Recovery of the Three-Dimensional Shape of an Object from a Single View*

Takeo Kanade
Department of Computer Science, Carnegie–Mellon University, Pittsburgh, PA 15213, U.S.A.

ABSTRACT

Given a single picture which is a projection of a three-dimensional scene onto the two-dimensional picture plane, we usually have definite ideas about the 3-D shapes of objects. To do this we need to use assumptions about the world and the image formation process, since there exist a large number of shapes which can produce the same picture.

The purpose of this paper is to identify some of these assumptions—mostly geometrical ones—by demonstrating how the theory and techniques which exploit such assumptions can provide a systematic shape-recovery method. The method consists of two parts. The first is the application of the Origami theory which models the world as a collection of plane surfaces and recovers the possible shapes qualitatively. The second is the technique of mapping image regularities into shape constraints for recovering the probable shapes quantitatively.

Actual shape recovery from a single view is demonstrated for the scenes of an object such as a box and a chair. Given a single image, the method recovers the 3-D shapes of an object in it, and generates images of the same object as we would see it from other directions.

1. Introduction

It is a common experience for us that, given a single two-dimensional picture of an object, we have definite ideas about its three-dimensional shape, in spite of the fact that a large number of possible shapes exist which can produce the same picture. This fact indicates that we use some assumptions or knowledge about the objects and about image formation. The purpose of this paper is to identify some of these assumptions—mostly geometrical ones—by demonstrat-

*This research was sponsored by the Defence Advanced Research Projects Agency (DOD), ARPA Order No. 3597, and monitored by the Air Force Avionics Laboratory under Contract F33615-78-C-1551. The views and conclusions contained in this document are those of the author, and should not be interpreted as representing the official policies, either expressed or implied, of the Defence Advanced Projects Agency or the U.S. Government.

Artificial Intelligence **17** (1981) 409–460
0004–3702/81/0000–0000/$02.50

ing how the theory and techniques which exploit such assumptions can provide a systematic shape-recovery process.

The process consists of two parts: qualitative shape recovery and quantitative shape recovery. The first part uses a model of the Origami world [8]. It labels a line drawing and recovers the geometrically possible shapes by exploiting the assumption of planar-surfaced objects. When images, either in monochrome or in color, are given, edge profiles taken across lines in the image can be used in order to constrain line labels in the search of plausible interpretations. The second part adopts a technique of mapping image regularities (in particular, parallelism of lines and 'skewed symmetry') into shape constraints. It quantitatively recovers the probable shapes by exploiting the assumptions which exclude accidental alignments and regularities in the picture.

Actual shape recovery from a single view is demonstrated for the scenes of an object such as a box and a chair. Given an image, the shape recovery process generates a 3-D shape description of the object in terms of plane surfaces, and the description is supplied to a display program which can synthesize images of the same object as we would see it from other view directions. Throughout the paper we will assume orthographic projection rather than perspective projection.

In brief, the paper is outlined as follows: in the next section the problem of shape recovery from a single view is addressed. It is shown that the shape recovery can be either qualitative or quantitative. The previous research is briefly reviewed: what results have been obtained from what assumptions. Section 3 presents example scenes which are used throughout the paper. Sections 4 through 6 provide descriptions of tools and basic theories: the gradient space in Section 4; the theory of the Origami world in section 5; and the mapping of image regularities into constraints in the gradient space in Section 6. Then, in Section 7 these are put together to obtain quantitative 3-D shape descriptions. Example scenes are processed. Sections 8–10 deal with the shape recovery from a real image using an example of 'chair' scene: especially use of color edge profiles is presented in Section 9 to reduce the number of possible labelings. Section 11 discusses implications of the results in a broader context.

2. The Problem of Shape Recovery

2.1. Qualitative vs. quantitative shape recovery

Suppose that the drawing of Figs. 1(a) and 1(b) are given. Most commonly, they both appear as the corner of a solid object, made of three plane surfaces, coming out toward the viewer. This level of descriptions qualitatively characterizes the shape of objects. In this sense, the two figures in Fig. 1 are qualitatively equivalent.

