A THEORY OF ORIGAMI WORLD

Takeo Kanade

Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pa. 15213, USA
and
Department of Information Science
Kyoto University
Kyoto, JAPAN

September 20, 1978

Abstract

The recovery of three-dimensional configurations of a scene from its image is one of the most important steps in computer vision. The Origami world is a model for understanding line drawings in terms of surfaces, and for finding their 3-D configurations. It assumes that surfaces themselves can be stand-alone objects, unlike conventional models, such as the trihedral world, which assume solid objects. We have established a labeling procedure for this Origami world, which can find the 3-D meaning of a given line drawing by assigning one of the four labels, + (convex edge), - (concave edge), ←, and → (occluding boundary) to each line. The procedure uses a filtering procedure not only for junction labels as in the Waltz labeling for the trihedral world, but also for checking the consistency of surface orientations. The theory includes the Huffman-Clowes-Waltz labelings for the trihedral solid-object world as a subset. It shows great potential for the application of recovering 3-D configurations from region-segmented images; other information (such as spectral information) available from images can also be incorporated smoothly. This paper also reveals interesting relationships among previous research in polyhedral scene analysis.

This research was supported by the Defense Advanced Research Project Agency under contract no. F33615-78-C-1551 and monitored by the Air Force Office of Scientific Research.

I Introduction

Origami is the Japanese traditional manual art of making various shaped objects (e.g., animals) by folding a sheet of paper. Figure 1 is a typical example of Origami. It is easy to see that Figure 1 is an Origami crane. This process of seeing and understanding may be divided into two processes: one is to determine the possible three-dimensional configurations from the picture, and the other is to match them with some known concepts (such as "crane"). This paper deals with the first process. Thus the problem is: how do we understand the possible three-dimensional configurations from a collection of lines?

One solution is: first, model a world (I will call it the "Origami" world), where surfaces themselves can be stand-alone objects, rather than the conventional trihedral solid-object world; secondly, establish a procedure which can assign a 3-D meaning to each line. The procedure developed uses a filtering method both for finding consistent combinations of labels and for testing the consistency of surface orientations based on the gradient space representation. Not only does this surface-oriented Origami world include the cases of the solid-object world, studied by Huffman [Huffman, 1971], Clowes [Clowes, 1971], and Waltz [Waltz, 1972], as a subset, but it also demonstrates various features that have the potential to be used in image understanding tasks of real-world images.

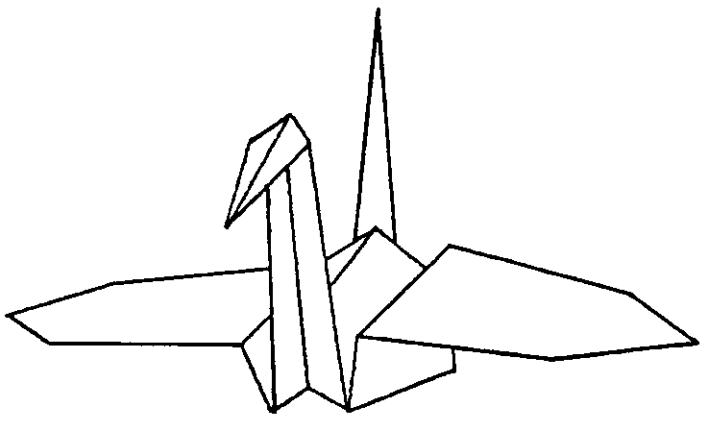


Figure 1 Origami crane.

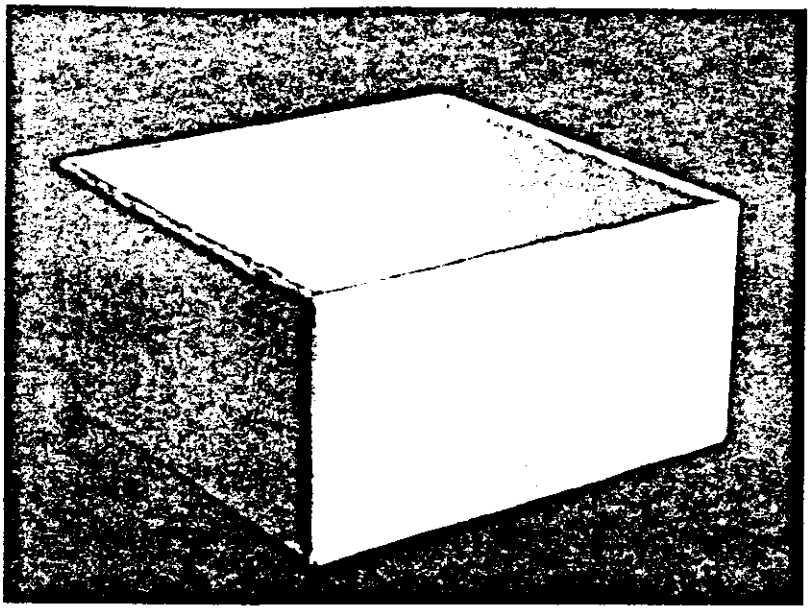


Figure 2 Photograph of a carton paper box.

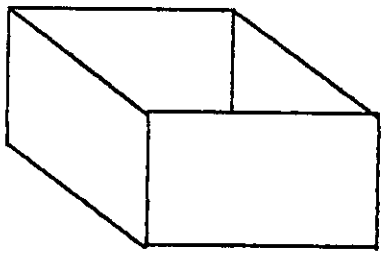


Figure 3 Line drawing of a carton paper box.

II Key Ideas and Related Work

Illustrative Examples

Let us have several illustrative examples of simple line drawings for the following discussions. Suppose that an image of a box case made of carton paper (Figure 2) is given. How do we recognize that the object in the image has a shape of "box" (i.e., an open-faced cube)? A line drawing derived from the image such as Figure 3 has long been an important product of the initial feature extraction process. In fact, we can imagine the three-dimensional shape of "box" from Figure 3. As other examples, the drawings in Figure 4 usually convey to the viewer the meaning intended by the artist: (a) a cube, (b) a W-folded paper, and (c) two coordinate planes intersected. However, take Figure 4(a) for example: other possible configurations, such as those in Figure 5, are imaginable.

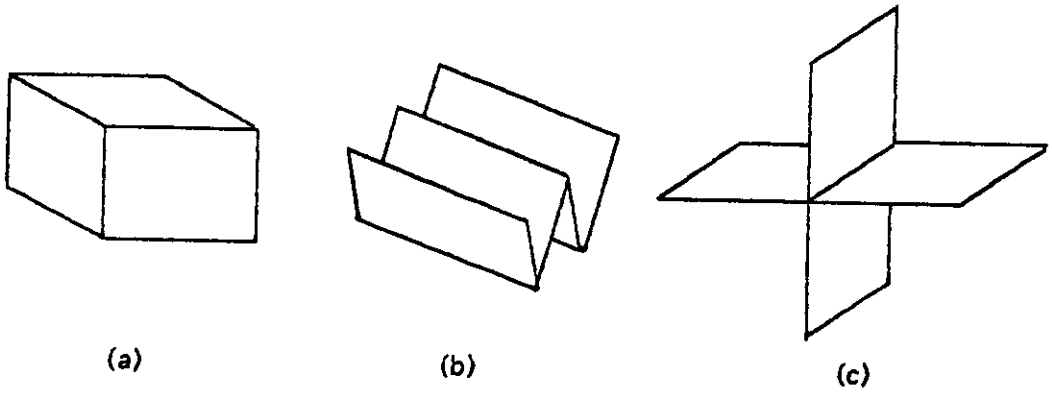


Figure 4 Simple line drawings.

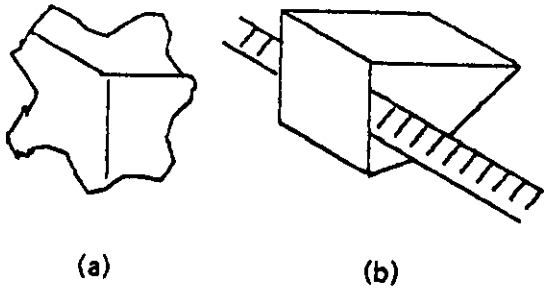


Figure 5 Other interpretations of Figure 4(a): (a) a concave corner; (b) a folded paper "roof" covering another paper.

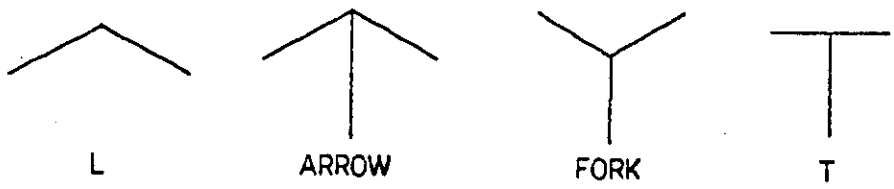


Figure 8 Junction types treated in this paper.

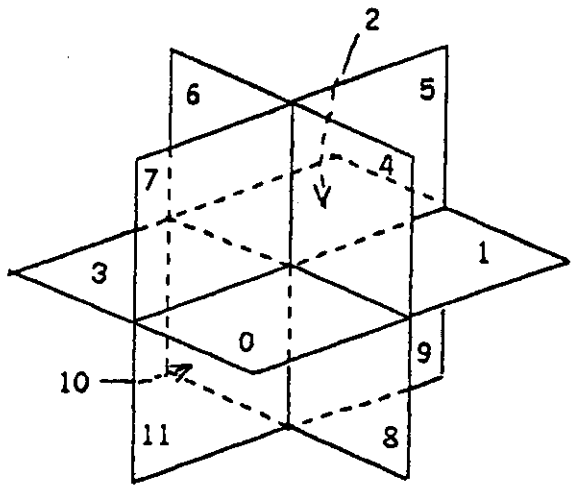


Figure 9 Twelve quadrant planes.

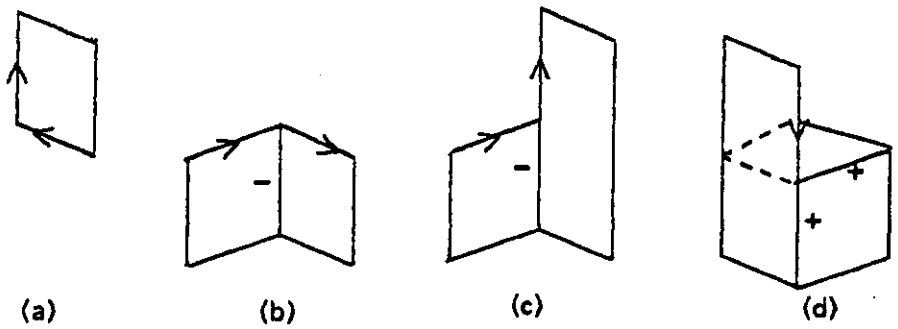


Figure 10 Examples of configurations at vertices: (a) one quadrant plane which generates an L junction; (b) two quadrant planes which generate an ARROW junction; (c) three quadrant planes which generate a T junction; (d) four quadrant planes which generate a T junction.

III The Theory of Origami World

The presentation of the theory of the Origami world consists of seven subsections:

(1) Terminology;
(2) The enumeration of legal combinations of line labels at junctions;
(3) The links between regions;
(4) The problem concerning consistency of surface orientations;
(5) A test method for consistency of surface orientations;
(6) The actual labeling procedure;
(7) Examples of labeling.

III-1 Origami World and Terminology

The world is assumed to be made of a collection of surfaces. A line drawing is a *picture* (orthographic projection) of such a composite in the *scene*. For the time being the surface in our Origami world are assumed to be planar; i.e., the orientation is constant throughout a surface (actually, the restriction to plane surfaces will be relaxed a little in the later sections). In this respect it is *not* the paper-surface (i.e., developable surface) world investigated in [Huffman, 1976].

The terminology we will use for the Origami world parallels that for the Waltz labeling theory for the trihedral world [Waltz, 1972]. An *edge* is a straight boundary of a plane surface. A *vertex* is a point where edges of the surface(s) meet. A *line* is an orthographic projection of an edge to the picture plane. A *junction* is a point in the picture where lines meet. A junction can be the projection of a vertex or the point where an edge is interrupted by an occluding surface. A *region* is an area in the picture surrounded by lines, and it corresponds to (a visible part of) a surface.

An edge can be classified according to its three-dimensional physical meaning in the scene. We will use the following terms and labels:

convex	+	: edge along which two surfaces meet and form a convexity
concave	-	: edge along which two surfaces meet and form a concavity
occluding	← or →	: edge along which one surface occludes another

The directions of arrows of occluding boundaries are given in such a way that the occluding surface is on their right hand side. A line can therefore be labeled with one of the four labels (+, -, ←, and →), according to its physical meaning of the corresponding edge; thus we can say "a line is convex" to mean that it depicts a convex edge. Therefore, to give line

labels to the lines in the drawing is to give a three-dimensional meaning to the drawing. A set of assignments of line labels to the lines in the drawing is called an *interpretation* of the drawing. For example, the labeling shown in Figure 6 is an interpretation of Figure 4(a).

Junctions are classified according to the number of lines meeting at the junctions and their geometrical configurations in the picture. In this paper we will confine ourselves to L, ARROW, FORK and T junctions shown in Figure 8.

III-2 The Enumeration of Legal Junction Labels

The physical world imposes constraints on the labels that lines can take at a particular type of junction. A combination of line labels for one junction type is referred to as a *junction label*. The crucial observation which was made by Huffman and Clowes, and which was exploited to a great extent by Waltz, is that not all the combinatorially possible junction labels can appear (are *legal*) in the picture. For example, for the ARROW junction, only three junction labels out of the 4x4x4 possible combinations can occur in the trihedral world. Needless to say, unless we assume a certain restriction on the three-dimensional configurations allowable at the vertices, the resultant constraints on junction labels will be too weak to be useful. We need to confine ourselves to a reasonably limited world which corresponds well to the real world images.

The confinement we adopt in the Origami world is that surfaces meet edge to edge, that *no more than* three surfaces of different orientations can meet at a vertex, and that the combination of the three orientations is "general", in the sense that they span the three-dimensional space (i.e., each orientation has a vector component perpendicular to the other two). Thus, no more than three edges of different directions are involved at a vertex. Let us call such vertices *up-to-3-surface* vertices. This restriction corresponds to the trihedral vertices in the solid-object world. Note, however, that the up-to-3-surface vertices generate a richer world than the world generated by the trihedral vertices, since the former can include 1- and 2- surface vertices; that is, it allows free extending surfaces as stand-alone objects.

Possible junction labels for the up-to-3-surface vertices in the Origami world can be enumerated in the following way. The planes of three general orientations intersect and divide each other into 12 partial planes. Thus we can think of 12 quadrant plane surfaces around the vertex point as shown in Figure 9.

