edge be unobstructed so as to permit a clear projection. Such obstructions manifest themselves as generalized T joints (see vertex type TO+ in Fig. 2C). If the condition does not apply to a particular edge, then there may be abitrarily many TO+ vertices along it and arbitrarily many rays outside the end vertices. The arbitrary rays are depicted in Fig. 2C, and lead to the modified labels A1+, A2+, LO+, L1+ and TO+. An arbitrary number of TO+ vertices along an edge are depicted by a * on the transition.

With these modified labels we are able to take into account accidental alignment.

It should be noted that there is possible ambiguity at some vertices with regard to type. An F vertex, for example, could also be interpreted as an L1+. Only one of these vertices, however, will work on the FSM, which forces the proper interpretation of such vertices.

It turns out that projectability of a plane does not guarantee that an object can thereby be realized. The two impossible objects in Fig. 4A and 4B taken from Huffman (6) each have projectable planes labeled A. If the vertices in 4A are restricted to be trihedral, then edges e1, e2 and e3 must meet in a single point, yet they do not. The example in 4B is clearly nonsense, so that although projectability is more global than vertex hacking, it is still local to a particular portion of an object.

To check the realizability of such objects, Huffman (6) has developed a unity gain criterion for the cyclically ordered set of edges of a plane. This criterion can also be applied to locate the position of hidden rays as indicated by I vertices. If there are two consecutive L vertices, there is one degree of freedom in locating their respective rays. The requirement of two consecutive vertices from the set (A0,A1+,A2+,F) mentioned earlier has been imposed to reduce the possible degrees of freedom in locating such rays. Ordinarily this restriction will be of no consequence.

The projectability of a plane does not automatically guarantee that its use will bring about the best description of an object. For example, in Fig. 4C the FSM interprets vertices v1, v2, and v3 as L1+, I0+, and L1+ respectively, and renders plane A projectable. By its use the decomposition in 4D is brought about. As is pointed out in a later section, what is actually done in such circumstances is to investigate all planes encompassed by the projection of A for the best description. In so doing the more reasonable decomposition in 4E is obtained.

If we equate "objectness" with projectability, it can be seen that "objectneess" is a simple problem, since a solution can be modeled by a FSM and with only a few states in that machine. This definition of projectability is superior to that given in Vision Flash 31, where there was a requirement for parallel rays and no accidental alignments. It seems that humans have an easy facility for recognizing objects without parallel rays, and that accidental alignment causes us no particular trouble. This ability has apparently been captured in the present definition.

3. Separation

The finite state machine in itself is inadequate to parse a scene into bodies, since a body may be obstructed in such a way as to leave no visible projectable planes. There usually are some bodies in the scene, however, that can be immediately identified by projection. Deletion of such bodies will unobstruct other bodies, and the process can be repeated. Of course when a body becomes unobstructed, some conjecture must be made as to the identity of the hidden part, and it is possible to concoct simple rules for this purpose.

Given this general scene parsing procedure, the first problem is to select a place in the scene to start. One possibility is to find all projectable planes and to form the appropriate bodies from them. For purposes of uniformity with description considerations discussed in a later section, I have chosen to start with the largest non-background plane. If this plane can be projected to form a body, then the body is deleted and the scene reconstructed. The parsing continues with the new largest plane.

If the plane is not projectable, there is an obscuring body that must be removed before the plane can be projected. It is easy to locate the obscuring body by noting what rays ruin the projectability of the plane. For example, in fig. 5A plane 1 is the largest, but is not projectable. Looking at vertex v, we notice that it can only be interpreted as L1+, and this ruins the possibility of having two consecutive vertices from the set (A0,A1+,A2+,F). The focus of attention is now placed on rays r1 and r2, and on the plane they commonly bound, plane 4. It turns out plane 4 is projectable in two different directions, towards plane 2 and towards plane 3. Depending on which direction is chosen, the parsing in 5B or 5C is obtained.

It should be noted that in 5D plane 4 is projectable only in the direction of plane 2 because of colinearity of an edge of planes 3 and 4. The parsing would then be like that in 5B, which seems to correspond with human preference for the scene.

If plane 4 is projected towards plane 2 and the corresponding body deleted from the scene, we are left with the situation in 5E. The reconstruction of planes 1 and 3 takes place by application of the reconstruction rules, which are listed below. These rules are applied in the order given.

1. Join a split edge (e.g., fig. 6A).
2. Extend two lines to a corner when this makes sense (6B).
3. Extend parallel lines between neighboring regions (6C).
4. Hypothesize a best completion when lines are parallel or do not meet at a reasonable spot (6D).

Application of rule 4 to fig. 5E gives us the familiar cube in 5B.

The procedure of finding what obstruction destroys the projectability of a plane and of removing it is recursive, since the obstruction may itself be obstructed. The recursion continues until an unobstructed body is found, at which point we are able to remove it and work backwards. Consequently in removing a particular obstruction, it may be necessary to remove a number of them.

As an example of the application of the procedure to a more complicated scene, consider fig. 7A. Plane 1 is the largest, but is rendered unprojectable by a number of obscuring objects. There are several planes that would be indicated as possibly belonging to an obscuring object, and suppose our attention is turned to plane 3. Unfortunately plane 3 is not projectable because of the presence of plane 15, and plane 15 is not because of plane 7. But plane 7 can be projected to form a cube 7-8-9, and its deletion leaves the scene as depicted in 7B, whereupon application of the reconstruction rules yields 7C.

In subsequent steps the deletion and reconstruction depiction will be combined. Now plane 15 is projectable, but an obstruction is met in the form of planes 11-13. However, this obstruction can be removed by projecting plane 11 to yield the cube 11-12-13, and the scene appears as in 7D. The projection of 15 can now be completed to yield the wedge 15-6-10 (7E). Finally plane 3 can be projected to yield 3-5-14, and we are left with a projectable plane 1 (7F).

This procedure has been applied to most of the scenes in

Guzman (5), and is successful on them. It is particularly encouraging that the simple reconstruction rules do so well in creating the obscured parts of objects. To be sure, one can construct examples where the reconstruction rules do not yield the most desirable interpretation, and it is likely these rules will have to be augmented with higher level considerations.

When scenes have shadows the procedure fails, since all lines are assumed to represent valid edges of an object. Moreover, there does not seem to be an easy way of extending it to handle shadows. Some indication exists nevertheless that it can profitably be coupled with Waltz's scheme (7) to yield a better parser for shadowed scenes. This comes about because the projective approach works well with aligned objects but poorly with shadows, while Waltz's scheme works well with shadows but not as well with alignment.

There might be some criticism of the approach as being too dependent on a perfect line drawing, which as everyone knows is hard to come by. Allowance can be made for missing lines in the projective approach, and in fact missing lines can sometimes be easily predicted. In fig. 8A, for example, at the end of the projection of plane A a missing edge is detected. Such information could be sent to a line finder, or the assumption of an edge could be made and the processing continued. In fig. 8B, neither plane is projectable, so alternate descriptions are set up using each of them. Of the two, plane A clearly yields the better description, and so an edge is hypothesized between v1 and v2.

4. What The Features Are

The simplest types of objects to describe and recognize are those that can be considered to be projections of their most complex plane, which class of objects will henceforth be called projectable surfaces. They form the atomic building blocks from which more complex object are constructed, and it was shown in VF 31 (2) how to systematically decompose complex objects into such parts. The answer then to the question of what are the features of a scene is that they are projectable surfaces.

Identification comes about when perceived features are matched with those of a model. All models are consequently expressed in terms of projectable surfaces. For example, simple objects such as block and wedge can be expressed as arbitrary projections of rectangles and triangles (fig. 9A and 9B), respectively. A pyramid can be modeled as projection of a triangle with linear scale change as a function of distance projected. The complex object in fig. 9D could be modeled as an L-shaped object with a cube on it.

A rectangle under an arbitrary projection into 2-space rarely appears as a rectangle, of course, but as a parallelogram when there is no perspective deformation and as a trapezoid or quadrilateral with deformation. Therefore when perceived features are matched against models, auxilliary considerations are required to equate the complex planes of the models against the deformed planes found in the scene.