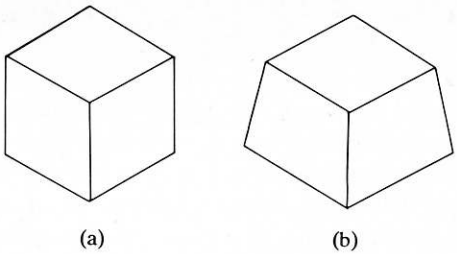


FIG. 1. Simple line drawings: (a) a 'cube' scene; (b) a 'trapezoid-block' scene.

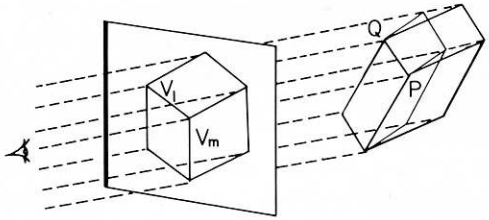


FIG. 2. Many shapes can produce the line drawing of Fig. 1(a).

Usually, though, the pictures seem to convey more quantitative shape information. Fig. 1(a) appears as a right-angled block, whereas Fig. 1(b) appears as a trapezoidal block, both viewed from the same direction with respect to the top face. That is, we feel that we can recover the surface orientations with respect to the view direction, or we can quantitatively describe the shape. Let us call this the quantitative shape recovery. (Note, however, that we cannot know any absolute information, such as the absolute distance to the object or its absolute size without knowing camera parameters.)

Notice that the shapes of our example scenes need not be as described above. Take the object in Fig. 1(a) for example. It need not be a cube. As shown in Fig. 2, the object can be any of the shapes made by three plane segments which intersect at P on the middle view line (which goes through the middle junction V_m in the picture), and the intersection edge (e.g. PQ) of each pair lies in the plane defined by the corresponding line in the picture (e.g. V_mV_1) and the view line. We will see later that the object in Fig. 1(a) can be of even a qualitatively different shape; i.e., it may not necessarily be a corner of a block. These considerations suggest that the assumptions we use in our shape recovery process are very strong ones.

2.2. Related work

Stereoscopic shape recovery has been studied quite extensively, but there is not much work on shape recovery from a single view, especially on quantitative shape recovery.

Interpretation of line drawings as 3-D scenes is the research domain which has been most actively studied. Guzman [3] first defined types of junctions (e.g., ELL, ARROW, and FORK), and developed many heuristics concerning probable association of regions suggested by each junction type. His program SEE decomposes a given line drawing into object regions, but it works totally in the 2-D picture domain without any explicit 3-D shape representation.

Huffman [6] and Clowes [1] gave an important theoretical framework. They observed that lines in the picture can have different physical meanings: a line can be a convex edge (signified by a label +), a concave edge (signified by −), and an occluding edge (signified by ↑ with the occluding region on its right side). In the trihedral world in which exactly three planes are assumed to meet at every corner, the possible combinations of line labels for each junction type can be catalogued in a dictionary. To find 3-D shapes, a labeling procedure can assign line labels to the lines in a given line drawing according to the dictionary. Waltz [18] extended this idea further to include cracks, shadows, etc., and devised an efficient labeling procedure, called filtering. This body of work can be named *qualitative* analysis for *qualitative* shape recovery, because they used tools, such as junction types and line labels, which only qualitatively describe 3-D shapes of polyhedra.

The next group is *quantitative* analysis for qualitative shape recovery. Huffman [6] introduced the gradient space to represent surface orientations. Mackworth [12] employed it as a central tool to test the consistency of labelings by the method of constructively locating the gradient of each surface. In the labeling procedure of the Origami world [8], employed in this paper, the constraints in the gradient space are all maintained and tested symbolically on a graph called a Surface Connection Graph. Theoretically the constraints in the gradient space provide only necessary conditions for planar realizability of a labeled line drawing. Huffman [7] presents a $\phi(\phi')$-point test as the necessary and sufficient condition for a 'cut set' of labeled lines (equivalently, set of regions incident to those lines). Still, unfortunately, it is not the sufficient condition for a whole configuration, though. Falk [2] and Sugihara [17] directly investigated the algebraic properties of the linear equation system which represents the projection from a 3-D space object to the 2-D picture plane: Falk related degrees of freedom in a projection with the concept of mergeability of his Face Adjacency Graph; Sugihara presented the conditions for realizability of a labeled line drawing including hidden lines. All these researches quantitatively analyze a picture for its possible interpretations as 3-D scenes, but the recovered shape is still qualitative.