Let us fix our eye position in one of the eight octants separated by the quadrant planes, say, the octant bounded by the quadrants 0,4, and 7. Next, we generate one by one all the possible (4096) combinations by setting each quadrant plane to be either occupied or vacant, and check how the vertex formed at the origin appears when viewed from the eye

position fixed as above. Then we can give a label to each line at the junction based on its meaning, and obtain a legal junction label. Figures 10(a) through 10(d) show examples of the vertex configurations and their derived junction labels. As previously stated, we consider only the combinations which result in the junction types shown in Figure 8. The number of junction labels thus obtained is: 8 for L, 15 for ARROW, 9 for FORK, and 12 for T.

For the junction type T, the four additional junction labels shown in Figure 11 are included as legal. They do not correspond to actual vertices, but to the cases in which the junction is caused because the upper half plane is in the front and occludes the edge behind. Table 1 compares the number of legal up-to-3-surface junction labels thus obtained with that of legal trihedral junction labels. It gives an idea of the degree of constraint imposed by the up-to-3-surface Origami world compared with the Huffman-Clowes trihedral-junction world. The appendix gives a complete list of legal junction labels in the Origami world; for each junction label, it includes an illustrative figure of the configuration which the label represents, and the *links* which will be explained next.

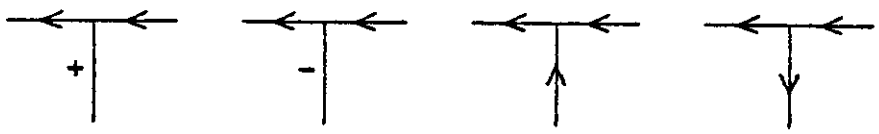


Figure 11 Legal T junction labels not corresponding to vertices.

Table 1. Comparison of the size of the Origami junction dictionary with the Huffman-Clowes dictionary.

Junction Type	Huffman-Clowes Dictionary	Origami World Dictionary
L	6	8
ARROW	3	15
FORK	3	9
T	4	16

III-3 The Links

Each junction label derived in the previous subsection implies which surfaces are connected at which edge in order to form that junction label. For later use, this information is also stored explicitly in the dictionary by means of a *link*, which links a pair of connected regions and points to the line at which they intersect. For example, the link in a legal FORK junction label shown in Figure 12 represents that the regions R_1 and R_2 are connected at the convex line L. Note that since the region R_3 is totally occluded by the other two regions (in other words, it is the background), it has no relationship to others at this junction.

In the case of junction labels which involve partially occluded regions, a subtle situation occurs. Take the ARROW junction label shown in Figure 13(a) as an example. This junction label was originally derived from the configuration shown in Figure 13(b); the surfaces S_1 and S_2 connect at the edge BC. However, note that the junction label itself can mean other cases such as those shown in Figure 13(c) and 13(d): (c) is the case where S_1 and S_2 intersect within the angle ABC, and (d) is the case where S_1 and S_2 will intersect outside of the angle ABC, when they are extended.

In the Origami world we will assume that the situation shown in Figure 13(c) is what is happening near the vertex. This assumption allows more configurations than assuming merely the case of Figure 13(b). It seems reasonable to exclude those situations like Figure 13(d), because they are accidental cases caused by a particular relationship between the view direction and the vertex. In fact, if we move our view direction a little to the left, the vertex of Figure 13(d) may appear like Figure 13(e), even though we are looking at the same sides of the same surfaces.

Therefore, the link for the junction label of Figure 13(a) is given as shown in Figure 13(f); it represents that the region R_1 and R_2 are connected at an "occluded intersection" line L' (its label is ⊛), which is located within the angle ABC. Note that the line L' can overlap

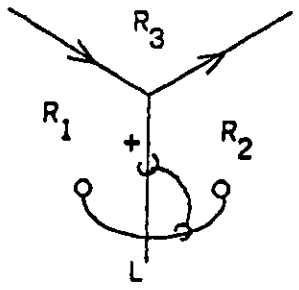


Figure 12 Link between regions.

with BC, but not with AB, because if it did, the label of AB will change from ← to +. In this way, appropriate link information is added to each junction label in the dictionary as shown in the appendix.

III-4 Global Consistency of Surface Orientations

The Problem

Let us take the line drawing of Figure 14(a) as an illustrative example. Consider an interpretation shown in Figure 14(b), which implies that two surfaces corresponding to the regions R_1 (ABED) and R_2 (ACBEFD) are connected at two convex edges AD and BE. All the junctions, A through F, are given legal junction labels. However, it can be easily seen that this configuration is not realizable by the two planar surfaces; one or the other should be bent. Why is this inconsistency not detected by the junction dictionary? The reason is shown in Figure 14(c). The junction dictionary tells that the line labels given to AD, BE, AB, BC, and CA are consistent because they form legal combinations *locally* at each junction A, B, or C. However, no use is made of the fact that the three shaded regions in Figure 14(c) are

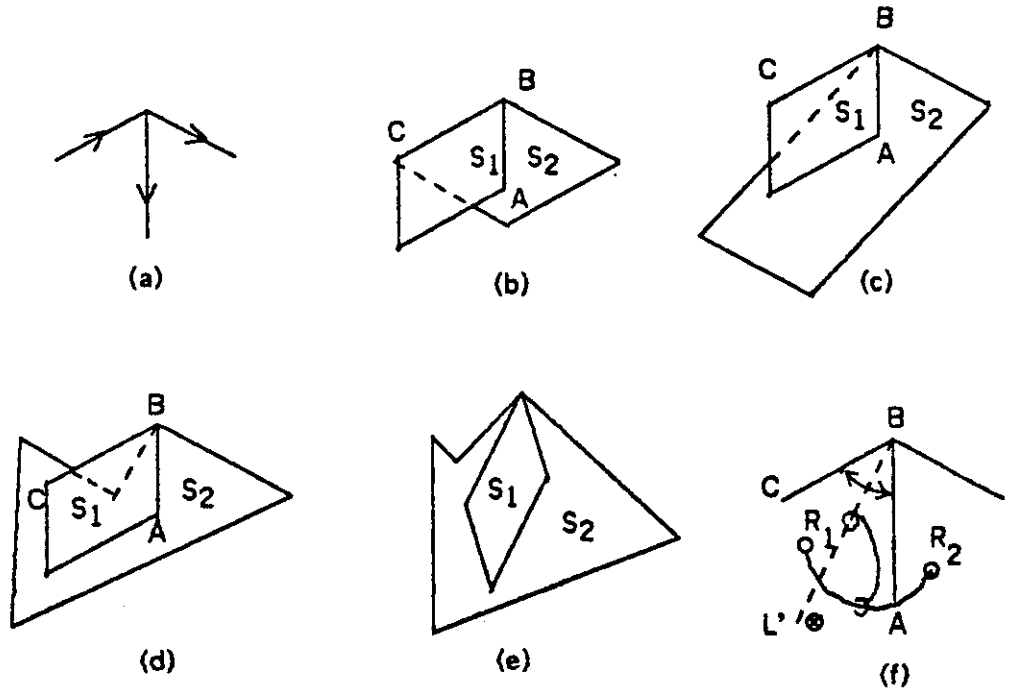


Figure 13 Link for an occluded intersection line.

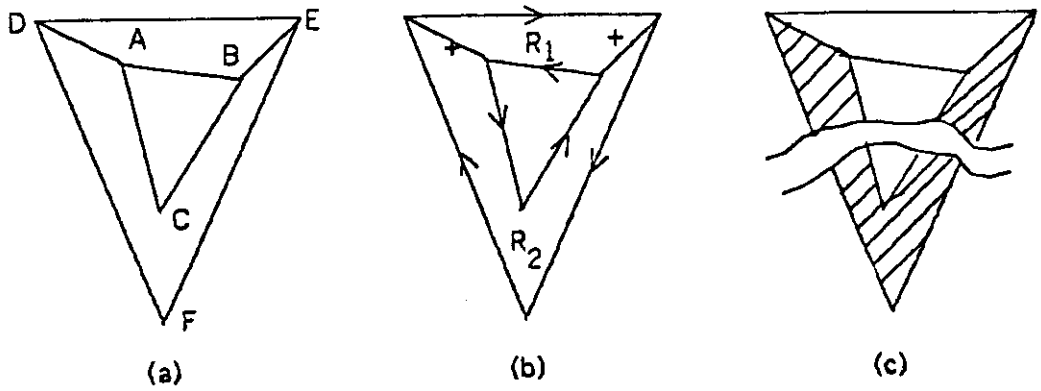


Figure 14 Interpretation inconsistent with respect to surface orientations.

actually the *same* surface and therefore have a *unique* surface orientation in the scene. This example demonstrates that we need a provision to check such global consistency of surface orientations.

It should be noted that the kinds of anomalies illustrated above, which are caused by relying solely upon the junction dictionary, have also occurred in the Huffman-Clowes-Waltz labeling for the trihedral solid-object world. But because they occured "less frequently", they did not show up as a very serious problem. Figure 15 is an example of such an anomaly shown in [Mackworth, 1977]. All the junction labels in it are legal in the trihedral world, but it can be seen that the configuration is not realizable in that world.

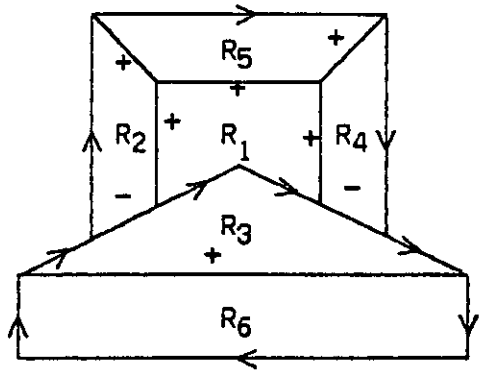


Figure 15 Anomaly in the Huffman-Clowes-Waltz labeling.

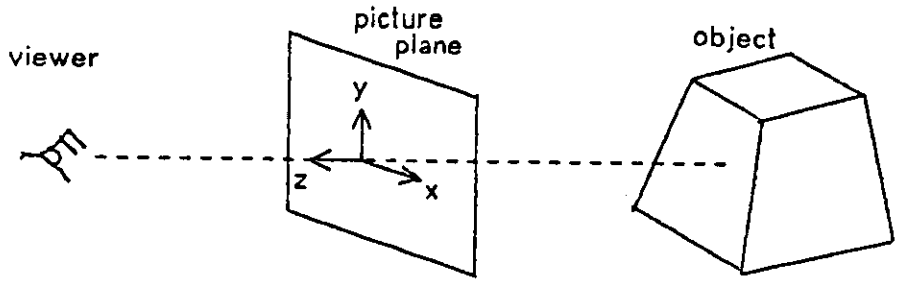


Figure 16 Geometry involving the viewer, picture plane, and object.

A Tool

In order to carry out consistency checks of surface orientations it is necessary to represent surface orientations in the scene in connection with their properties projected onto the picture. The *gradient space* introduced by Mackworth [Mackworth, 1973] provides a good tool for it.

Let Figure 16 be the geometry involving the viewer, the picture plane, and the object in the scene. A plane in the scene whose surface is visible from the viewer can be expressed as

$$-z = ax + by + c. \tag{1}$$

The two-dimensional space made of the ordered pairs (a,b) is called the *gradient space* G. Let us assume for our convenience that we align the directions of the coordinates of (x,y) in the picture with those of (a,b). All planes in the scene which have the same values of a and b are mapped into the point (a,b), called the *gradient*, in G.

The values of (a,b) represent how the planes are slanting relative to the view line (z axis). For example, the origin O_G=(0,0) of G corresponds to those planes (-z=c) perpendicular to the view line. P_1=(1,0) corresponds to the planes (-z=x+c) which are slanting horizontally to the right. Mathematically,

$$a = \partial(-z)/\partial x, \quad b = \partial(-z)/\partial y, \tag{2}$$

which is why (a,b) is called the gradient. Thus the length $\sqrt{a^2+b^2}$ of the vector from O_G=(0,0) to P=(a,b) is the tangent of the angle between the picture plane and the planes corresponding to P; and the direction $\tan^{-1}(b/a)$ of the vector is the direction of the steepest change of -z (depth) on the plane.

One of the useful properties of the gradient space is the following [Mackworth, 1973]. Consider two planes meeting at an edge and the orthographic picture made of regions R_1 and R_2 and a line L, as shown in Figure 17. Then in the gradient space, the gradients G_1 and G_2 of the two planes should be on a line which is perpendicular to the picture line L. Moreover, if the edge is convex (+), G_1 and G_2 are ordered in the same direction as are the corresponding regions in the picture. If the edge is concave (-), their order is reversed.

In the Origami world, we additionally have the case of an occluded intersection (⊗) as shown in Figure 18. This case is treated similarly. If we remove the part of R_1 which is occluding R_2 below L, the rest of R_1 and R_2 will form a convex line at L, and L can be anywhere in the angle ABC. Therefore, as shown in Figure 18, the gradient G_2 should be inside of the fan-shaped area whose origin is at G_1, and is bounded by the lines which are perpendicular to AB and BC.

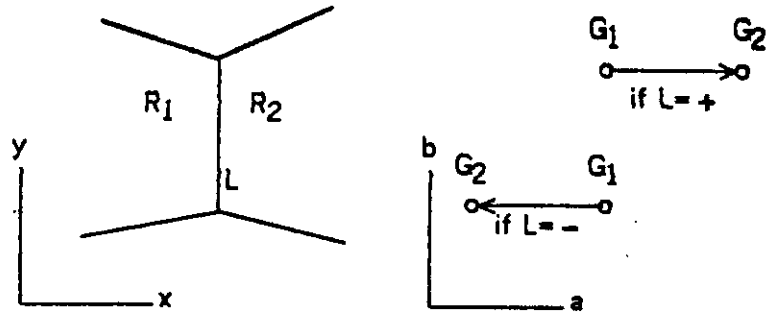


Figure 17 Constraint on the gradients of regions connected by a convex (+) or concave (-) line.

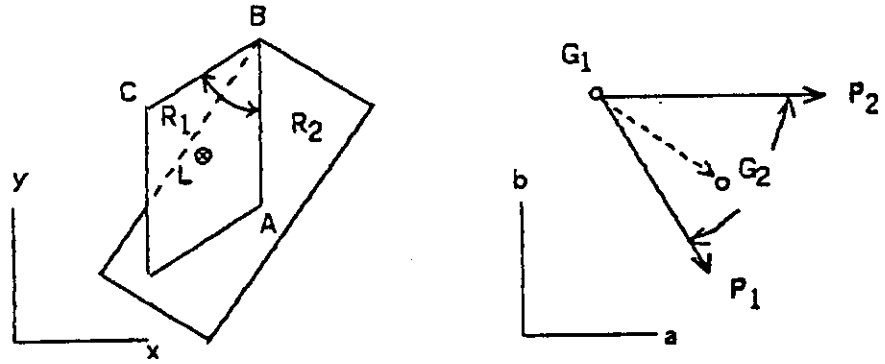


Figure 18 Constraint on the gradients of regions connected by an occluded intersection (⊗) line.