Given that objects are essentially identified by their complex plane, it is not always desirable to describe this feature exactly. For example, the objects in fig. 10 are all projectable surfaces, and each is describable by specifying the shape of plane A. Yet perhaps the better description of 10A is a block with an indentation and of 10B a block with a protrusion. For the object 10C in which there is a group of similar indentations, an appropriate adjective like jagged or saw-toothed is probably best applied to the modified edge of the block. What is required in such cases is to simplify the features by considering the plane to be a modification of a simpler plane. The question then becomes to determine what is the simpler plane and what is the modification. The answer to this question is independent of the projective approach, although as will be seen in the next section the two can be coupled together to obtain better descriptions.

The determination of simpler planes can depend on circumstances, but in general they will be such regular planes as square, rectangle, triangle, L-shape, T-shape, etc. There are basically two ways to modify a plane and still leave a suggestion of its original shape: make an indentation or add a protrusion. How to recognize under what circumstances a part of a plane is an indentation (I), protrusion (P), or an integral part of it (N for neither) was the subject of a poll of Vision Group membes and friends. Systematic modifications were made to squares, and people were asked to categorize each modification as an I, F or N. A poll was felt necessary to average out individual biases and inconsistencies, and in fact consistent results were thereby obtained. Sample results on some modifications are presented in fig. 11.

On the basis of these results, a parameterization of the figures was sought that would split them correctly into the above three groups. By plotting the ratio of the depth of a protrusion to the height of the completed square versus the ratio of the total gap on either side of the protrusion to the protrusion breadth, the figures were split as indicated in fig. 12. When two or more protrusions eminated from the side of a rectangle, the parameters were obtained by considering only the largest.

This parameterization has a simple interpretation. Protrusions must be sufficiently isolated from the rest of the object to resist integration as part of an indentation, which happens when the gap:breadth ratio becomes large enough. Yet the protrusion must not be so large as to become significant with respect to the size of the rest of the object, as indicated when the depth:height ratio approaches one. In this case the object is composed of at least two roughly equal and distinct pieces, and hence receives an N categorization.

Some interesting anomalies arose from the poll that could not be explained by the parameterization. These anomalies disappeared, however, when the simplifying effect of symmetry was taken into account. For example, the protrusions of objects 11O and 11P, 11Q and 11R, and 11S and 11T, respectively, are proportionately equal. Yet the symmetry in objects 11O, 11Q and 11S cause the gaps to be seen as indentations, while the asymmetry of 11P, 11R and 11T cause the protrusions to be seen as protrusions.

Another interesting result is the discrepancy between objects 11U and 11V, and between 11Q and 11W. Once again, the top protrusions are of proportionately equal size, yet in one case the protrusion is symmetrically placed and in the other it isn't. They were interpreted, respectively, as P and N. A final mystifying result was obtained for 11X, which because of its symmetrical shape resisted decomposition.

The conclusion drawn from these anomalies is that in describing a feature, we are more likely to interpret irregularities as indentations than protrusions because they are visually simpler. Similarly, we are more inclined to interpret a modification as a protrusion than to assume that the feature has an atomic but more complex shape consisting of the protrusion and the remainder. These proclivities should be incorporated into a descriptive mechanism to render the descriptions more compatible with human preference. A possible explanation for these proclivities will be offered in a later section.

5. What The Pertinent Features Are

The selection of the proper features for identification of an object may depend on semantics, i.e., on the exact function of a feature with respect to the whole object. The keyhole, for example, in fig. 13 is certainly more important to the identity of the padlock than is the chunk missing from the casing. Yet to a large extent the choice of features can be done on purely syntactic grounds, and this will be the approach taken.

The basis for syntactic selection is size. A large object is noticed before small objects, and this observation is mirrored in the description. Thus when describing a feature as a modification of a simple plane, the simple plane is more prominent in the description since it is larger than the modification. What is suggested then is that description ought to be based on a hierarchy of detail. There exists one or more primary centers of attention of an object, identified as the largest solid components. Secondary centers of attention are then located and related to the primary centers. This process can be continued recursively down to the smallest detail, or a choice can be made to stop at some particular level of detail.

Because of projective distortions, it is unclear what the largest component is without a priori knowledge of the object's identity. It has been decided to consider the largest component as that component which includes the apparent largest plane. That is to say, a plane is sought that when projected encompasses this largest plane (the projected plane may be the largest, although this is not necessary). This plane may be smaller in 3 dimensions than some other plane, and as a result different descriptions might be obtained if the object is viewed from a different perspective, although this is probably not a serious problem. Nevertheless, the choice seems justified on the basis of human judgement of volume, which is often based on apparent area.

It is now possible to outline a general procedure to produce a hierarchical description of a complex object.

(1) Find the largest plane.

(2) Form from it a body by projecting some plane.

(3) Smooth out indentations and protrusions while projecting.

(4) If there are protrusions, go to (1) with each of them.

The coupling of projection and feature simplification takes place in step (3), as was intimated in the previous section. When an obstruction is met during the projection, a choice is made on the basis of shape and relative size of the obstruction with respect to the projected plane whether the projection should be carried past the obstruction, or whether it should stop and break the object into two parts at that point. The basis for this decision is from feature simplification considerations like those presented in the previous section. In step (2) feature simplification of the plane occurs after its projection.

It is perhaps informative to apply the procedure to an example. Plane A in fig. 14A is evidently the largest, and so becomes the center of attention from which the main component will be fashioned. The FSM tells us that A is not projectable, and that the source of difficulty lies with plane C. Plane B is projectable and can encompass A in the projection, but as seen in 14B the resultant main body would have to be modified with an indentation formed by removing the block with plane D. On the basis of feature simplification, it can be determined that due to the relative sizes of planes C′ and D, it is better to describe B as a block with protrusion formed from C than the present interpretation.

Thus a component is formed by projecting C, and it is removed from the body (14C). But feature analysis indicates C is best considered to be a rectangle with indentation, and leads to the description of the protrusion as a block with a cubic indentation at a corner.

Analysis continues by recognizing that A is now projectable, and yields the decomposition in 14D. Once again, feature analysis indicates plane A should be described as a rectangle with an indentation, and the description of the main body is now a block with a smaller block missing from a side. It should be noted that in projecting A an indentation and protrusion were smoothed out. Plane E can be projected nearly an arbitrary length to form the protrusion, but the simplest assumption is that it extends along the whole length of one side.

The hierarchical description of 14A is as follows. At the top level, the complex object is basically a block. The two major modifications on the basis of size are a block protrusion resting at one corner, and a block indentation from a side. The remaining protrusion and two indentations form the third level of detail.

In the section on separation, the objects were implicitly assumed to be projective surfaces. It is a simple matter, however, to integrate complex object description with object separation. Once part of a complex object is separated, the largest of its planes that has been located becomes the main center of attention. All of the complex object might be identified by application of the procedure, or there might remain some residue that awaits removal of an obstruction.

6. Why The Projective Approach Wins

It will be noticed that the description methodology developed thus far works without the use of models. Insofar as the methodology is completely guided by the features of a particular scene, it is procedural in nature. If a link can be made between intellectualizing and modelling, then this methodology suggests that object recognition is a primitive human process not dependent on thinking.

H.A. Simon (8) categorizes descriptions as being either state or process. A state description of a cube, for example, would be the location of its vertices. Process descriptions, or procedural definitions in local terminology, are particularly suitable for implementation on computers, since they not only make explicit pertinent features, but also how to search for them. It is now clear that the expression of models as projections of planes is a process description, since to find an instance of a model merely requires locating a similar plane in the scene that has the same projective characteristics.

The identity of an object is essentially determined by its most complex plane. The quadrilaterals or whatever that are encompassed by the projection of this plane are unimportant to the object's identity, and serve merely as 3-D filler to give the object extent. A canonical representation of an object is thus essentially two-dimensional. Some earlier attempts at object description recognized the 2-D nature of description, but were unable to account for and ignore the 3-D filler, and consequently failed.