For *quantitative* shape recovery from a single picture we need to introduce more assumptions. The support hypothesis, first described by Roberts [15], is an example of such an assumption, which enables us to determine the depth (distance) of object points on a table. Mackworth [13] assumed rectangularity at every corner to uniquely determine the gradients of surfaces.

More commonly, model descriptions of objects which can appear in the scene are given to the system, and the system interprets a picture by identifying (possibly modified) occurrences of the models. In his pioneering work [15], Roberts used models of a cube, a wedge, and a hexagonal prism represented as a set of 3-D space coordinates of their vertices. The models are scaled, rotated, translated, and projected to test matching with the junctions in the picture. Falk's INTERPRET [2], which interprets imperfect line drawings, has fixed-size models of nine prototypes.

When one can assume the conditions in the image formulation process (such as the imaging geometry, the illumination, and the surface photometry, etc.), the intensity of a picture point can provide constraints on the surface orientation at the corresponding 3-D point. Horn [5] formulated a basic shape-from-shading theory employing the gradient-space representation. Woodham [19] identified two general rules (uniqueness and continuity) and demonstrated the recovery of local orientations in several cases such as a Lambertian sphere and a cone.

The quantitative shape recovery in this paper does not assume a predefined set of objects. Our approach resembles Mackworth's method and Horn–Woodham's shape-from-shading method in that we seek a unique determination of gradients, and in that some picture properties are mapped into constraints in the gradient space. But we focus on geometrical properties of objects and of the picture-taking process, and our assumptions are much more general than Mackworth's rectangularity assumption.

3. Example Scenes

The example scenes we will use in the rest of the paper are the following.

- 'Cube' scene: a line drawing (Fig. 1(a)).
- 'Trapezoid-block' scene: a line drawing (Fig. 1(b)).
- 'W-folded paper' scene: a line drawing (Fig. 3).
- 'Box' scene: a line drawing (Fig. 4).
- 'Chair' scene: a color image and a line drawing (Figs. 5(a) and 5(b)).

Each has its own difficulty, which illustrates an inherent problem in the shape recovery. The names of the scenes are given just for the purpose of referencing

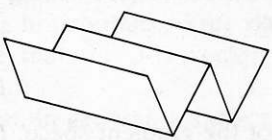


FIG. 3. A line drawing of a 'W-folded paper' scene.

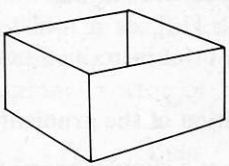
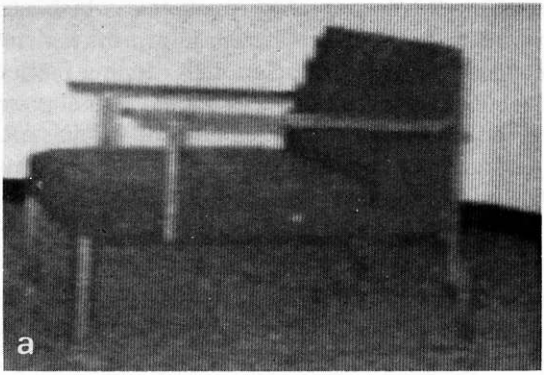
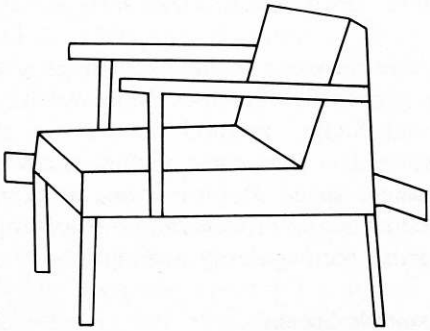


FIG. 4. A line drawing of a 'box' scene. Though it looks 'perfect', the trihedral labeling does not work for it.