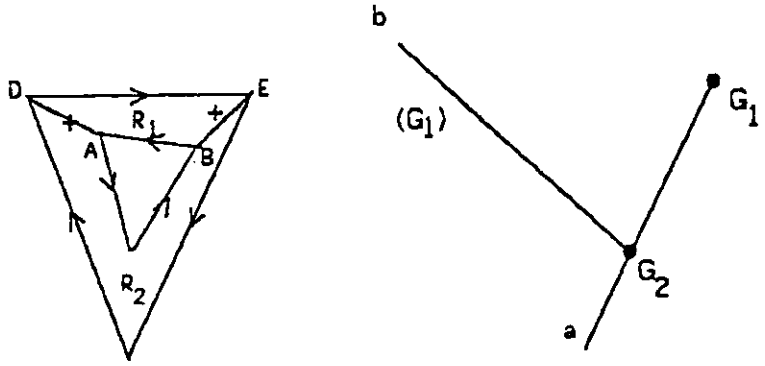


Figure 19 Trace of gradients of the regions in Figure 14.

An Example of Nonconsistency of Surface Orientations

Now we are ready to test the example of Figure 14. Imagine a gradient space and refer to Figure 19. Let G_1 be the gradient of the region R_1. R_1 and R_2 are connected at the convex line AD in the picture. Thus, the gradient of R_2 should be somewhere on the half line G_1a, which is perpendicular to AD and extends toward left, because of the property of the gradient space. Suppose it is at G_2. Again because R_2 is connected with R_1 at a convex line BE, the gradient of R_1 should be somewhere on the half line G_2b as shown. Since we have fixed the gradient of R_1 at G_1, this half line G_2b should pass G_1, which is impossible wherever we select G_2 on the half line G_1a. This means that there is no combination of gradients for the regions R_1 and R_2 which results in the configuration of Figure 14(b); therefore the configuration is inconsistent.

III-5 The Test Procedure in the Origami World

The above example has demonstrated the necessity of and the method for global consistency checks of surface orientations for a set of regions. This section will present an algorithm which indicates on what sets of regions the consistency checks are to be performed and which tells whether they can have consistent surface orientations. Given an interpretation, the method consists of first constructing a labeled graph called a *Surface Connection Graph*, and then performing a type of filtering operation on the constraints in the gradient space. The test procedure to be presented is closely related to the idea of the dual graph of Huffman [Huffman, 1971] and the POLY program of Mackworth [Mackworth, 1973]. In fact, the Surface Connection Graph represents by and large the topological properties of the dual graph, and the filtering procedure uses constraints in the gradient space in a more thorough and systematic way than does POLY.

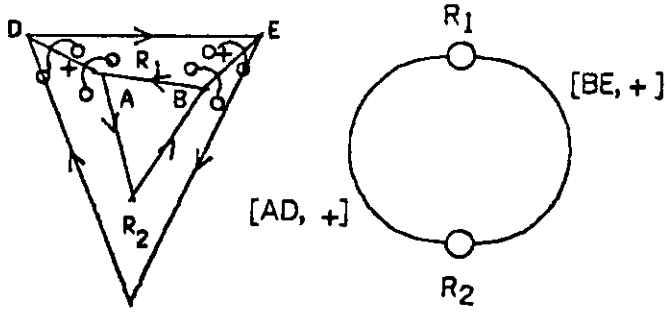


Figure 20 Surface Connection Graph of the interpretation in Figure 14 (b).

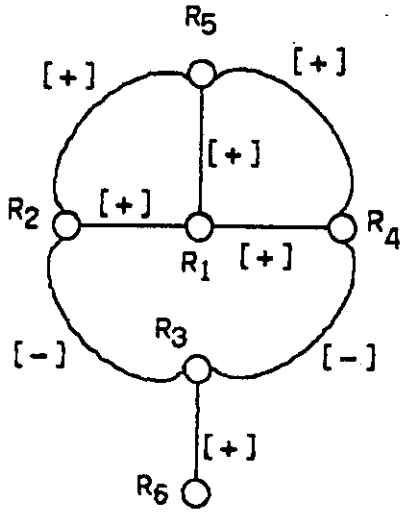


Figure 21 Surface Connection Graph of the interpretation in Figure 15.

Surface Connection Graph (SCG)

A *Surface Connection Graph (SCG)* is a labeled graph constructed in the following way from *a given interpretation* of the line drawing. The nodes of the SCG are regions in the drawing. A pair of nodes are connected by an arc if a junction involving the corresponding pair of regions has been given a junction label which has a link between them. To each arc is attached the corresponding link information, which includes the kind of intersection line (+, - or ⊛) and its position. A pair of regions can be connected by multiple arcs if there are several junctions having the link between them. Because a link is given between a pair of regions when they intersect at a junction, the SCG represents which surfaces are connected with which surfaces at which edges. As examples, the SCG's for Figure 14(b) and Figure 15 are shown in Figure 20 and Figure 21, respectively. Duplication of the same link is eliminated in the SCG, such as the links between R_1 and R_2 at the vertices A and D in Figure 20.

Spanning Angles and their Computation

Let G_1 and G_2 denote the gradients of R_1 and R_2, respectively. Suppose the nodes R_1 and R_2 in an SCG are connected by an arc whose attached link information is a convex intersection line L. Because of the properties of the gradient space shown in Figure 17, the constraint imposed by the arc can be represented as

$$G_2 - G_1 = k \cdot P. \tag{3}$$

Here P is a two-dimensional unit vector ($|P|=1$) with a direction perpendicular to L, and pointing from R_1 to R_2. The k is a positive scalar constant. If the kind of intersection is concave(-), then the constraint between G_1 and G_2 is also represented by the form (3), but in this case P points from R_2 to R_1, as shown in Figure 17. If the intersection line is an occluded intersection (⊛), the constraint between G_1 and G_2 is represented as

$$G_2 - G_1 = k_1 \cdot P_1 + k_2 \cdot P_2. \tag{4}$$

The direction of P_1 and P_2 are defined as in Figure 18, and $k_1>0$ and $k_2\geq 0$. (Since the occluded intersection line can coincide with BC, k_2 can be 0.) The non-negative linear combination of P_1 and P_2 spans the fan-shaped area in which G_2 is located with respect to G_1.

Notice that if we have two arcs $R_1 \rightarrow R_2$ and $R_2 \rightarrow R_3$, and concatenate them to form a path ($R_1 \rightarrow R_2 \rightarrow R_3$), then it is readily understood that the gradient of node R_3, relative to that of R_1, should be within the area spanned by a set of vectors which is the union of the set for $R_1 \rightarrow R_2$ and the set for $R_2 \rightarrow R_3$.

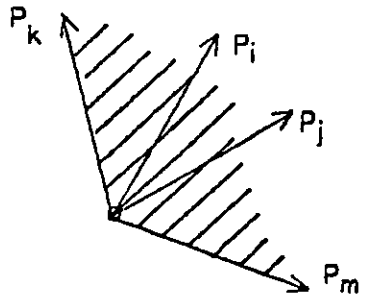


Figure 22 Spanning angle of a path.

In any event, the constraint which the SCG imposes between a node pair (p,q) connected by a path γ=(p→q) is that if the gradient of one node is fixed, the gradient of the other node should be within a fan-shaped area (or, in a special case, on a half line as in equation (3)), which is spanned by a non-negative linear combination of vectors. Let us call the fan-shaped area the spanning angle S_γ of the path (see Figure 22).

We can define the following computations on spanning angles: inverse, union, and intersection. Suppose that a path (p→q) has a spanning angle $S_{(p\to q)}$. If we traverse the path inversely as (q→p), the corresponding spanning angle $S_{(q\to p)}$ is the angle spanned by a set of vectors obtained by inverting the vectors which define $S_{(p\to q)}$. Graphically, as shown in Figure 23(a), the spanning angle $S_{(q\to p)}$ is the angle opposite to $S_{(p\to q)}$. This operation is called the inverse of spanning angle, and is denoted by $S_{q\to p} = -S_{p\to q}$.

We can concatenate two arcs, or more generally, two paths (p→q) and (q→r) to form a

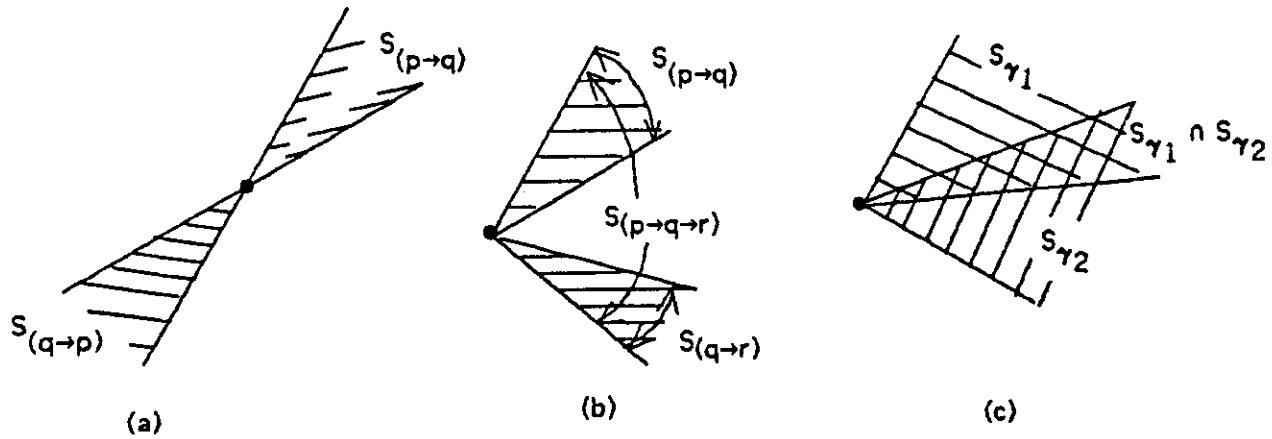


Figure 23 Computation on spanning angles: (a) inverse; (b) union; (c) intersection.

longer path ($p \rightarrow q \rightarrow r$). The operation of spanning angles corresponding to this concatenation is the union. Graphically, as shown in Figure 23(b), $S_{(p \rightarrow q \rightarrow r)}$ is the angle of the area which either belongs to one of $S_{(p \rightarrow q)}$ and $S_{(q \rightarrow r)}$ or belongs to the area which is separated by $S_{(p \rightarrow q)}$ and $S_{(q \rightarrow r)}$ and which has an angle less than 180^{o}. Let us denote this operation by $S_{(p \rightarrow q \rightarrow r)} = S_{(p \rightarrow q)} \cup S_{(q \rightarrow r)}$. By the union operation, the resultant spanning angle can become 360^{o}: the whole 2-D space. This happens when the set of vectors includes more than three vectors and the angles made by a neighboring pair of vectors are all less than 180^{o}. In such a case, the path does not give to the node pair any constraints regarding the relative locations of their gradients.

Now, if there are two paths γ_1 and γ_2 from the node p to q, then they impose constraints on G_p and G_q simultaneously; i.e.,

$$G_q - G_p = \Sigma\, k_{i1} \cdot P_{i1} = \Sigma\, k_{i1} \cdot P_{i2}, \qquad (5)$$

where $\{P_{i1}\}$ is from γ_1 and $\{P_{i2}\}$ from γ_2. This means that G_q should be within the overlapping area of S_{γ_1} and S_{γ_2} (Figure 23(c)). This overlapping area is the intersection of spanning angles, and is denoted by $S_{\gamma_1} \cap S_{\gamma_2}$.

Loop-Free SCG and Elementary Paths

The operation of intersection of spanning angles suggests that we can reduce an SCG into a simpler form on which our test will be applied. First of all, it is easily seen that only those parts of the SCG which include a loop or circuit need to be actually considered. This implies that if the SCG can be separated into two subgraphs by cutting a single arc, then each subgraph can be considered independently. In particular, leaf nodes can be eliminated from the consideration. Thus, we can "prune" and "cut" the SCG into a set of *leaf-free connected SCG (LF-SCG)*'s, each of which is independently subject to the consistency check. In Figure 21, the leaf (R_3, R_6) can be pruned, and the remainder is the LF-SCG.

Further, it is understood that the gradients of the nodes which are connected to exactly two other nodes (i.e., their node degree is two) in an LF-SCG, such as the nodes s and t in Figure 24, are relatively less important. They do not affect other nodes beyond p or q; it is only required that the relation (vector) between G_p and G_q is kept the same. This implies that we can divide an LF-SCG into a collection of paths, each of which begins and ends with nodes of degree more than two, and each of which contains only nodes of degree two in between. Let us call such a path an *elementary path*. In Figure 21, paths such as ($R_2 \rightarrow R_3 \rightarrow R_4$) and ($R_2 \rightarrow R_1$) are elementary paths. We can associate a spanning angle with each (directed) elementary path. What we need now is a computational procedure on the spanning angles of the elementary paths of an LF-SCG, to see whether the constraints on surface orientations can be satisfied.

Filtering Operation on Spanning Angles of Elementary Paths

Now we are ready to describe the filtering operation defined on the SCG. We assume that the SCG for a given interpretation has been simplified and decomposed into LF-SCG's, and that each LF-SCG is decomposed into elementary paths. If there is no LF-SCG, there is no need for performing the filtering procedure, and the test trivially succeeds.

(1) For each elementary path γ, associate an initial spanning angle $S_\gamma^{(0)}$ which is computed from a set of vectors defined for the arcs belonging to the path. Set n←1.

(2) For each elementary path γ, let $\{\Gamma_j\}$ be a set of all the paths that connect the same node pair in the same direction as γ connects. Since the LF-SCG is decomposed into elementary paths, each Γ_j is a concatenation of several elementary paths $\{\gamma_{jk}\}$; i.e., $\Gamma_j = \gamma_{j1} \cdot \gamma_{j2} \cdots \gamma_{jk}$. The spanning angle of Γ_j, $S_{\Gamma_j}^{(n-1)}$, is computed from the union of the spanning angles of the component paths, $S_{\gamma_{jk}}^{(n-1)}$; i.e.,

$$S_{\Gamma_j}^{(n-1)} = \bigcup_{\gamma_{jk} \in \Gamma_j} S_{\gamma_{jk}}^{(n-1)}.$$

Then, filter the $S_\gamma^{(n-1)}$ by the intersection of $S_{\Gamma_j}^{(n-1)}$, and set the result to $S_\gamma^{(n)}$; i.e.,

$$S_\gamma^{(n)} \leftarrow \left(\bigcap_{\Gamma_j \in \{\Gamma_j\}} S_{\Gamma_j}^{(n-1)} \right) \cap S_\gamma^{(n-1)},$$

If any $S_\gamma^{(n)}$ becomes null, then the test fails.