That an object is identified by a characteristic plane has been a part of some art styles, and strengthens the credibility of the projective approach. Early Christian art, for example, is characterized by a lack of depth or perspective, since the vogue at the time was to represent objects by a frontal and hence essentially two-dimensional view (9). Drawings of younger children also exhibit a tendency towards canonical representation of objects by a characteristic plane.

A hypothetical model of human vision can be drawn that explains some of the previous results. It is not claimed that any distinct physical process corresponds to elements of the model, but only that there is a general tendency of human vision to follow the model. Certain features of a visual scene impose a preferred direction on eye movement, such as decreasing intervals. Gibson (10) argues effectively that texture has this property; for example, the coarse texture of a plowed field gives way to finer texture as the field becomes more distant. The oblique lines eminating from a frontal plane also induce a preferred direction of scanning. When our eyes follow a path of decreasing intervals or a set of oblique lines, we have the sensation of moving back into the picture, i.e., of putting in the third dimension. This selfsame effect is obtained by my methodology when the edges from a plane are followed during projection.

Scanning the eyes in a straight line is the simplest form of eye movement possible, so that objects that can be comprehended in a straight scan are visually the simplest. This class of objects is just the projectable surfaces, and their use as the atomic building blocks for complex objects is thereby corroborated. For, when there is some obstruction to the line of sight, the direction of eye movement must change to scan the obstruction. The process of shifting the direction of eye movement is analogous to splitting the object into two pieces at that point.

It is now possible to explain why indentations are visually simpler than protrusions. Indentations encountered while scanning do not change the line of sight. Instead, a decision must be made to stop at that point or to continue scanning, thereby implicitly filling in the indentations. On the other hand, protrusions do force a change of direction as mentioned above and make objects more complex.

7. Curved Objects

The application of the projective approach to a subset of curved objects called quasi-rectilinear by Guzman (5) is immediate. Examples of such objects are the casing of the padlock in fig. 13, a violin case, a suitcase, and a cylinder. The description of the feature, i.e., the projected plane, of such an object is probably a little more difficult than for polyhedra, but it is expected that the same principle of feature simplification and modification apply.

The projective approach, however, is not very applicable to the general class of curved objects. It is not necessary for curved objects to have planes, which are integral to the success of the approach. There is consequently no readily identifiable feature corresponding to 3-D filler, and one is committed to working almost exclusively with the outline. Even if planes exist, they are often less important to the identity of an object than the path of projection. For example, the bottle in fig. 15 has a circular plane A on top, but the identity of the bottle is not revealed by this plane, even though there exists a projection path using A that describes the body. Rather, we recognize the bottle by its outline, almost ignoring A.

When describing such curved objects by their outlines, a fundamental assumption is made; namely, the object is round. Plane A in fig. 15 is an affirmation of the roundness of the bottle, rather than an important part of the description. There are other ways in which the roundness of a curved object can be inferred, such as texture, highlight and shadows.

Whenever a curved outline is in fact perceived, the roundness of the object is almost automatically inferred, since flat curved surfaces such as a disc are relatively rare in nature. The inner surface is induced by the contour, and unless there is contrary evidence assumes the smallest and simplest shape imaginable, namely circular or round. A parallel may be drawn with children's drawings and with physics (9). Circles have priority in very young children's drawings, since they depict nearly everything as being approximately round. Finally, there is a tendency towards simplest possible surfaces in physics; e.g., a dip wire in a soap solution gives a soap film of smallest possible surface.

The problem then is to specify the shape of the outline, and to modify such shape on the basis of internal perceived features. Once again simplification is necessary to render descriptions useful, and it is likely that the only atomic components required for this purpose are cylinder, cone and ellipsoid. A cylinder corresponds to an outline with essentially parallel sides, a cone to an outline with converging sides, and an ellipsoid to an outline with curved sides. The bottle, for example, can be considered to be composed of a cylindrical neck and main portion connected by an ellipsoidal part. The generation of the appropriate simplifications to an outline is the main difficulty in this approach, and is the subject of further reasearch.

References

(1) Roberts, L., "Machine Perception of Three-Dimensional Solids," Technical Report 315, Lincoln Laboratory, M.I.T., May, 1963.

(2) Hollerbach, J., "An Approach to Three-Dimensional Decomposition and Description of Polyhedra," Vision Flash 31, A.I. Laboratory, M.I.T., July, 1972.

(3) Guzman, A., "Some Aspect of Pattern Recognition by Computer," Project MAC Technical Report MAC TR 37, M.I.T, February, 1967.

(4) Winston, P.H., "Learning Structural Descriptions from Examples," A.I. Technical Report 231, A.I. Laboratory, M.I.T., June, 1971.

(5) Guzman, A., "Computer Recognition of Three-Dimensional Objects in a Visual Scene," Project MAC Technical Report MAC TR 59, M.I.T., December, 1968.

(6) Huffman, D., "Impossible Objects as Nonsense Sentences," Machine Intelligence 6, pp. 295-323 (ed. Meltzer, B. & Michie, D.), Edinburgh University Press, Edinburgh, 1971.

(7) Waltz, d., "Shedding Light on Shadows," Vision Flash 29, A.I. Laboratory, M.I.T., July, 1972.

(8) Simon, H.A., "The Architecture of Complexity," Nature, 1962.

(9) Arnheim, R.A., Art and Visual Perception, University of California Press, Berkeley, 1954.

(10) Gibson, J., The Perception of the Visual World, Houghton Mifflin Company, Boston, 1950.

(11) Waltz, D., "Understanding Scenes with Shadows," Vision Flash 21, A.I. Laboratory, M.I.T., November, 1971.

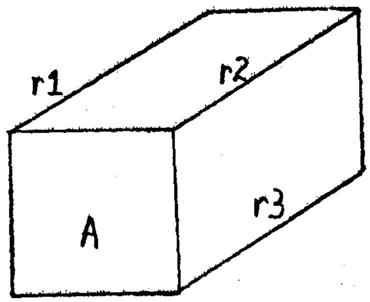


Fig. 1. A block.

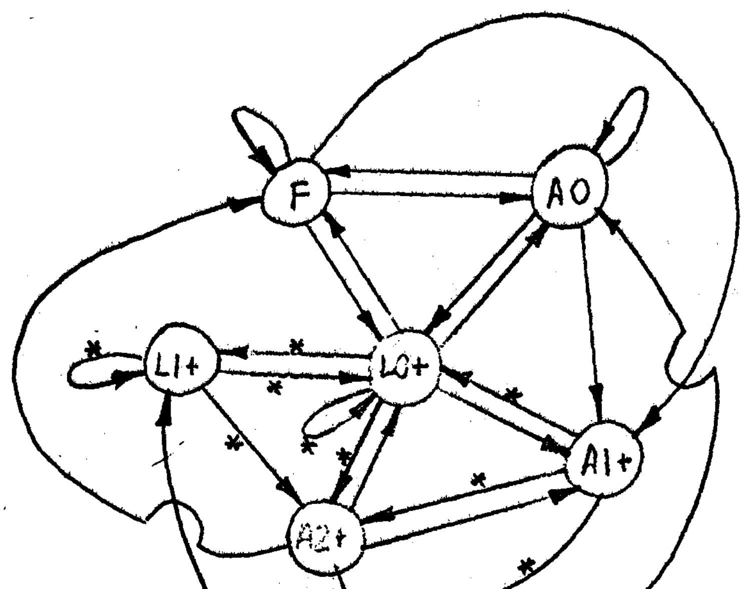


Fig. 3. FSM to recognize projectability.