(3) If there exists an elementary path such that $S_\gamma^{(n)} \neq S_\gamma^{(n-1)}$, then n←n+1 and go to (2). Otherwise, the test terminates with success.

Four things should be noted about the test procedure. First, if an LF-SCG consists of a single circuit (i.e., the degree of all the nodes is 2), then we can pick up any pair of nodes and regard the two paths connecting them as elementary paths. Or, alternatively, this is equivalent to testing whether the spanning angle corresponding to the circuit is the entire 360^o.

Second, the iteration in the procedure always terminates, since all the $S_\gamma^{(n)}$'s monotonically decrease in step (2) and the number of possible spanning angles which $S_\gamma^{(n)}$'s can take is finite. (It is bounded by the number of subsets of the vectors involved in the LF-SCG under the test.)

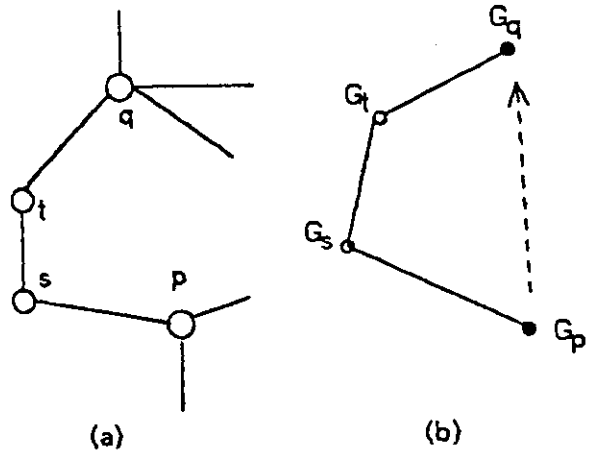


Figure 24 Elementary path: (a) SCG; (b) Gradient space.

Third, the procedure is a necessary condition for surface orientations to satisfy all the constraints represented in the SCG, but it is *not* a sufficient condition for that. An example of this will be given in IV-1.

Fourth, the presented algorithm is a conceptually straightforward one, but it is inefficient. Implementation of the algorithm can exploit several properties of the SCG to increase efficiency.

If all the LF-SCG's pass the above test procedure, then the given interpretation is said to pass the test for the surface orientation consistency. The above test procedure, together with the up-to-3-surface junction dictionary, defines the nature of the Origami world. An interpretation of a line drawing is called *plausible* in the Origami world, if all the junctions are given legal junction labels contained in the dictionary, *and* if its SCG passes the above test.

For Figure 20, it is easy to see that the SCG consisting of two nodes does not pass the test. Let us next consider the SCG in Figure 21. The path $\gamma=(R_2 \rightarrow R_3 \rightarrow R_4)$ is an elementary path, and the path $\Gamma_1=(R_2 \rightarrow R_1 \rightarrow R_4)$ is one of the paths which connect R_2 and R_4. The spanning angles $S_\gamma^{(0)}$ and $S_{\Gamma_i}^{(0)}$ are calculated as shown in Figure 25. Since their intersection is null, the spanning angle $S_\gamma^{(1)}$ will become null in step (2), and the test fails.

On the other hand, if we change the concave line labels (-) between R_2 and R_3 and between R_4 and R_3 in Figure 15 to occluding boundaries (→) so that R_3 occludes R_2 and R_4, then the corresponding SCG will consist of two subgraphs: one includes nodes $\{R_1,R_2,R_4,R_5\}$, and the other $\{R_3,R_6\}$. It is easy to see that the latter subgraph has no LF-SCG and that the former passes the test; therefore the test succeeds this time. As another example,

consider Figure 26 (a) and one of its interpretations Figure 26(b). It is a figure obtained by adding a convex line CF to Figure 14(b). It is a paper-sided, truncated pyramid viewed from above. Its SCG consists of a single circuit (Figure 26(c)), and passes the test. Therefore the configuration of Figure 26(b) is plausible.

Two points should be mentioned concerning the properties of SCG. Suppose that in Figure 21 we are filtering the spanning angle of the elementary path $\gamma=(R_4 \rightarrow R_5)$ against all the paths that connect R_4 and R_5. If we have filtered γ by the path $\Gamma_1=(R_4 \rightarrow R_1 \rightarrow R_5)$, then we need not filter by such paths which travel a component elementary path of Γ_1, say $\gamma_1=(R_4 \rightarrow R_1)$, through a nonelementary path. For example, $\Gamma_2=(R_4 \rightarrow R_3 \rightarrow R_2 \rightarrow R_1 \rightarrow R_5)$ travels γ_1 through the nonelementary path $(R_4 \rightarrow R_3 \rightarrow R_2 \rightarrow R_1)$. The reason is that since $\gamma_1=(R_4 \rightarrow R_1)$ is itself an elementary path, the partial path $(R_4 \rightarrow R_3 \rightarrow R_2 \rightarrow R_1)$ of Γ_2 has been or will be used in filtering γ_1, and therefore, Γ_2 does not add any constraint different from Γ_1.

Another point is that $\Gamma_1=(R_4 \rightarrow R_1 \rightarrow R_5)$ does surely have an overlapping spanning angle with that of $\gamma=(R_4 \rightarrow R_5)$, for they form a circuit which surrounds a single junction. If we stop and think, this is what the junction label means: it assures a local consistency around a junction. These two facts can reduce the number of paths which are actually used in filtering. In our example, γ needs to be filtered only by $\Gamma_3=(R_4 \rightarrow R_3 \rightarrow R_2 \rightarrow R_5)$.

III-6 The Labeling Procedure in Origami World

In the previous four subsections we have first enumerated the legal junction labels in the Origami world, and have stored them together with link information in the Origami junction dictionary. Then, after introducing the concept of spanning angles, we have defined a test procedure on the SCG which can decide whether a given labeled interpretation of a line drawing is consistent with respect to surface orientations. This subsection will present a labeling procedure which, *given a line drawing*, finds all its plausible interpretations; that is, all the combinations of assignments of line labels which result in legal junction labels at all the junctions in the drawing, and which pass the test procedure for surface orientations.

Strategy

The labeling procedure must include two tasks:

(1) Using the junction dictionary of the Origami world, interpretations are generated in which the labels given to lines constitute legal junction labels at all the junctions. This could be done in two steps:

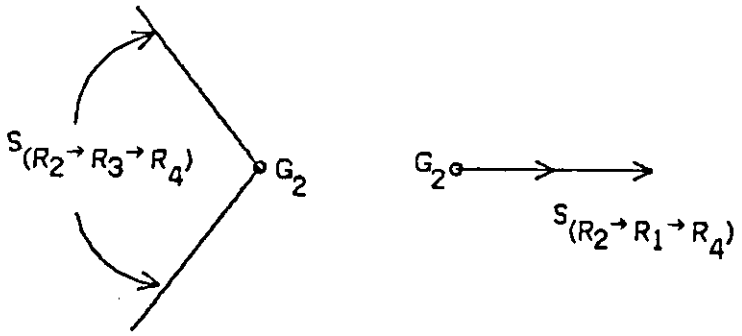


Figure 25 Spanning angles of paths in the SCG in Figure 21.

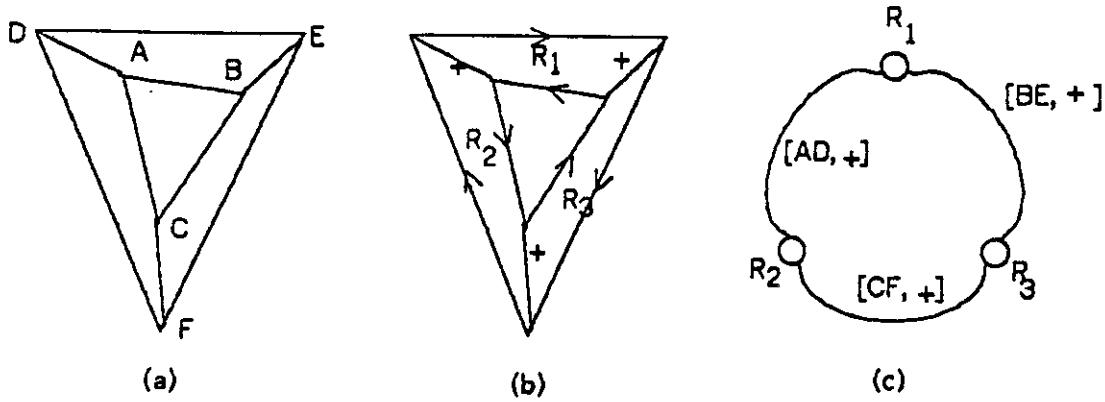
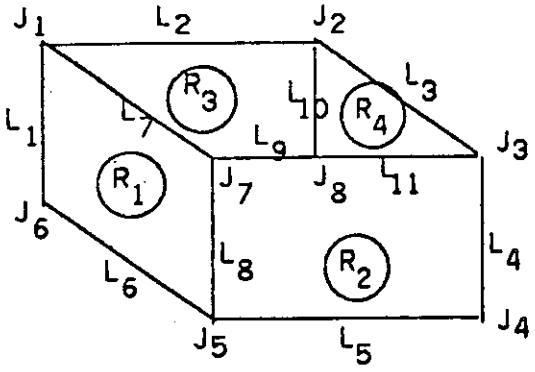


Figure 26 Paper-sided, truncated pyramid: (a) line drawing; (b) labeling; (c) SCG.

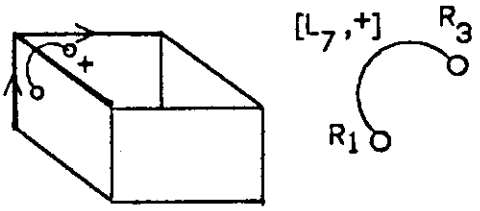
(1a) Filtering of junction labels (as in the Waltz labeling [Waltz, 1972]),
(1b) Tree searching to obtain the individual consistent combinations of junction labels

(2) For each interpretation obtained in (1), the SCG is constructed and the consistency of surface orientations is tested by filtering the spanning angles of elementary paths.

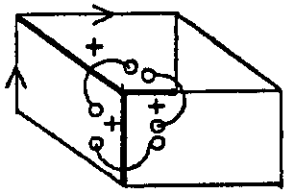
However, the number of interpretations which are generated in (1) as being consistent with respect to the junction dictionary is very large, and most of them are inconsistent with respect to the surface orientations. Also, if a certain subconfiguration is inconsistent with respect to surface orientations, any interpretation which includes that subconfiguration is never plausible. Therefore, in the actual implementation of the labeling procedure the steps (1b) and a part of (2) are combined into one process: while assigning a junction label to a



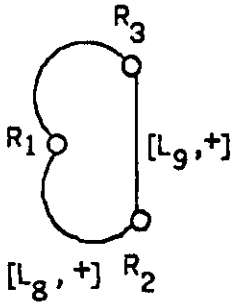
(a)



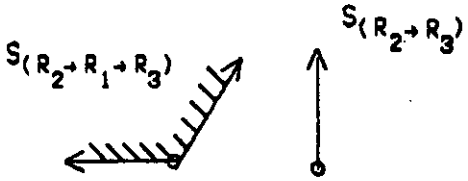
(b)



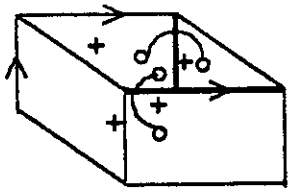
(c)

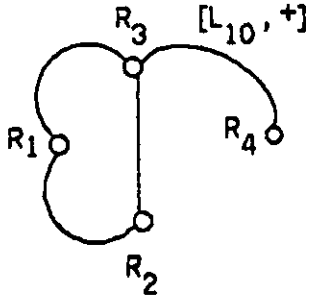


(d)

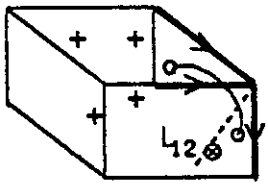


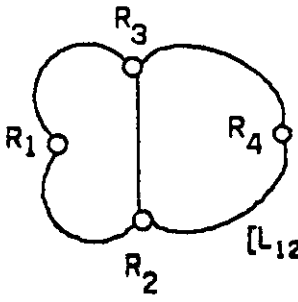
(e)



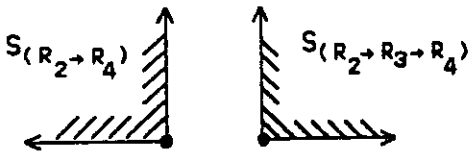


(f)

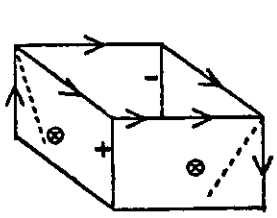


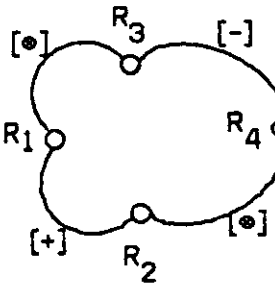


(g)



(h)





(i)

Figure 27 Explanation of labeling process for Figure 3.

junction in a depth-first manner, the process constructs the SCG incrementally and checks the spanning angles as far as possible. When any inconsistency, either in the junction labels or in the surface orientations, is detected, the process backtracks the search for the next combination. This combined process greatly increases efficiency in the labeling procedure.

As a result, the labeling procedure consists of three phases:
(1) Filtering of junction labels,
(2) Tree searching combined with filtering of spanning angles on a partial SCG,
(3) Final filtering of spanning angles of elementary paths.

Phase (1): Filtering of Junction Labels

This process is exactly the same as the Waltz method [Waltz, 1972]. At first, each junction is given a set of possible labels drawn from the dictionary according to its junction type. Initial constraints are that the outer boundaries should have the label, ← or →, with such a direction that the object is on their right hand side; the arrows surround the object clockwise. Then the Waltz filtering procedure is executed repeatedly to eliminate those junction labels incompatible with neighboring ones until no further elimination is obtained.

Let us work with a "box" line drawing of Figure 27(a). It has 11 lines, L_1 through L_{11}, and 8 junctions: J_1, J_2, J_3, and J_5 are ARROW's; J_4 and J_6 are L's; J_7 is a FORK; J_8 is a T. As the result of filtering, 4 junction labels remain for J_1 out of the original 12 possible ones contained in the dictionary for the ARROW type junction. Similarly, the number of junction labels which remain are: 4 out of 12 for J_2; 4 out of 12 for J_3; 1 out of 8 for J_4; 4 out of 12 for J_5; 1 out of 8 for J_6; 19 out of 19 for J_7 (since a Y junction is rotational, 19 possible assignments exist although there are only 9 entries in the dictionary); and 16 out of 16 for J_8.

Phase (2): Tree Search with Filtering of Spanning Angles

Each junction now has a set of possible junction labels which have been filtered through the previous process. Only those combinations which give consistent labels to lines are to be explicitly selected. This is done by a depth-first tree search. For the first junction, one junction label is selected from the set of possible junction labels for it. For the second junction, one junction label is selected which is consistent with the labels given to the first junction. Similarly, for the third junction, etc. If no more consistent label exists for the present junction, the search backtracks.

Combined with this recursive tree search process is the filtering of spanning angles. Whenever a junction label is assigned to a junction, the corresponding part of the SCG is

constructed. If the link of the present junction label adds an arc and results in the formation of a new circuit in the SCG, the spanning angle of this arc is checked against all the existing paths which connect the same node pair. If they have a null intersection, the partial interpretation is inconsistent.

If any inconsistency in the configuration is detected, either in the combination of junction labels or in the surface orientations, the tree search process backtracks one step and searches for the next combination. If all the junctions are labeled consistently, the resultant interpretation is handed to the final phase (3).

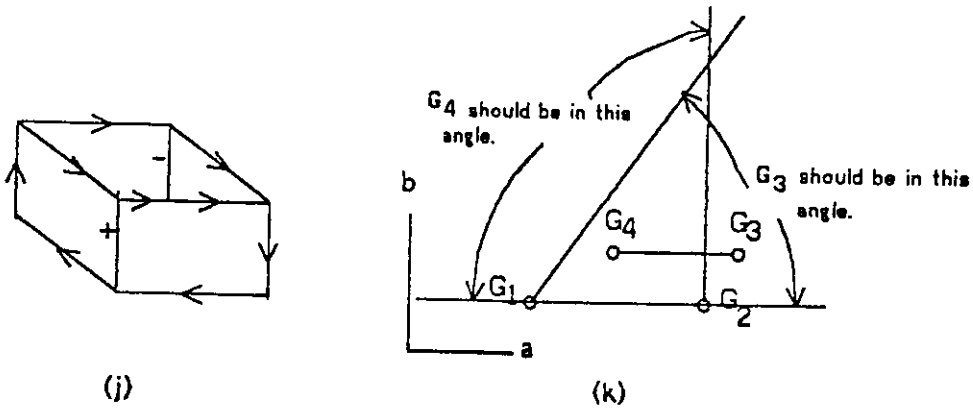
Let us see the example of Figure 27(a). Suppose that the junction J_1 is first given a junction label as shown in Figure 27(b); the corresponding partial SCG consists of a single arc $(R_1 \rightarrow R_3)$. Then, J_7 is given a junction label (see Figure 27(c)). Since the link it has between R_1 and R_3 is the same as the existing one, two new arcs are added to the SCG: first $(R_1 \rightarrow R_2)$, and then $(R_2 \rightarrow R_3)$. A circuit is formed when $(R_2 \rightarrow R_3)$ is added (Figure 27(d)). Its spanning angle is checked against the existing path $(R_2 \rightarrow R_1 \rightarrow R_3)$. As shown in Figure 27(e), the intersection is not null, and therefore the search proceeds to J_8. Suppose that the first choice of junction labels for it results in the interpretation and the corresponding SCG as in Figure 27(f). Since no new circuit is generated, the assignment of junction labels proceeds. J_3 is given a unique junction label determined by the line labels already given. As shown in Figure 27(g), when J_3 is given the junction label, it adds an arc $(R_2 \rightarrow R_4)$ and the SCG has a new circuit. Thus, the arc $(R_2 \rightarrow R_4)$ is to be checked against paths $(R_2 \rightarrow R_3 \rightarrow R_4)$ and $(R_2 \rightarrow R_1 \rightarrow R_3 \rightarrow R_4)$. Since the spanning angle of $(R_2 \rightarrow R_4)$ does not have an overlapping area with that of $(R_2 \rightarrow R_3 \rightarrow R_4)$ as shown in Figure 27(h), this interpretation turns out to be inconsistent. The process winds back to J_8, and the next choice for J_8 will be examined.

Consider another stage of the tree search in which the junction labels have been given as shown in Figure 27(i). When the junction J_3 is examined, it adds a new arc $(R_2 \rightarrow R_4)$. This time, the spanning angle of $(R_2 \rightarrow R_4)$ is compatible with that of $(R_2 \rightarrow R_1 \rightarrow R_3 \rightarrow R_4)$, and therefore the search proceeds. Since the rest of junctions do not add new arcs to the SCG, the interpretation shown in Figure 27(j) is handed to the final phase.

If we considered only junction labels, 90 interpretations would have been generated for the line drawing of Figure 27(a). However, by means of checking spanning angles in the course of tree searching, only 8 out of these 90 interpretations were passed to the final phase.

Phase (3): Final Filtering of Spanning Angles of Elementary Paths

The method of this phase is exactly the same as the test procedure described in III-5. The reason for the necessity of this phase is that the SCG is not completely constructed until



(j) (k)

Figure 27 (continued).

the tree search has completed the assignments of junction labels in an interpretation: we could not identify the elementary paths, and therefore we have only checked the spanning angle of *newly added arcs* against the existing paths which connect the same node pair. Partial duplication of this phase may appear inefficient, but usually most of the inconsistencies of surface orientations have been detected in the phase of tree searching, and only inconsistencies that involve very global relationships among regions remain undetected until this phase. This phase is also useful because it reveals the mutual relationships among the gradients of regions in the SCG; this information is used in reconstructing the 3-D shape of the scene.

In our example, all of the eight interpretations that are generated in the tree search pass this final phase. In the case of Figure 27(j), it is revealed that the gradients of the four regions should be placed in the gradient space as shown in Figure 27(k). The diagram represents the fact that the four surfaces form a convex opening in the 3-D space, which is probably a general description of a "box".

III-7 Examples of Labeling

A few examples of interpretations in the Origami world follows. Figure 28 shows three possible interpretations (without counting rotations) that a line drawing of Figure 4(a) can have: (a) a cube-like configuration; (b) a concave corner; and (c) a "roof" placed on a plane.

A "box" line drawing of Figure 3 has eight possible interpretations shown in Figure 29: (a) corresponds to an ordinary box; (b) is a "squashed" box whose front sides are pushed backward; (c) is another "squashed" box whose rear sides are pulled forward; (d) is a box with a triangular lid in the right corner; etc.

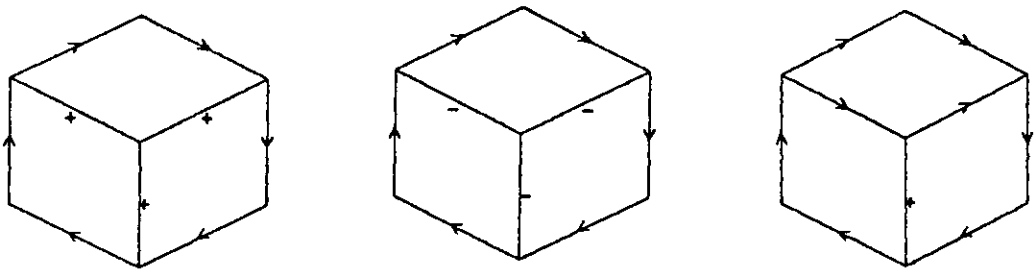


Figure 28 Interpretations of Figure 4(a).

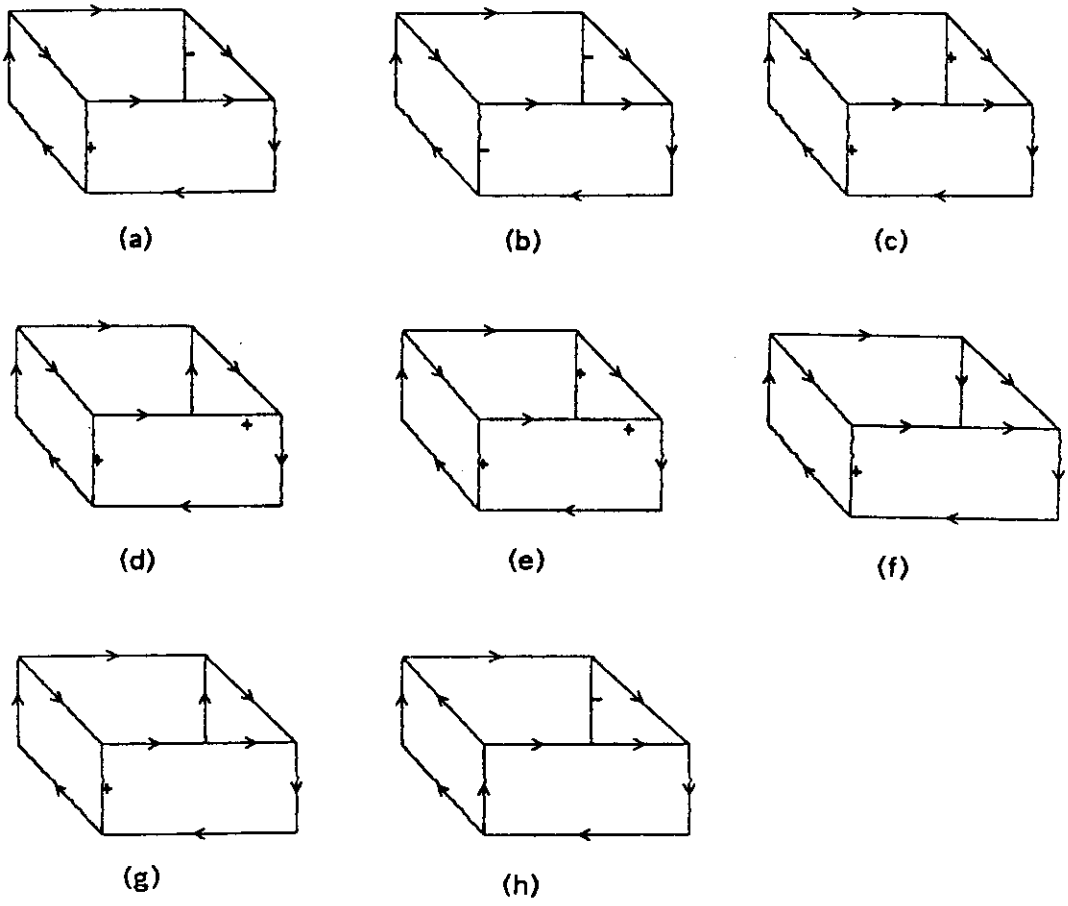


Figure 29 Interpretations of a "box" line drawing.

Figure 30 shows 11 interpretations of Huffman's "impossible" pyramid: (a) a solid, truncated pyramid; (b) a paper-sided, truncated pyramid; (c) a view of (a) from the bottom; (d) a view of (b) from the bottom; (d) a triangular "dome" with an opening in the lowest side; etc.

Figure 31 is another example of interpreting an "impossible" object. It has 10 interpretations. The interpretation (a) corresponds to three twisted rectangular bars.

Figure 32 includes 16 possible interpretations of Figure 4(b). Interpretation (a) corresponds to the W-folded paper.

IV Discussion

The discussion in this section is divided into three parts. The first part discusses how the test criteria employed for checking the surface orientations in the Origami world are related to plane surfaces. The second part reveals interesting relationships of the Origami world to other worlds dealt with in prior work on polyhedral scene analysis. The third part discusses how knowledge is used in understanding line drawings.

IV-1 Plane Surfaces and Origami World

We have noted that the test procedure, described in III-5, to check the consistency of surface orientations in the Origami world is not a sufficient condition for the constraints in the SCG to be satisfied simultaneously. But it should be remembered that the constraints in the gradient space themselves are not a sufficient condition for the configuration to be realized by plane surfaces. Consider again the configuration of Figure 26(b). The configuration made of three regions R_1, R_2, and R_3 has passed the test. However, it is a simple geometry problem to show that unless three lines AD, BE and CF meet at a single point, the configuration is not realizable by the three planes corresponding to R_1, R_2, and R_3.

The problem arises from the fact that the gradient space does not take into account the value of c in equation (1): a consistent trace in the gradient space means that the corresponding regions can take a consistent combination of (a,b) values, but it does not necessarily assure a consistent combination of (a,b,c) values. Huffman [Huffman, 1977] presents a $\phi(\phi')$-point test as the necessary and sufficient condition for a "cut set" of lines (equivalently, a set of regions separated by those lines) to be realizable by plane surfaces. Consider again the example of a paper-sided, truncated pyramid shown in Figure 33(a), and take the set of lines AD, BE and CF cut by the dotted loop. Each line belonging to the cut set of lines is given an orientation shown as a big arrow according to its label, either coming into the loop (if the label is +) or going out from it (if the label is -) (see Figure 33(b)). Then the

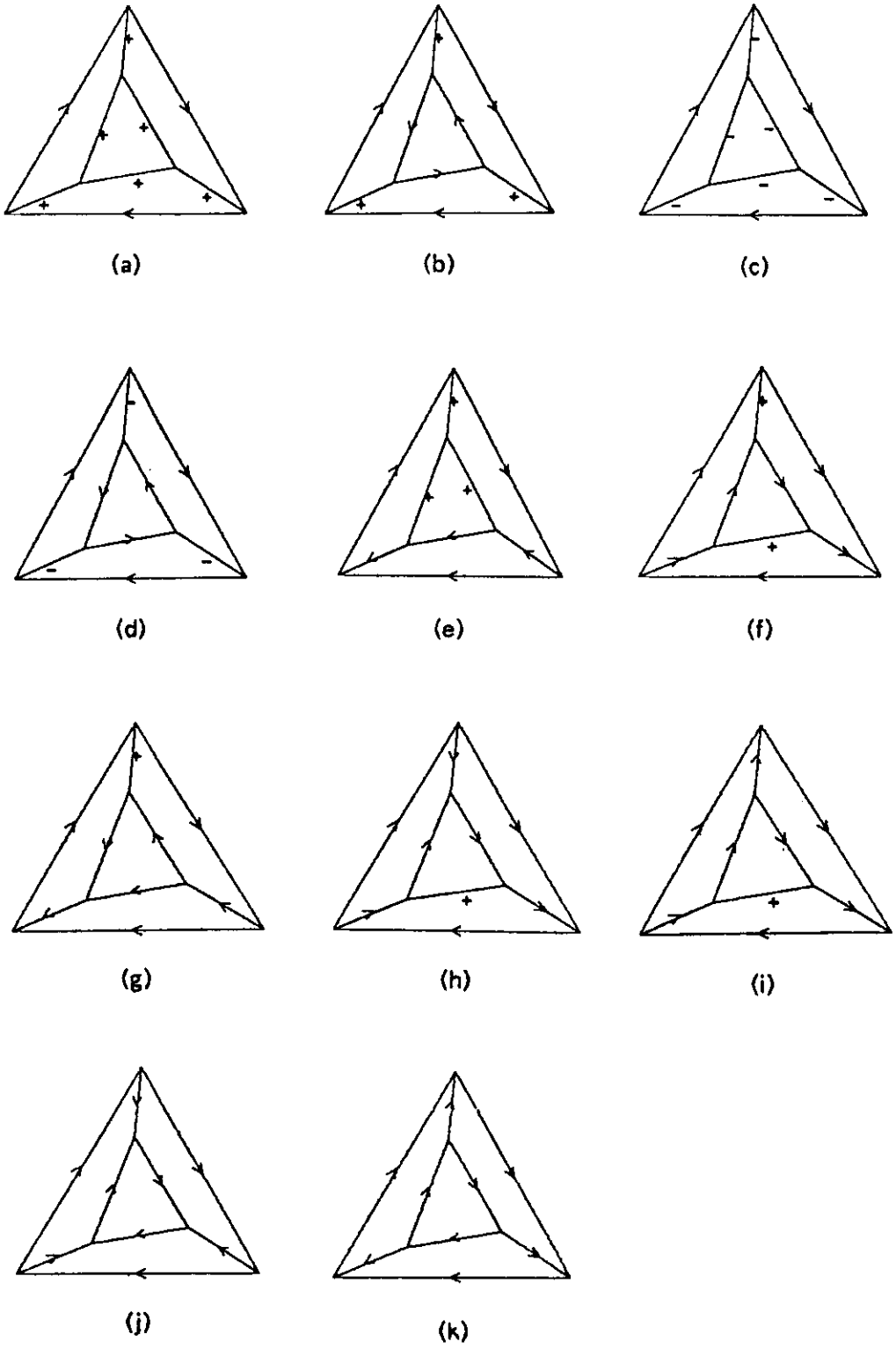


Figure 30 Interpretations of an "impossible" pyramid.