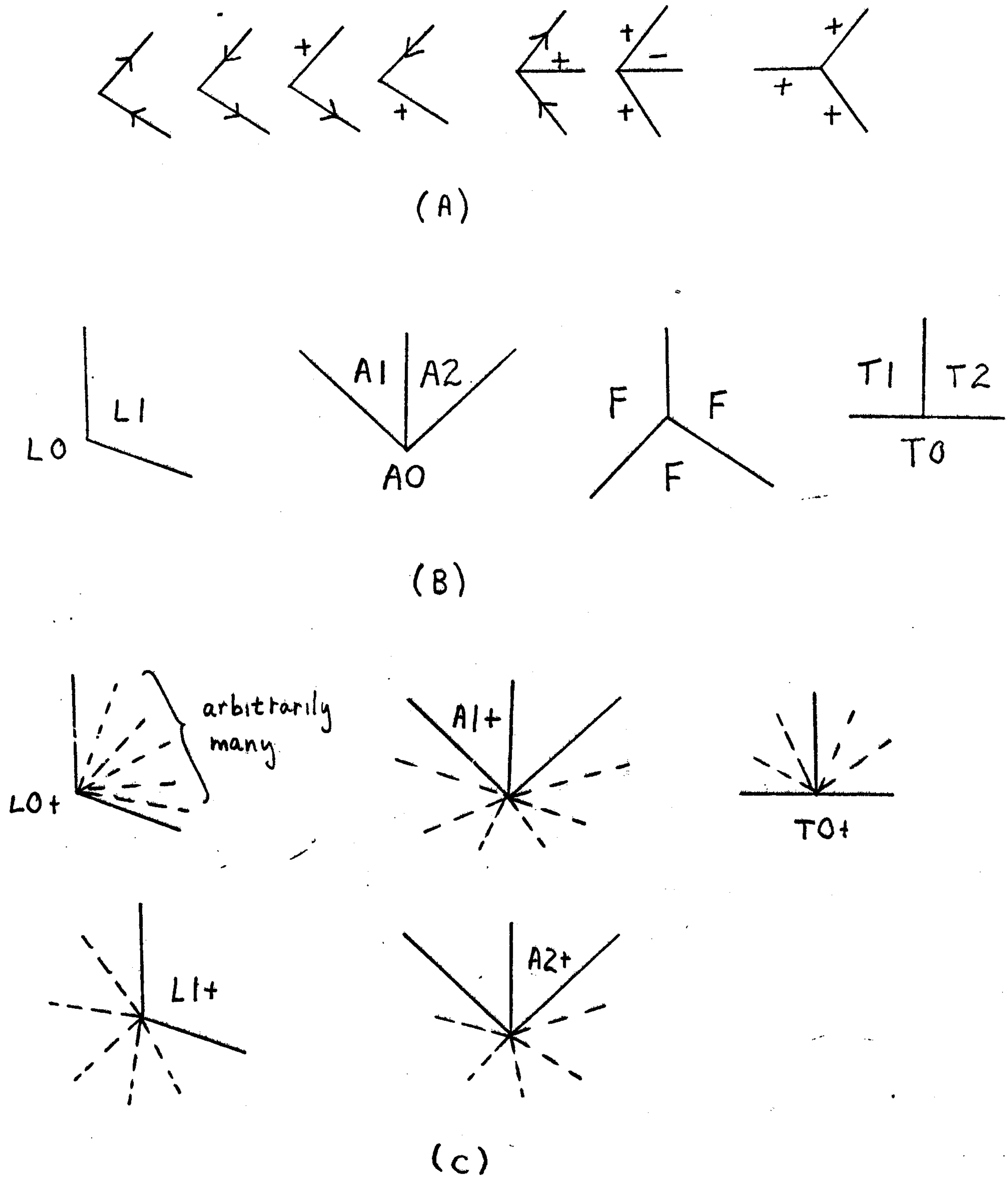


Fig. 2.

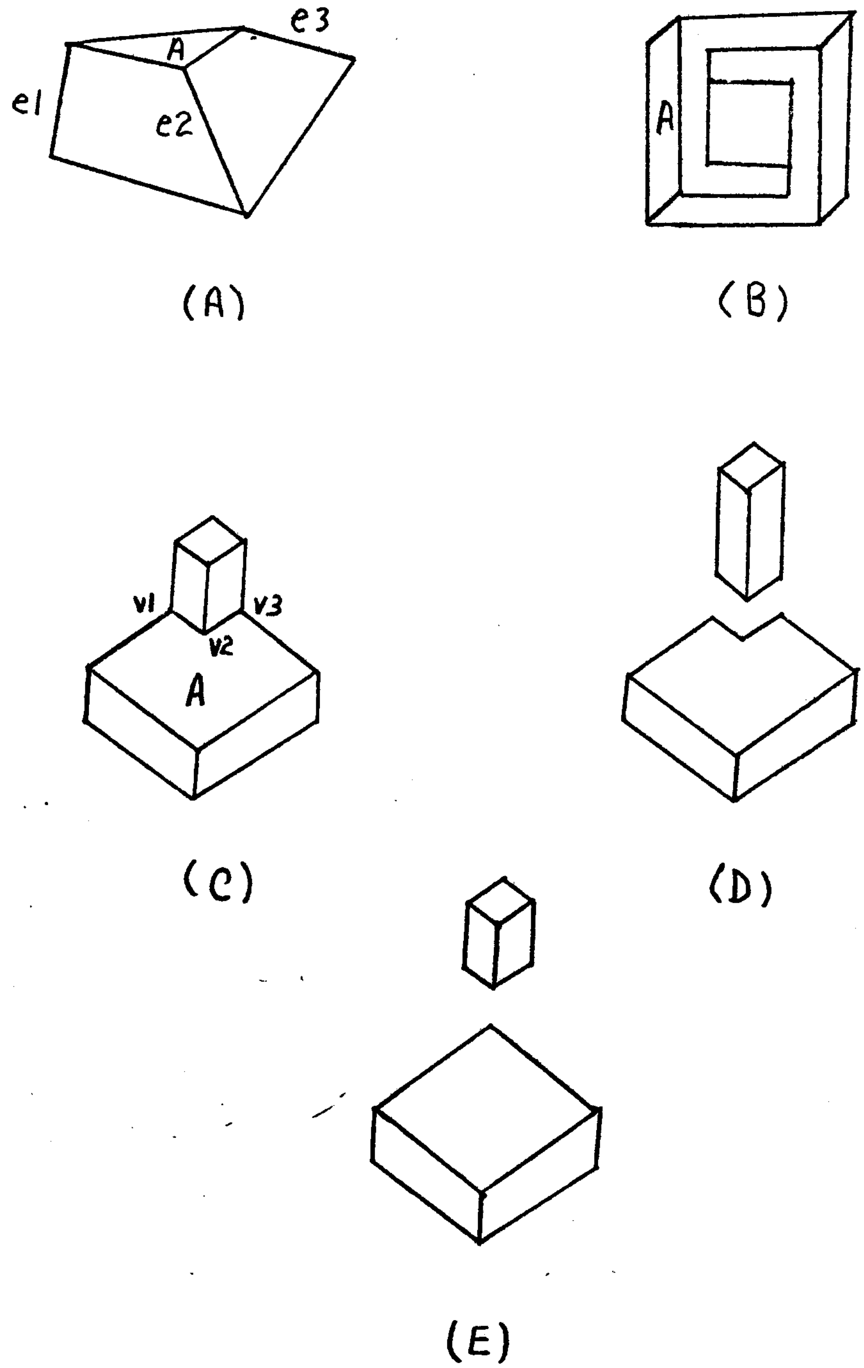


Fig. 4.