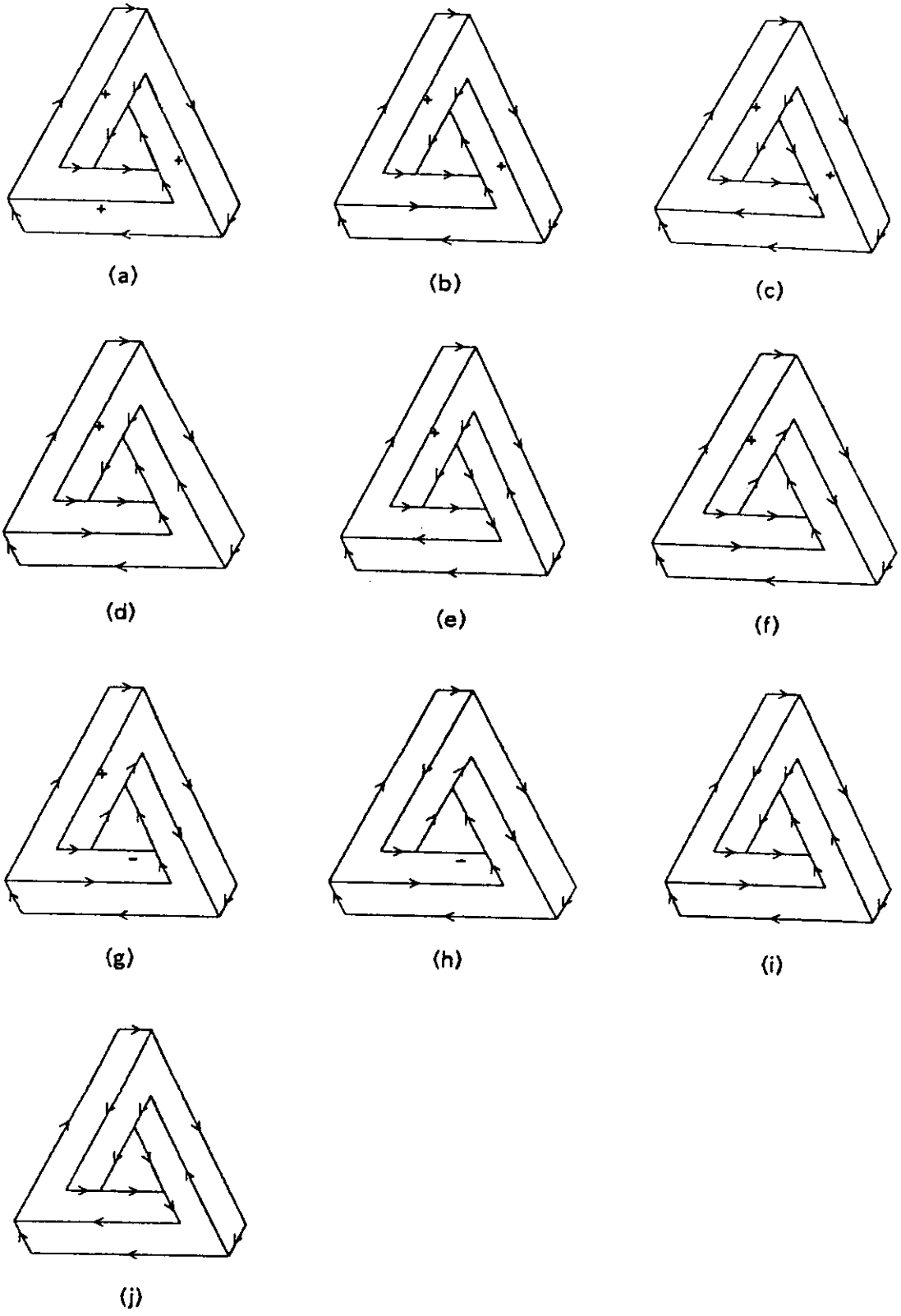


Figure 31 Interpretations of an "impossible" twisted object.

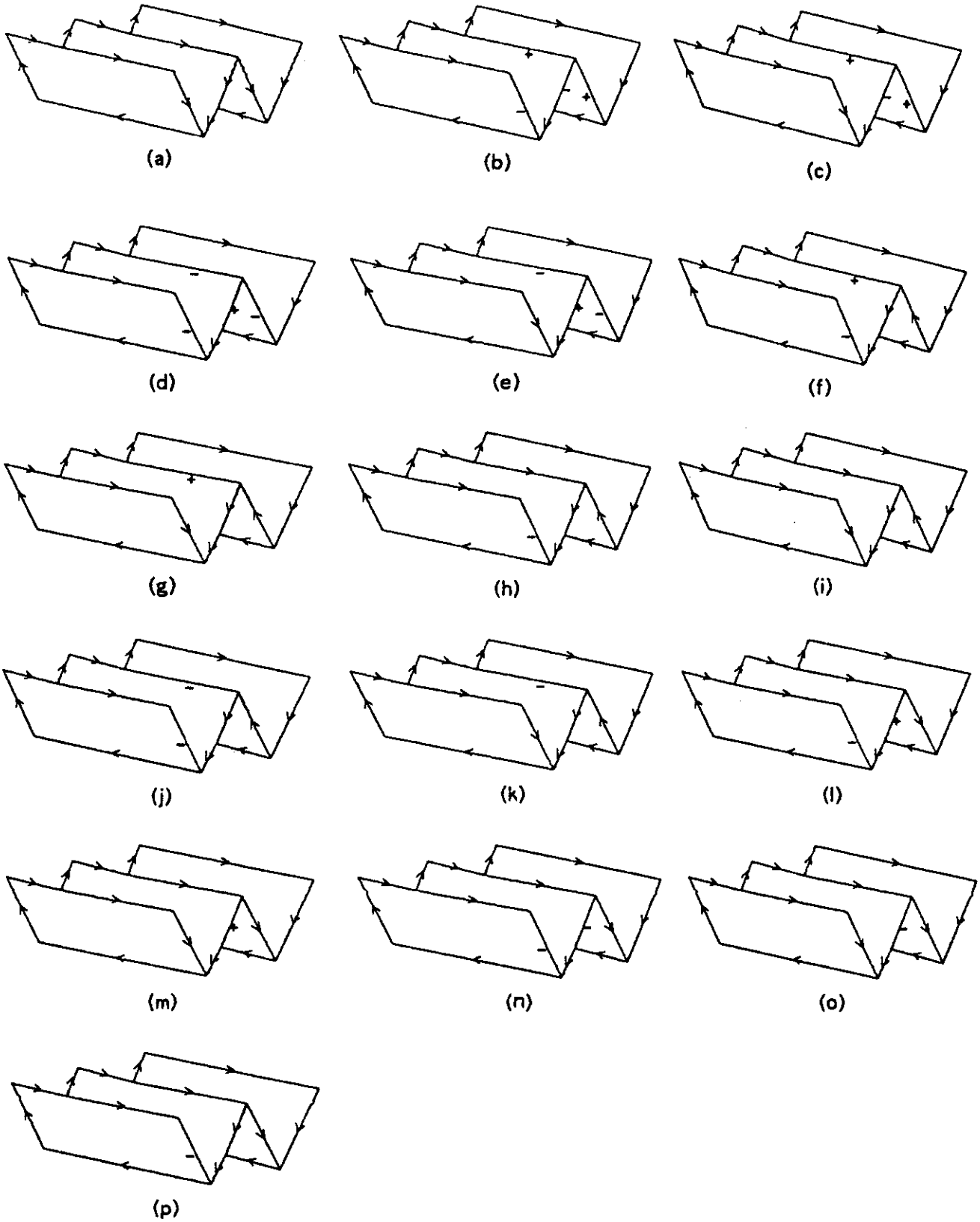


Figure 32 Interpretations of a "W-folded paper" line drawing.

$\phi(\phi')$-point is a point that is to the right (left) of some line of the cut set and that is not to the left (right) of any other lines. The $\phi(\phi')$-point test simply checks whether either a ϕ-point or a ϕ'-point exists, and if either one exists, then the cut set is unrealizable. In fact, unless AD, BE and CF meet at a single point, ϕ or ϕ' points exist, and therefore the configuration of Figure 33(a) is unrealizable. Unfortunately, it can *not* be said that if all the cut sets in the interpretation pass the $\phi(\phi')$-point test, then the whole interpretation is realizable by only plane surfaces. (Notice that the $\phi(\phi')$-point test is the necessary and sufficient condition for the realizability of a cut set of lines, not of the whole interpretation).

To human eyes, the configuration of Figure 33(a) appears quite reasonable; sometimes it takes time to persuade people that the configuration is not realizable by plane surfaces. Figures 34(a) through 34(c) are other examples of interpretations of simple figures which appear plausible but not actually realizable by only plane surfaces: they pass our test in the Origami world, but not the $\phi(\phi')$-point test.

The use of only the gradients (a,b) makes some sense mathematically when we consider the manner in which we view a picture. Note that the gradient (a,b) of a plane is invariant to the x-y translation of the picture plane (i.e. shift of eye position). In viewing an orthographically projected picture, we do not fix an absolute origin in mind. When we see a line where two surfaces actually intersect, we tend to "place" the origin on that line, which means we give the two surfaces the same c value. Therefore, constraints about c are automatically satisfied at that intersection. Also, when we see occlusions, we tend to attribute it to the difference of the c value rather than to any relations among a, b, and c. As we shift our eye and move the origin in the picture, it is easy to keep the gradient of a particular region in mind, but it seems difficult for us to "calculate" the value c which that region should have in the new coordinates. These observations seem to explain why the objects in Figure 33(a), Figures 34(a) through 34(c) do not look impossible at first glance.

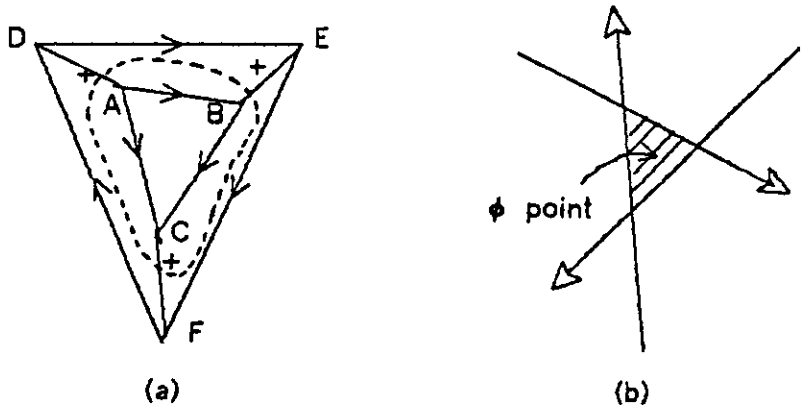


Figure 33 Huffman's $\phi(\phi')$-point test.

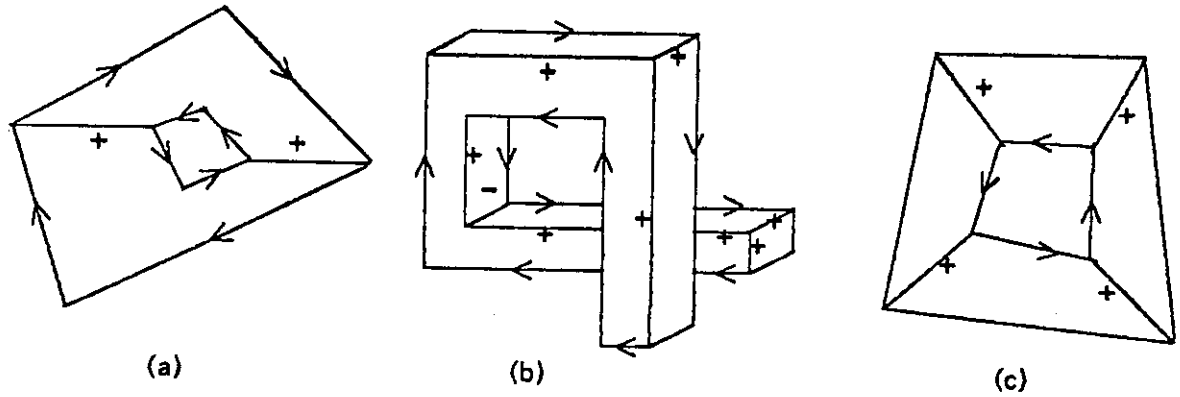


Figure 34 Examples of "plausible" interpretations which can not be made of plane surfaces (the unlabeled lines are occluding boundaries, ← or →, in the obvious direction).

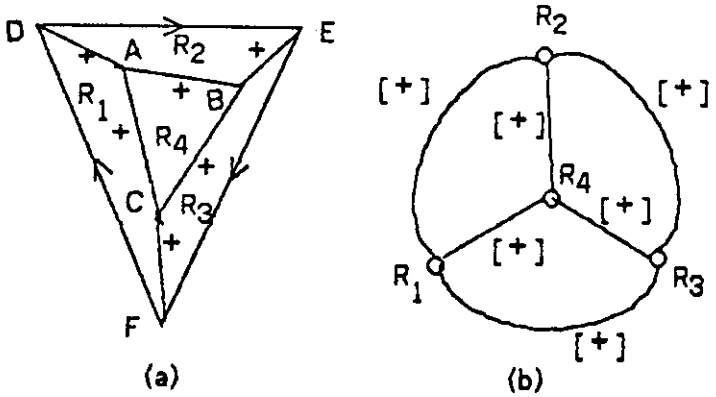


Figure 35 Solid truncated pyramid. An example in which all the constraints in the gradient space cannot be satisfied simultaneously, but all the pairwise constraints with respect to others can be satisfied.