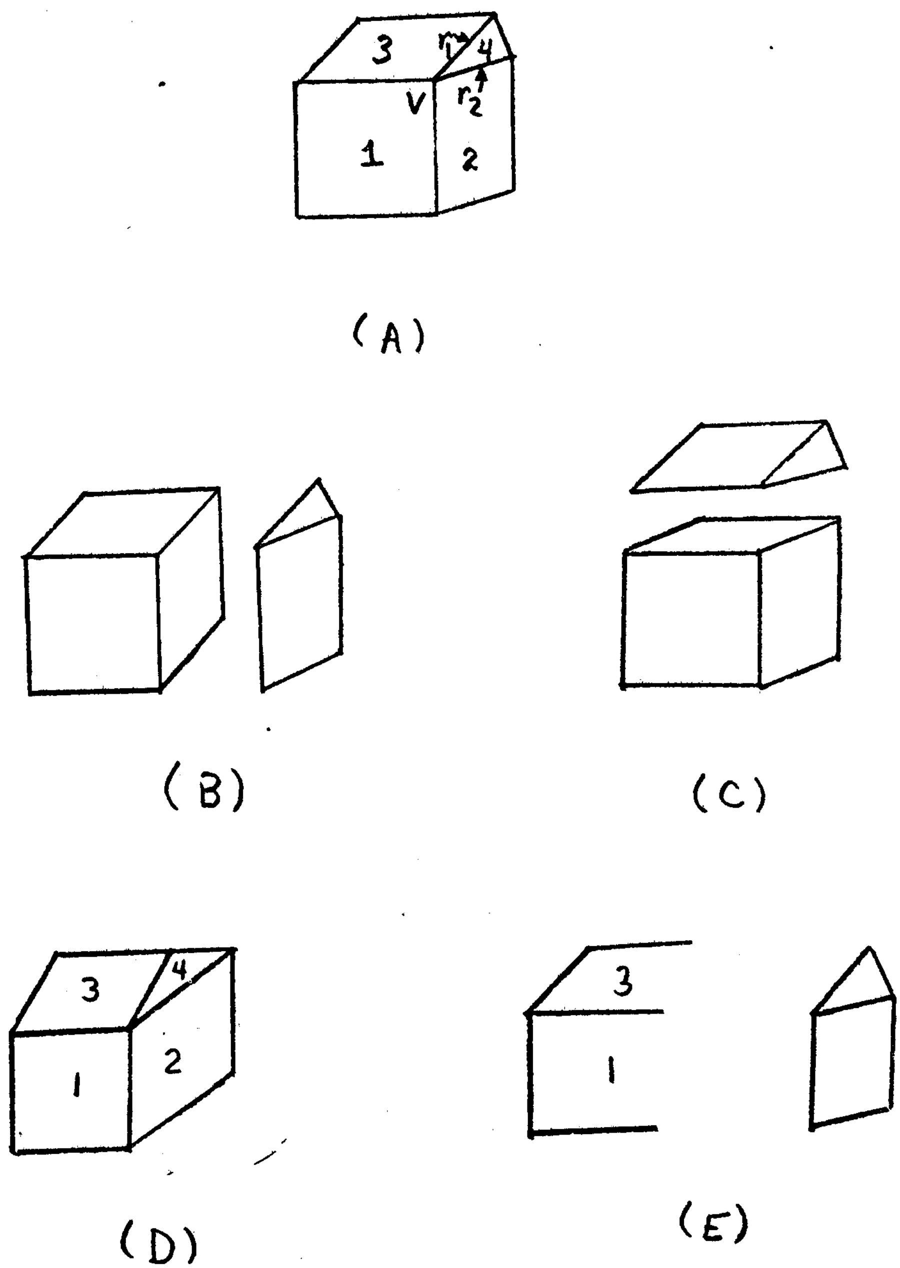


Fig. 5. Variations on a wedge and cube.

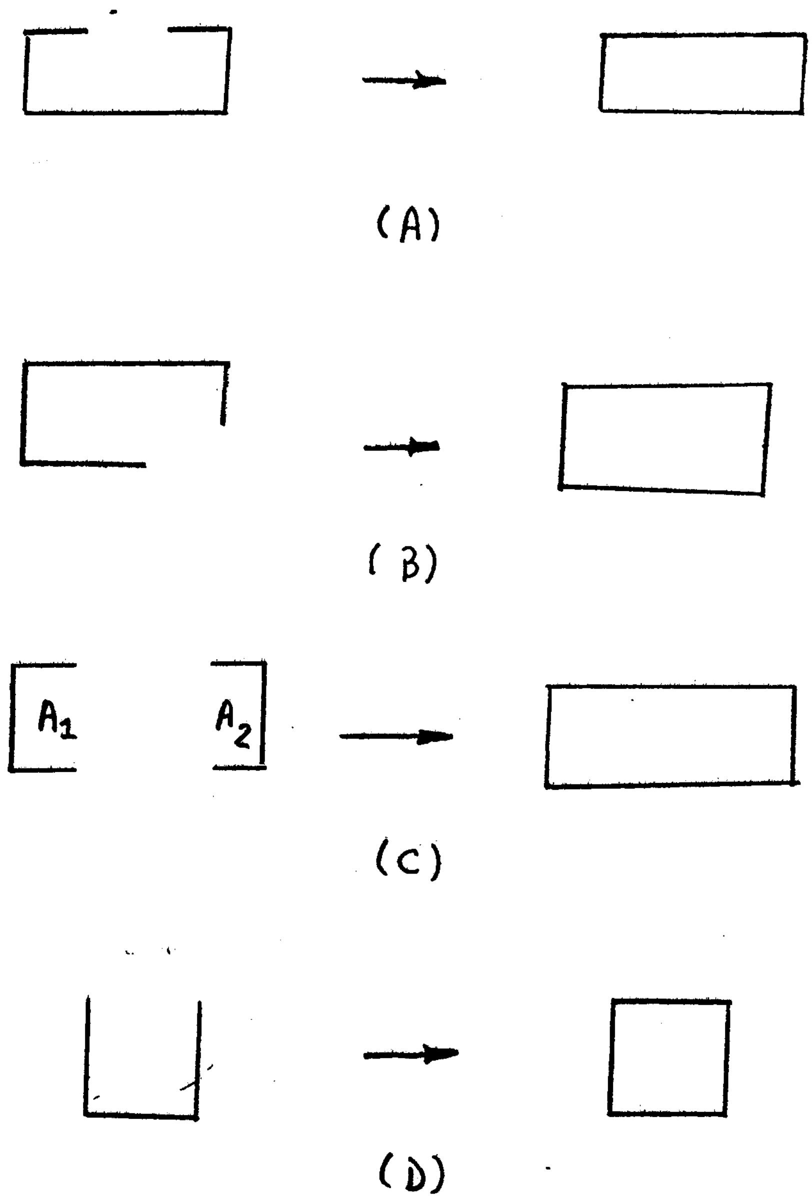


Fig. 6. Reconstruction Rules.