Another issue about the test procedure deserves comment: its insufficiency for assuring that all the constraints on the surface orientations represented in the SCG are satisfied simultaneously. The test procedure presented in III-4 merely assures that for any pair of regions which are connected by an elementary path (it means that they intersect directly or they are connected by a sequence of free-hanging surfaces between them), the constraints on the surface gradients that relate the pair of regions are all satisfied. In order to understand the difference, consider the interpretation of Figure 35(a), which is a solid, truncated pyramid viewed from above. The corresponding SCG is shown in Figure 35(b). The SCG passes our test procedure, but it is not possible to find a set of gradients for all the regions R_1 through R_4 so that all the constraints in the SCG are satisfied simultaneously, unless AD, BE, and CF meet at single point. This crucial difference stems from the fact that when we pick up a pair of regions, say, R_1 and R_2, the gradients of regions R_3 and R_4 are not fixed uniquely in checking whether the paths $(R_1 \rightarrow R_4 \rightarrow R_2)$ and $(R_1 \rightarrow R_3 \rightarrow R_4 \rightarrow R_2)$ have an overlapping spanning angle with that of $(R_1 \rightarrow R_2)$.

IV-2 Origami World and Various Worlds

The theory of the Origami world and the labeling procedure for it have interesting relationships with the work of Guzman [Guzman, 1968], Huffman [Huffman, 1971], Clowes [Clowes, 1971], Waltz [Waltz, 1972], Mackworth [Mackworth, 1973], and Huffman [Huffman, 1977]. Their work all concerned the problem of recovering three-dimensional configurations from line drawings. (The historical development in polyhedral scene analysis has been well reviewed by Mackworth [Mackworth, 1977].)

Assume we consider only a set of line drawings which are reasonably "likely" figures; that is, it does not include figures which show too anomalous behavior. We can consider a set of all the combinatorially possible interpretations (assignments of line labels) of those line drawings. A subset exists containing those interpretations which can be realized by plane surfaces. Let us denote that subset as the *Plane Surface World*, S_{psw}. We can also think of a subset consisting of interpretations in which all the constraints on the gradients of surfaces are completely satisfied. Let us call it the *Consistent Gradient World*, S_{cgw}. Obviously $S_{cgw} \supset S_{psw}$.

We can view a labeling procedure as a method consisting of a generator and a tester: given a line drawing, a generator generates interpretations in a certain manner, each one of which a tester accepts or rejects based on a certain method. Table 2 summarizes various labeling methods according to this taxonomy. Various subsets can be defined which are generated by generators, or are determined as legal by testers. We will discuss the relationships among those subsets, referring to Figure 36.

Table 2. Various labeling schemes as a method of a generator and a tester.

Method	Generator	Subset	Tester	Subset
Huffman Clowes Waltz	Trihedral junction dictionary Trihedral junction dictionary with cracks, shadows, etc	S_{tri}		
Mackworth	Sequential generation of most connected interpretations		Constructive test on coherence rules in the gradient space	S_{poly}
Huffman			$\phi(\phi')$ point test for all the cut sets in the line drawing	$S_{\phi(\phi')}$
Origami World	Up-to-3-surface junction dictionary	S_{up3}	Filtering of spanning angles on the SCG	$S_{origami}$

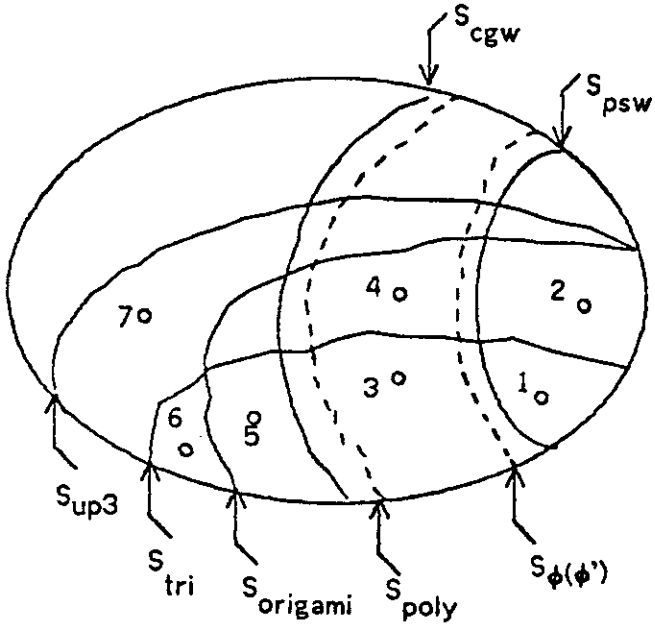
Huffman [Huffman, 1971], Clowes [Clowes, 1971] and Waltz [Waltz, 1972] used a trihedral junction dictionary as the generator and did not use any tester. Let us denote the subset of interpretations generated by the trihedral junctions as S_{tri} (Waltz used cracks, illuminations, shadows, etc., and the corresponding subset is different from S_{tri}, but because geometrically his dictionary is a trihedral one, it is included in this category). As we saw, S_{tri} is larger than the subset of solid trihedral objects bounded by plane surfaces.

The Huffman's $\phi(\phi')$-point test, when used on all the cut sets in the line drawing, can define a subset $S_{\phi(\phi')}$, which is larger than S_{psw}, but is the closest one to it articulated so far. However, since there is no appropriate efficient generator, it would be difficult, given a line drawing, to actually enumerate all the interpretations belonging to $S_{\phi(\phi')}$.

Mackworth's POLY used a generator which generates combinations of line labels based on some preferences. In particular, to achieve the most connected interpretations, it

generates first the interpretation in which all the edges are connected (i.e., all the lines are given either + or -), and then when such an interpretation fails to pass the test, it generates an interpretation with all edges but one as connected edges, and so on. The consistency in the gradient space is tested by a constructive method which actually tries to fix the positions of gradients corresponding to the regions, step by step. In this way, POLY avoids the use of predetermined interpretations for particular categories of junctions (such as L, ARROW, FORK, etc.), and thus, theoretically, the subset S_{poly} could be equal to S_{cgw}. However, since it is not practical to test all the interpretations, a certain selection criterion is needed to supply the generator with advice or preferences concerning the order of generation. The constructive test procedure also uses some heuristic rules, because the construction is not a trivial process. As the result, the actual nature of S_{poly} becomes a little unclear.

In the Origami world, the subset S_{up3} generated by the Origami junction dictionary properly includes S_{tri}. The subset $S_{origami}$, consisting of plausible interpretations which have passed the test procedure, is a little larger than the up-to-3-surface objects in the S_{cgw}, as shown in Figure 36. One feature of $S_{origami}$ is that it has a clear definition of the membership, which allows an efficient procedure to generate the member interpretations for a given line drawing; i.e., filtering of junction labels and spanning angles on the SCG. The locations of several example interpretations are indicated in the diagram, and they can serve to illustrate relationships among the subsets.



1: cube (Figure 6)
2: box (Figure 29(a))
3: Figure 34(b)
4: paper-sided, truncated pyramid (Figure 30(b))
W-folded paper (Figure 32(a))
Figure 34(a)(c)
5: solid truncated pyramid (Figure 30(a))
6: Figure 15
7: Figure 14

Figure 36 Relationship among various subsets of interpretations.

It is interesting to review Guzman's work[Guzman, 1968] on object recognition at this point. In its goal, his work is the forerunner of Huffman, Clowes and Waltz. He tried to find "good" associations of regions into separate 3-D bodies based on heuristics concerning junction types. The idea of a "region" (which is a projection of surface to the picture plane) is very close to the idea of the surface-oriented world. Also, the links he used represent the possible connections of regions like our links. However, since his links are defined for junction types rather than for junction labels, his link information is a kind of "average" over various combinations of surfaces at a particular junction type. Because of its heuristic nature, Guzman's method could not explicitly clarify the world for which it is intended.

Table 2 suggests that we can employ different combinations of generator and tester depending on the world in which we want to work. For instance, the trihedral junction dictionary together with Huffman's $\phi(\phi')$-point test is the closest to the plane-surface trihedral solid object world; the up-to-3-surface Origami junction dictionary together with the $\phi(\phi')$-point test expands the above world to allow up-to-3-surface objects; the trihedral junction dictionary together with filtering of spanning angles on the SCG defines the solid-object subset of the Origami world; and so on.

IV-3 Use of Knowledge in Understanding Line drawings

To divide a labeling procedure into a generator and a tester, as in Table 2, provides another interesting observation concerning the manner of using knowledge in understanding line drawings. The junction dictionaries rely on the junction categories. Thus, in order to make the generators work, we use geometrical information in a given line drawing, up to the precision necessary to classify junctions into the predefined categories. As long as a junction is classified, say as a FORK, precise angles between lines are *not* important. In contrast, the testers use precise values of angles and locations of lines. The junction dictionary is a form of precompiled knowledge about the geometrical constraints in the world, and it allows an efficient filtering method on junction labels to recover part of three-dimensional configurations. In contrast again, the testers need actual numerical computations which involve the global connections of junctions and which depend on the individual drawings.

In the case of the trihedral solid-object world, the labeling is based only on the generator, and the interpretations included in S_{tri} are the output. Still, most often a unique or a few interpretations are obtained and they appear reasonable to humans. This happens because the trihedral world is so constrained that a global test is not necessary, and the subset of interpretations which are in S_{tri} but which do not appear reasonable to us is not large for "likely" figures.

Minsky [Minsky, 1974] uses the side view of the "impossible" truncated pyramid as an example to show how little humans rely on numerical three-dimensional geometrical information in seeing objects. However, since the interpretation of Figure 15 *does* look unreasonable to us, a more precise statement would be that humans use the geometrical information to check certain consistencies in the gradients of regions, but not the total consistency among them.

In the case of the Origami world, the precompiled knowledge in the Origami junction dictionary can reduce the possibilities to S_{up3} by means of filtering the junction labels. The filtering of spanning angles on the SCG can reduce the possibilities further to $S_{origami}$. The same type of filtering procedure is used both for exploiting the precompiled knowledge and for the dedicated computation; one is used symbolically, the other numerically. It seems that the difference $S_{up3} - S_{origami}$ is fairly large in the surface-object world, and thus the tester is really needed. The following is to be noted. The junction labels hold information concerning a very local consistency, as was pointed out in III-5. They can propagate information to the neighboring junctions only through line labels. In contrast, the links can globally transfer information to any junctions through regions.

V Application of Origami Theory to Recover 3-D Configurations from Image

The theory of the Origami world has great potential in applications to image understanding tasks that recover three-dimensional configurations. That it can deal with scenes which include free extending surfaces is very attractive, because real world images include objects which are practically or conceptually flat. In fact, the Origami world corresponds well to the way in which we would interpret a picture which has been segmented into regions. Suppose that Figures 37(a) and 37(b) are obtained as the results of region segmentation of "chair" and "door" scenes, respectively. They are satisfying to us,

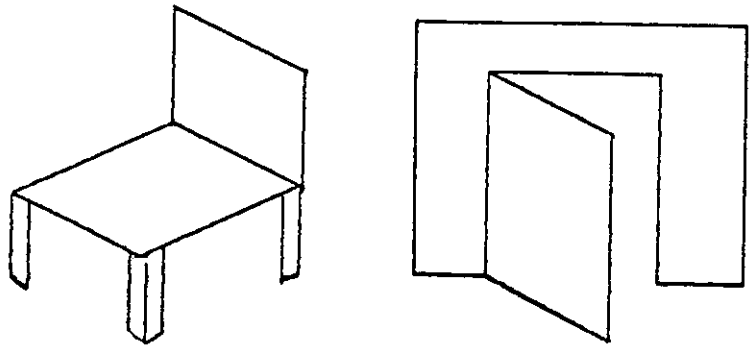


Figure 37 Region segmented picture of "chair" and "door" scenes.

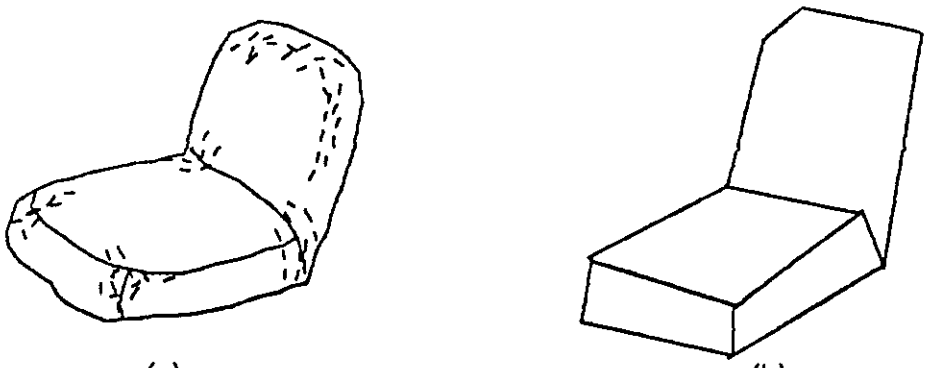


Figure 38 (a) Moderately curved chair seat; (b) Possible line drawing

because we interpret them in terms of surfaces. Needless to say, the Origami world includes the solid-object world as its subset.

Just as we generalized from solid volume to surface, we can go further and say that a "pencil" has conceptually a line shape, thus we need a "wire-frame" world, and further, a "dot" world. The more basic the unit of the world is, the broader class of pictures it can deal with, but at the same time the less constraints it provides. We feel that the Origami world is rich enough to accept a large class of line drawings, and at the same time it has enough structure to impose constraints on the possible label combinations.

Even though the Origami world is not intended for curved objects or imperfect line drawings, a certain class of line drawings including curved objects or imperfections can be accommodated within it. As an illustrative example, suppose that a moderately curved chair seat of Figure 38(a) yields a line drawing of Figure 38(b). While it is an "impossible" figure in the trihedral world, it has an interpretation in the Origami world corresponding to "a rectangular block with a flat sheet attached". Once this is hypothesized, further processing, probably involving image data analysis, can discover the detailed shape and know whether it is a square or curved, solid or flat object.