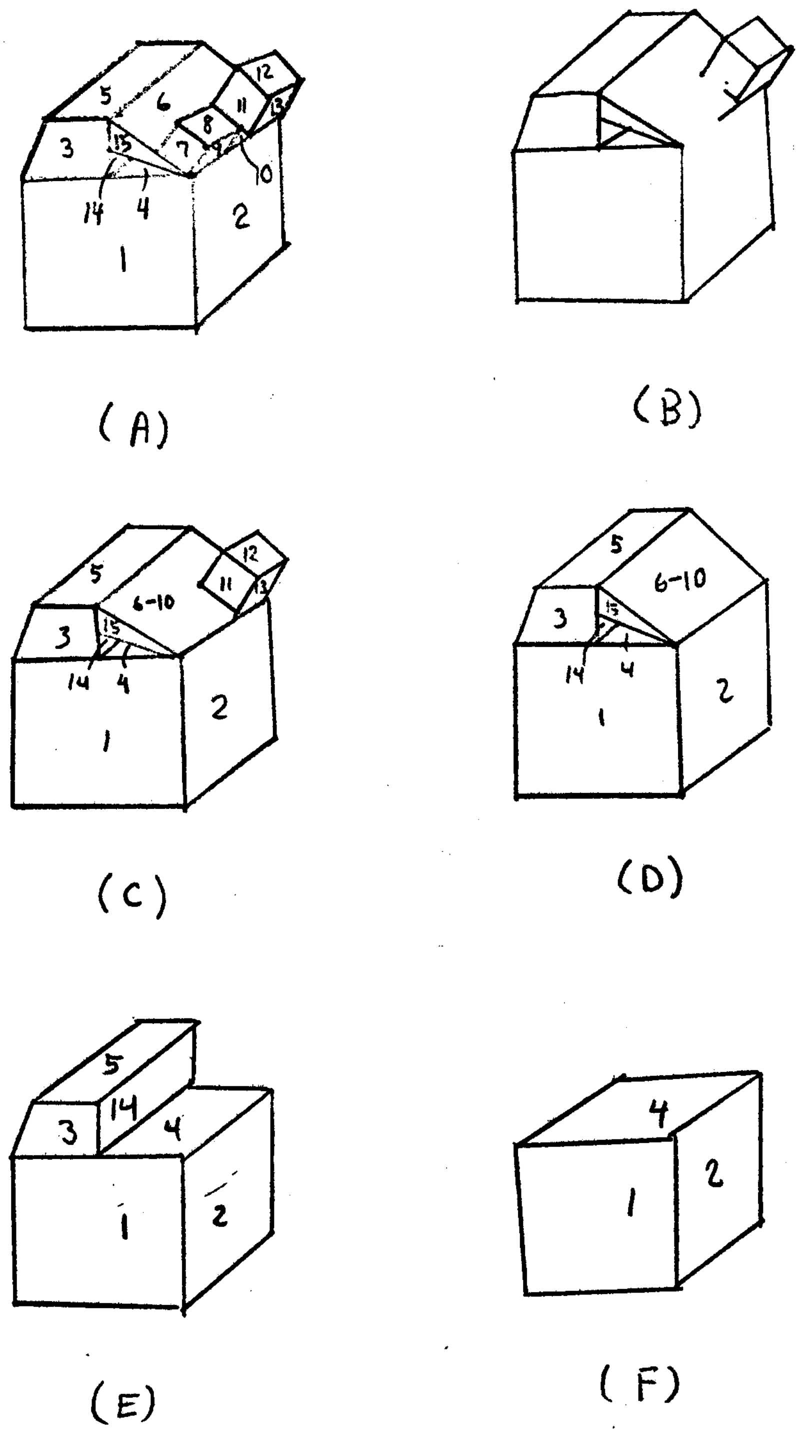

5
6
12
11
3
15
7
8
9
13
10
14
4
2
1
(A)
(B)
5
12
6-10
11
13
3
15
14
4
2
1
(C)
5
6-10
3
15
14
4
2
1
(D)
5
3
14
4
1
2
(E)
4
2
1
(F)

Fig. 7. PAPA.

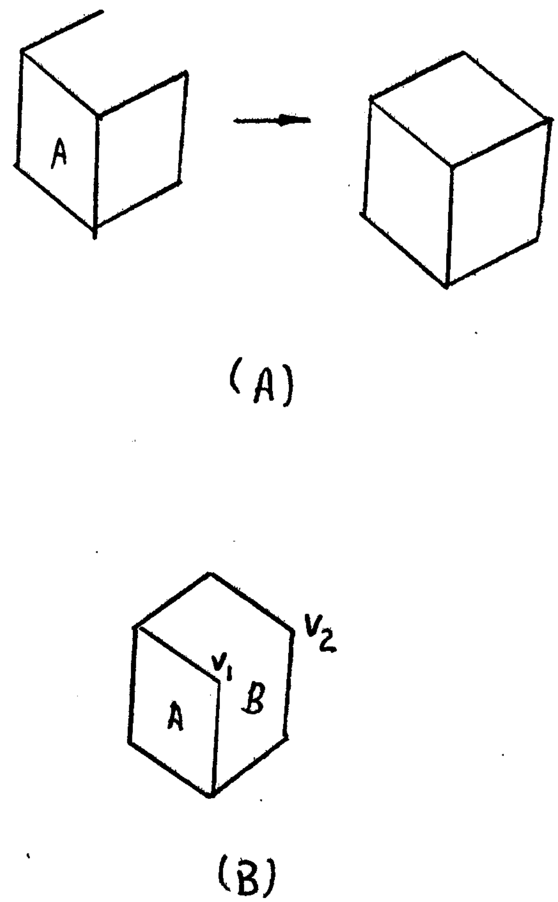
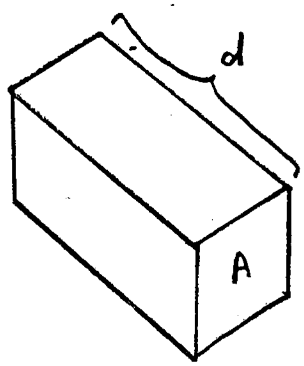
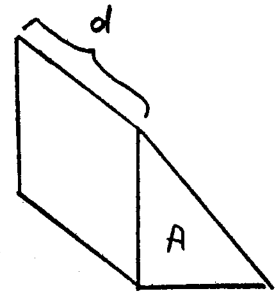


Fig. 8. Incomplete cubes.



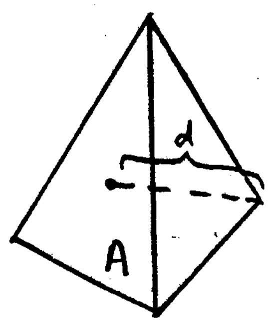
Block: project rectangle A
an arbitrary distance d.

(A)



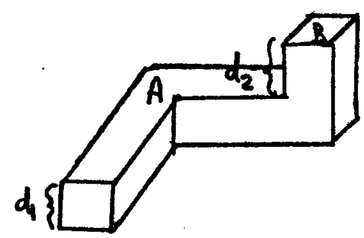
Wedge: project triangle A
an arbitrary distance d.

(B)



Pyramid: project triangle A
a distance d with scale
change $\frac{d-x}{d}$, where
x = distance of projection.

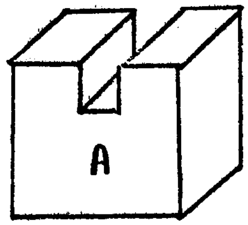
(c)



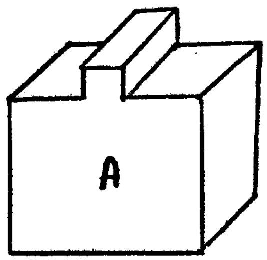
Squiggit: L-shaped object
(project A a distance d_1)
with a cube (project B
a distance d_2) on it.

(D)

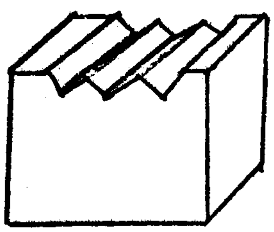
Fig. 9. Some models.



(A)



(B)



(C)

Fig. 10. Some approximate cubes.

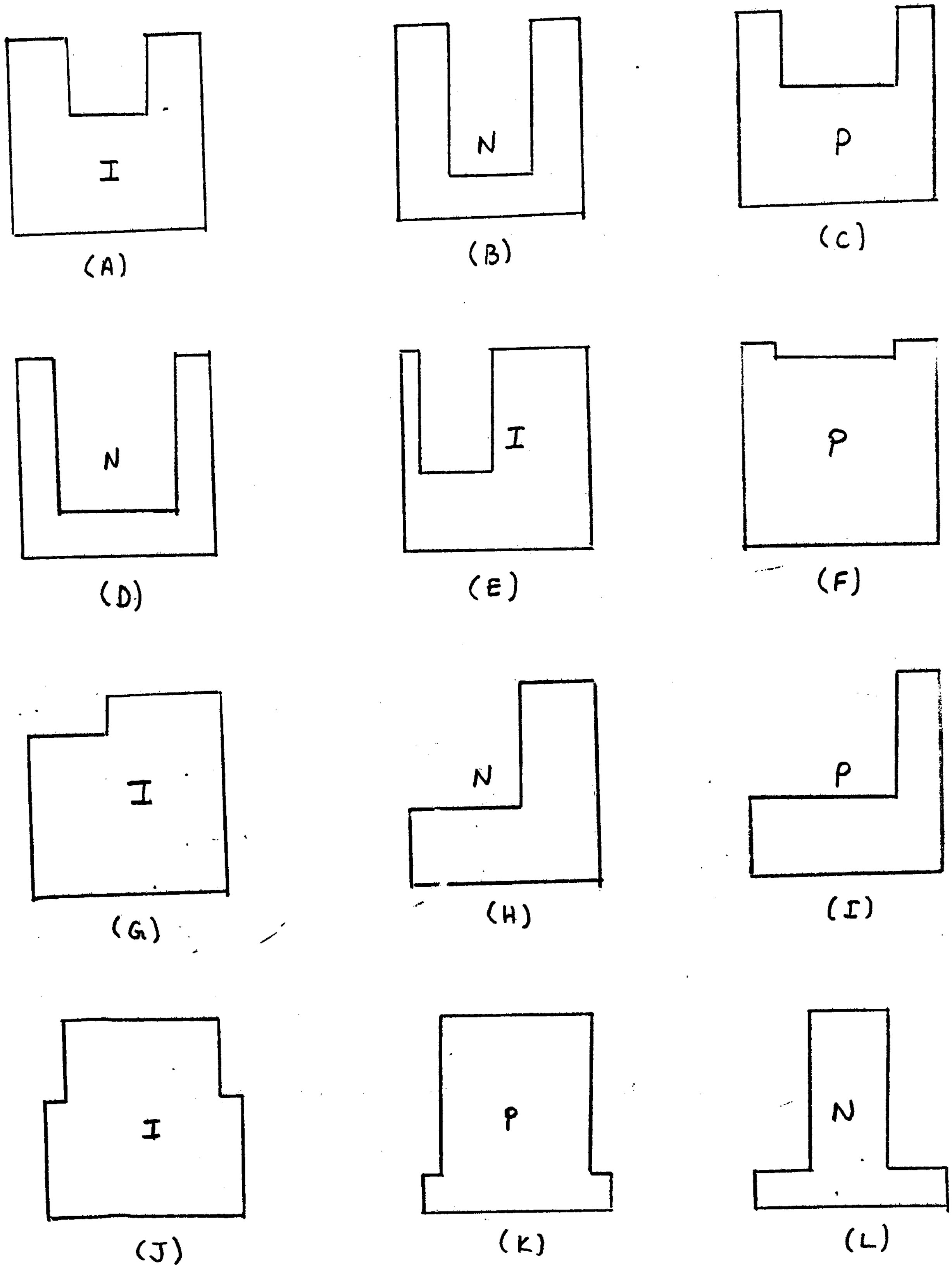


Fig. 11.

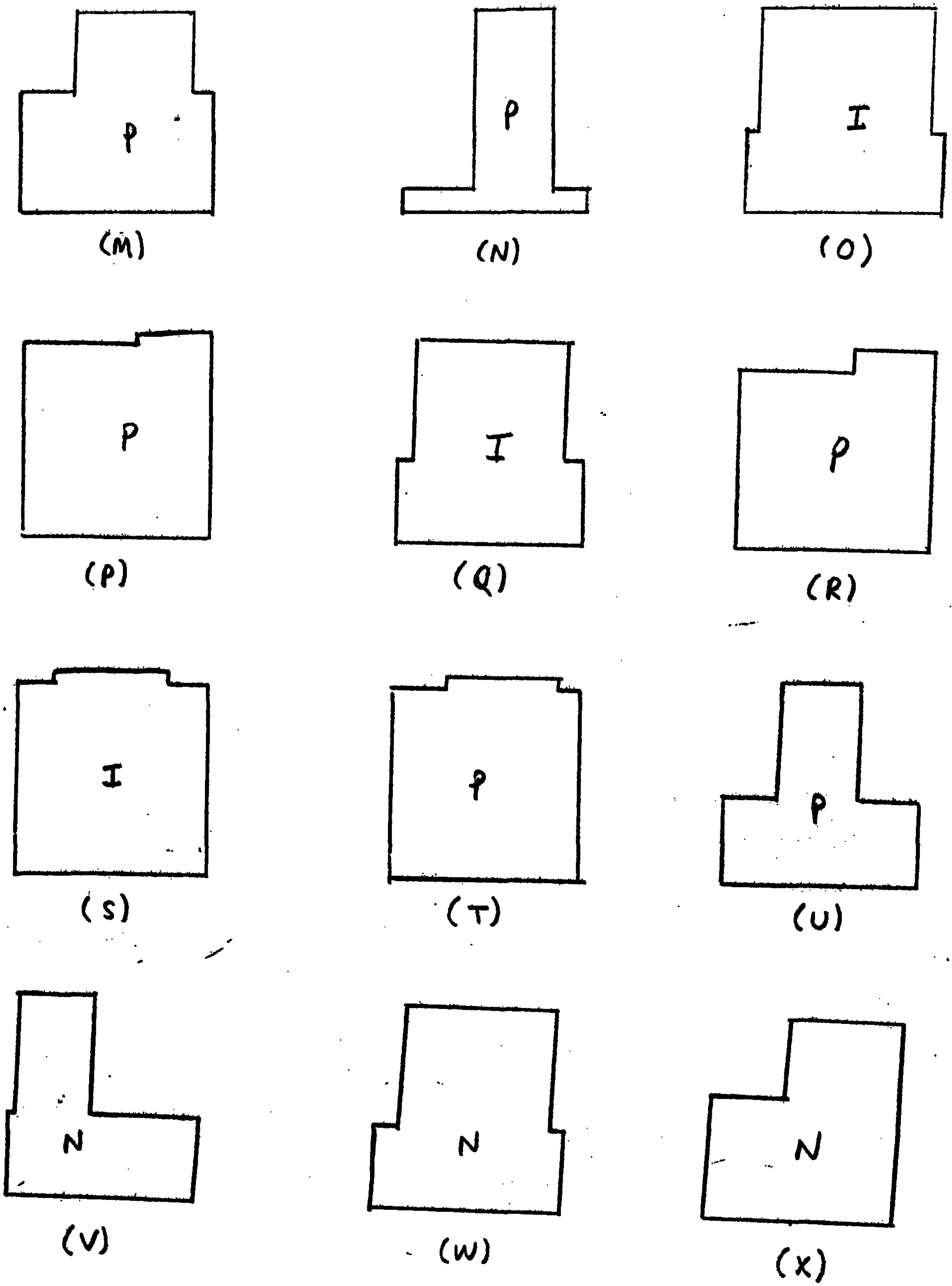


Fig. 11 (cont.).