In real image understanding tasks, the number of lines is large, and therefore the number of possible interpretations is also large. Even the line drawing of a box (Figure 3) has eight interpretations, for instance. However, spectral (shading, color, etc.) and geometrical (collinearity, parallelism, etc.) knowledge can be used here to reduce the number of possible interpretations. There is knowledge that relates the nature of edges and their

intensity profiles taken across the edge [Horn, 1977]; for example, a peak-shaped edge profile often suggests a convex edge. Another typical example which provides constraints on line labels is a so-called matched T shown in Figure 39. If the edge profiles of lines L_1 and L_2 are similar (and, preferably, if the edge profiles of L_3 and L_4 and those of L_5 and L_6 are also similar), then the labels of L_1 and L_2 are likely the same, and L_3 through L_6 are likely to be occluding boundaries, with the middle region R obscuring L_1 and L_2. All these constrain, conceivably in a probabilistic way, the possible combinations of line labels that a set of lines can take. Therefore, "best" or "most plausible" interpretations can be defined and searched. For example, in the case of Figure 27(a), if L_8 and L_{10} are known to take the same label, the number of possibilities reduces from 8 to 3 (see Figure 29).

Further, heuristics concerning surface orientations can be used to provide preferences for interpretations. For example, if interpretations in S_{tri} (i.e., all the junction labels given are trihedral ones) are to be chosen first, then in the cases of Figures 28, 30, and 31, the interpretations corresponding to a cube, a solid truncated pyramid, and three twisted bars are selected, respectively. Another heuristic is that parallel lines in the picture are also preferably parallel in the scene; this would prefer the interpretations corresponding to an ordinary box and a W-folded paper over others in Figures 29 and 32, respectively. This "parallel-in-the-picture/parallel-in-the-scene" heuristic seems very powerful for pictures which do not include strong perspective distortions. The subject of applying the Origami theory to real image understanding is further treated in [Kanade, 1978].

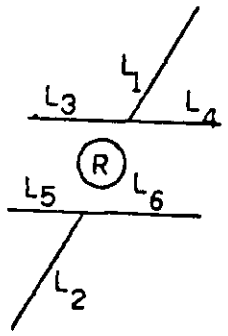


Figure 39 Matched T configuration.

VI Conclusion

The theory of the Origami world (up-to-3-surface junctions) has been developed. The contributions of this paper might be the following:

(1) The concept of selecting surfaces as basic components of the world, rather than the conventional solid polyhedra;

(2) The enumeration of the up-to-3-surface junction labels;

(3) The use of links to capture the global relationships of regions in the form of a Surface Connection Graph;

(4) The filtering procedure defined on the spanning angles;

(5) The discussion of relationships among various worlds dealt with in prior work on polyhedral scene analysis.

It seems that the Origami world defines a subset of interpretations which are interesting both from the standpoint of psychological perception of shapes and from that of practical analysis of region segmented pictures.

Acknowledgements

I would like to thank John Kender, Allen Newell, Raj Reddy, and Steven Shafer for very stimulating discussions in the development of the theory presented in this paper.

REFERENCES

Clowes, M. B. (1971), "On Seeing Things", *Artificial Intelligence*, Vol. 2, No. 1, pp.79-116.

Falk, G. (1972), "Interpretation of Imperfect Line Data as a Three-Dimensional Scene", *Artificial Intelligence*, Vol. 3, No. 2, pp.101-144.

Guzman, A. (1968), "Computer Recognition of Three Dimensional Objects in a Visual Scene", *MAC-TR-59*, MIT.

Horn, B. K. P. (1977), "Understanding Image Intensity", *Artificial Intelligence*, Vol. 8, No. 2, pp.201-231.

Huffman, D. A. (1971), "Impossible Objects as Nonsense Sentences", *Machine Intelligence* Vol. 6 (Meltzer, B. and Michie, D. eds.), Edingburgh University Press, pp.295-323.

Huffman, D. A. (1976), "Curvature and Creases: A Primer on Paper", *IEEE Trans. Computer*, Vol. C-25, No. 10, pp.1010-1019.

Huffman, D. A. (1977), "Realizable Configurations of Lines in Pictures of Polyhedra", *Machine Intelligence* Vol. 8, (Elcock, E. .W. and Michie, D., eds.), Edinburgh University Press, Edingburgh, pp. 493-509.

Kanade, T. (1978), "Task Independent Aspect of Image Understanding", *Proc. Image Understanding Workshop*, Cambridge, Massachusetts, May. Also an extended version will be available as Technical Report, Department of Computer Science, Carnegie-Mellon University.

Mackworth, A. K. (1973), "Interpreting Pictures of Polyhedral Scenes", *Artificial Intelligence*, Vol. 4, No. 2, pp.121-137.

Mackworth, A. K. (1977), "How to See a Simple World: An Exegesis of Some Computer Programs for Scene Analysis", *Machine Intelligence* Vol. 8, (Elcock, E. .W. and Michie, D., eds.), Edinburgh University Press, Edingburgh, pp. 510-537.

Minsky, M. (1974), "A Framework for Representing Knowledge", MIT AI Memo No. 306. Also in *Psychology of Computer Vision* (Winston, P. H., ed.), McGraw Hill, 1975.

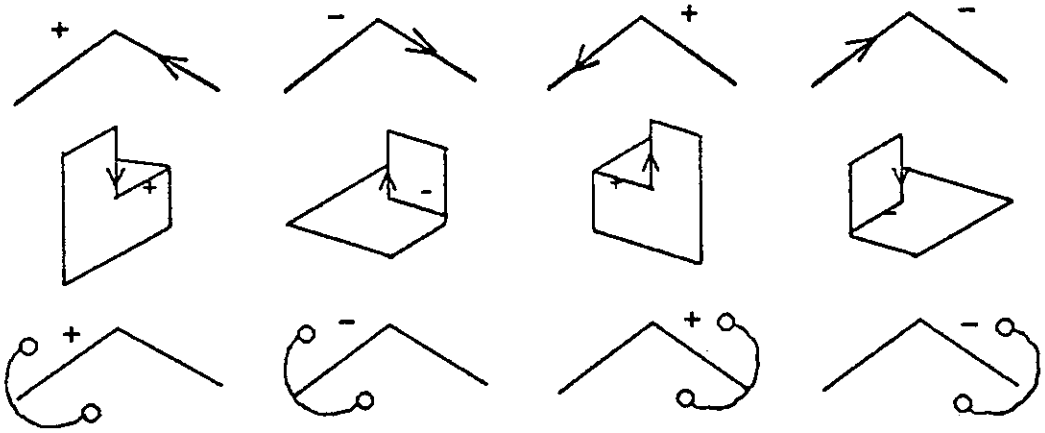
Turner, K., J. (1974), "Computer Perception of Curved Objectes Using a Television Camera", Ph.D. Thesis, School of Artificial Intelligence, Edinburgh University.

Waltz, D. (1972), "Generating Semantic Descriptions from Drawings of Scenes with Shadows", *MAC AI-TR-271*, MIT., reproduced in *The Psychology of Computer Vision*, (Winston, P. ed.), McGraw Hill, 1975.

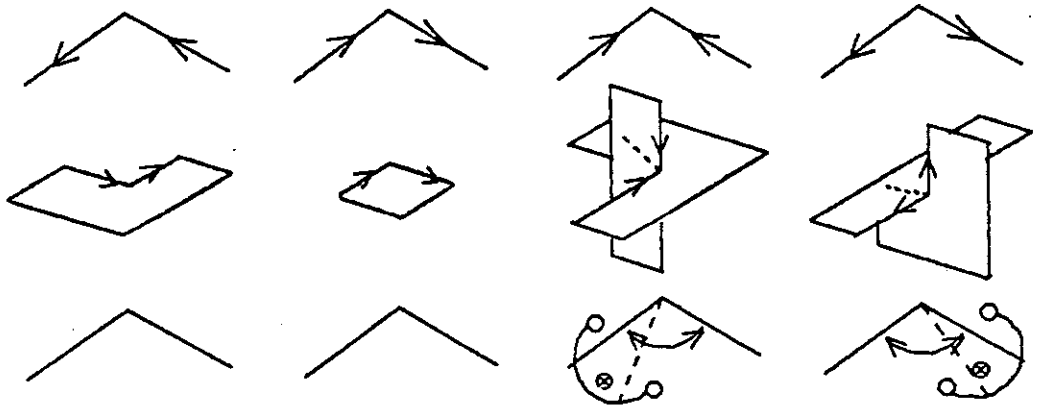
APPENDIX ORIGAMI JUNCTION DICTIONARY

JUNCTION TYPE = L

L1* L2* L3* L4*

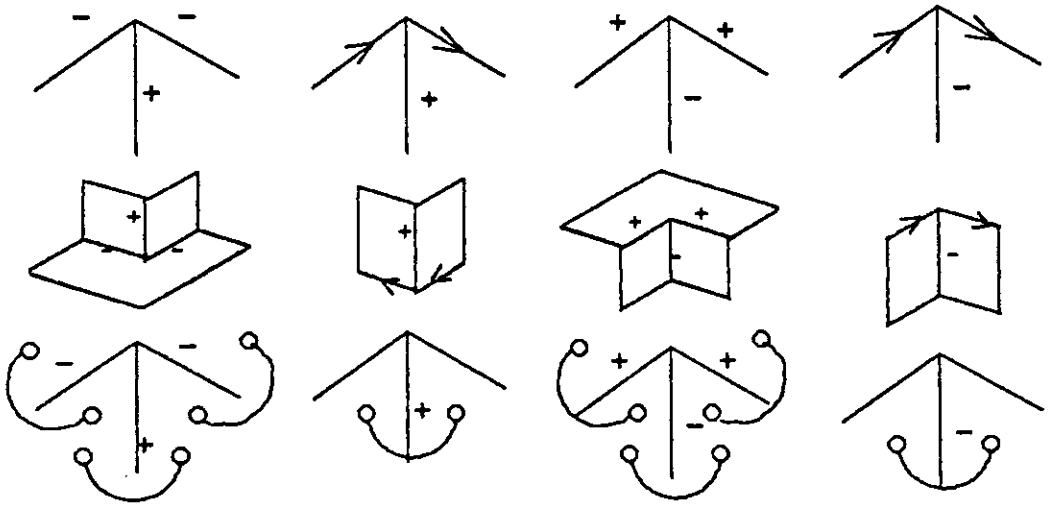


L5* L6* L7 L8



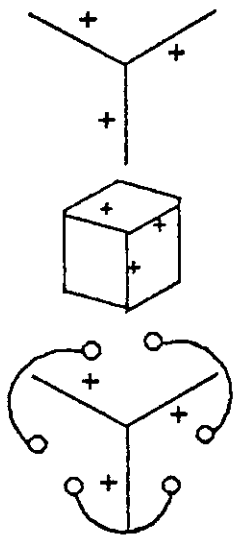
JUNCTION TYPE = ARROW

A1* A2* A3* A4

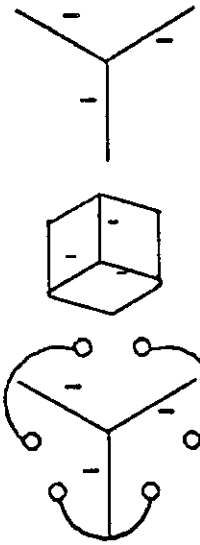


JUNCTION TYPE = FORK

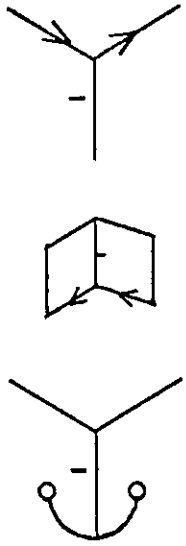
F1*



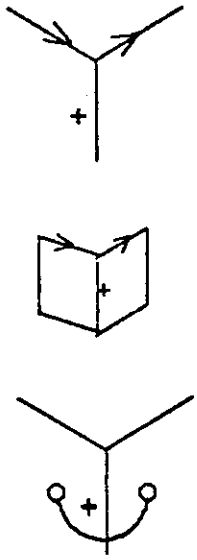
F2*



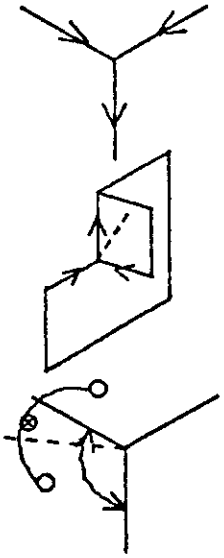
F3 *



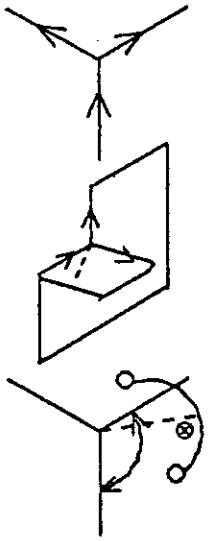
F4



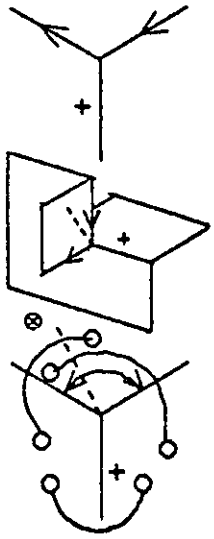
F5



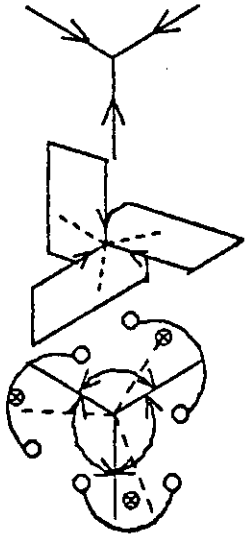
F6



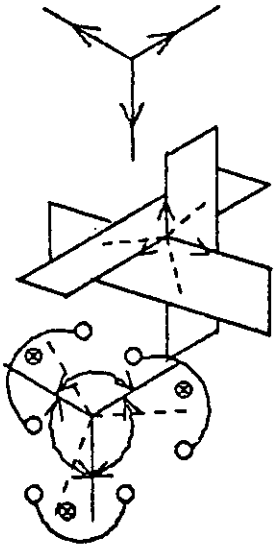
F7



F8

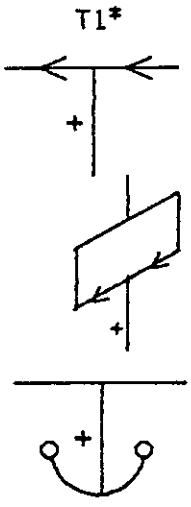


F9

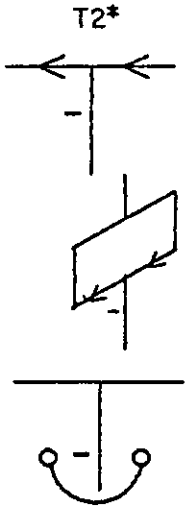


JUNCTION TYPE = T

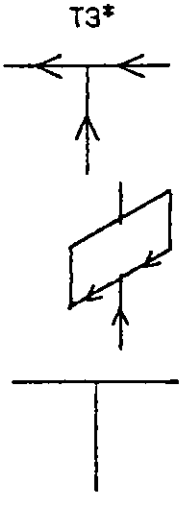
T1*



T2*



T3*



T4*