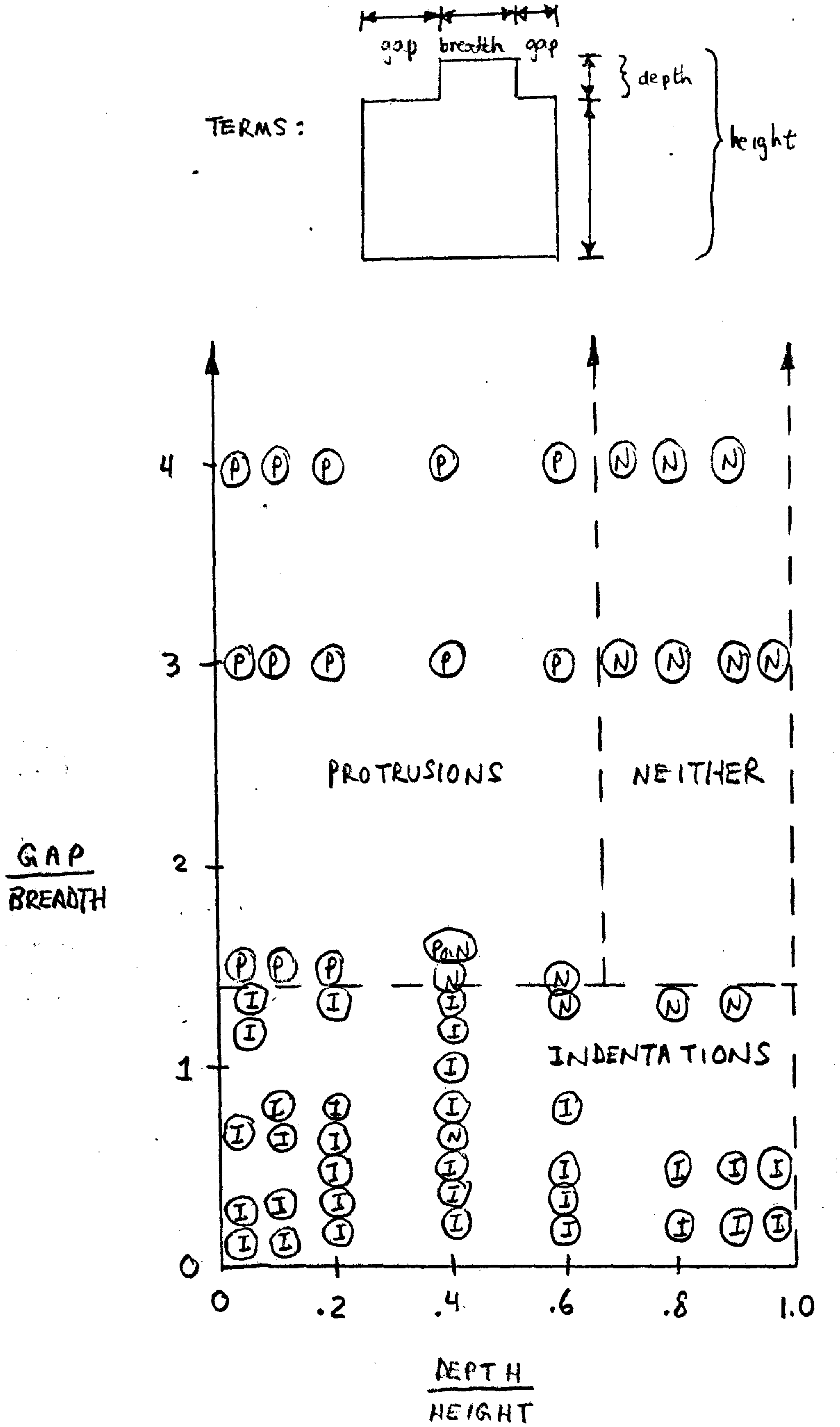

gap
breadth
gap
depth
height
TERMS:
PROTRUSIONS
NEITHER
INDENTATIONS
GAP / BREADTH
DEPTH / HEIGHT
0
1
2
3
4
.2
.4
.6
.8
1.0

Fig. 12

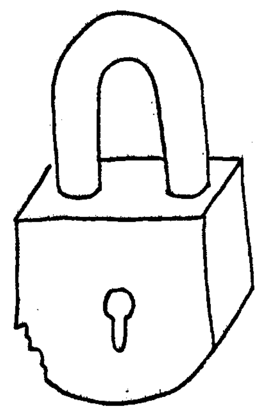


Fig. 13. Padlock.

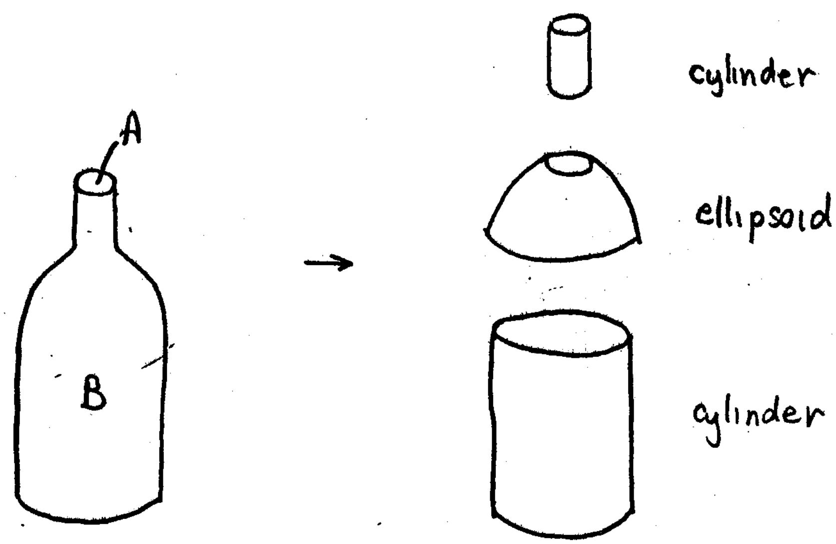


Fig. 15. Bottle

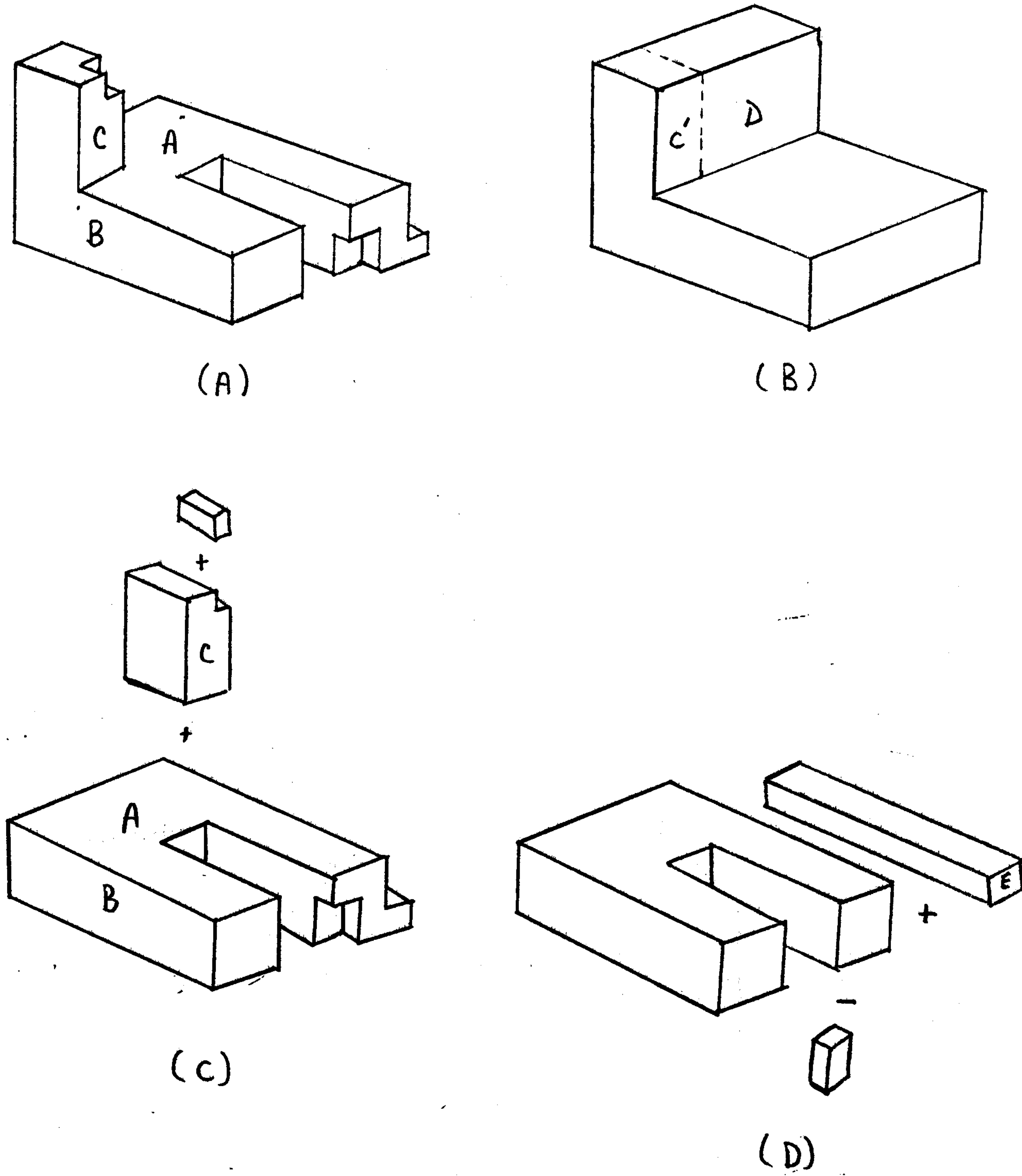


Fig. 14.