(a)



(b)

FIG. 5. A 'chair' scene: (a) a halftone image; (b) a line drawing.

them. They never imply any assumptions about the semantic nature of the scene.

4. The Gradient Space: Introduction

In this paper we will use the gradient space, popularized by Huffman [6] and Mackworth [12], as a tool for representing surface orientations. This section provides a brief introduction to the gradient space.

4.1. Definition of the gradient space

We will give a simplest version of the definition of the gradient space. Let Fig. 6(a) be the geometry involving the viewer, the picture plane, and the object in the scene. The z-axis is taken as parallel with the view line, and the x–y plane is on the picture plane, with the x-axis pointing to the horizontal right. We

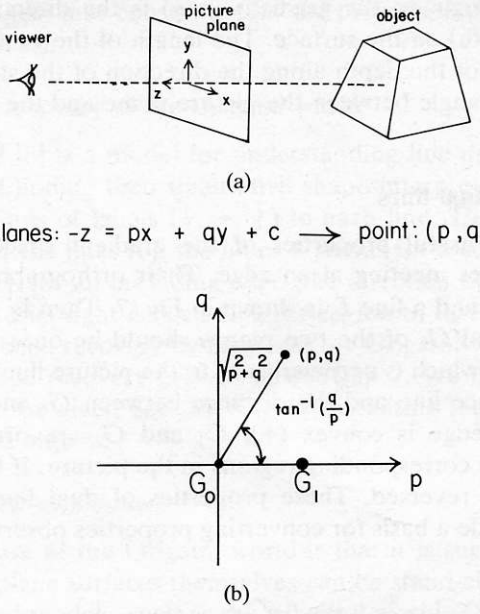


FIG. 6. The gradient space: (a) geometry including the object, the picture, and the viewer; (b) mapping of planes to a gradient.

assume orthographic projection here. A plane in the scene whose surface is visible from the viewer can be expressed as

$$-z = px + qy + c\,. \tag{1}$$

The two-dimensional space made of the ordered pairs (p, q) is called the *gradient space* G (Fig. 6(b)). Let us assume for our convenience that we align the directions of the coordinates of (x, y) in the picture with those of (p, q). All the planes in the scene that have the same values of p and q (i.e., the same orientation) are mapped into the point (p, q), called the *gradient*, in G.

The gradient (p, q) represents how the planes are slanting relative to the view line (z-axis). For example, the origin $G_0 = (0, 0)$ of G corresponds to those planes $(-z = c)$ which are perpendicular to the view line. Points on the positive p-axis, $(p, 0)$ with $p > 0$, correspond to the planes $(-z = px + c, p > 0)$ which are slanting horizontally to the right: the larger p is, the more slanted. From the (1) we see that the 3-D vector $(p, q, 1)$ is the directional vector of the surface normal.

When, in general, a surface is represented as $-z = f(x, y)$, then

$$p = \partial(-z)/\partial x\,, \quad \text{and} \quad q = \partial(-z)/\partial y\,, \tag{2}$$

which is why (p, q) is called the gradient. Thus the direction $(\tan^{-1}(q/p))$ of the

vector from the origin to the gradient (p, q) is the direction of the steepest change of $-z$ (depth) on the surface. The length of the vector $(p^2+q^2)^{1/2}$ is the rate of the change of the depth along the direction of the steepest change; i.e., the tangent of the angle between the picture plane and the planes corresponding to (p, q).

4.2. Properties of dual lines

One of the most useful properties of the gradient space is the following. Consider two planes meeting at an edge. Their orthographic picture made of regions R_1 and R_2 and a line L is shown in Fig. 7. Then in the gradient space, the gradient G_1 and G_2 of the two planes should be on a gradient-space line (called a dual line) which is perpendicular to the picture line L, but the location of the gradient-space line and the distance between G_1 and G_2 are arbitrary. 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. These properties of dual lines are quite useful because they provide a basis for converting properties observable in the picture

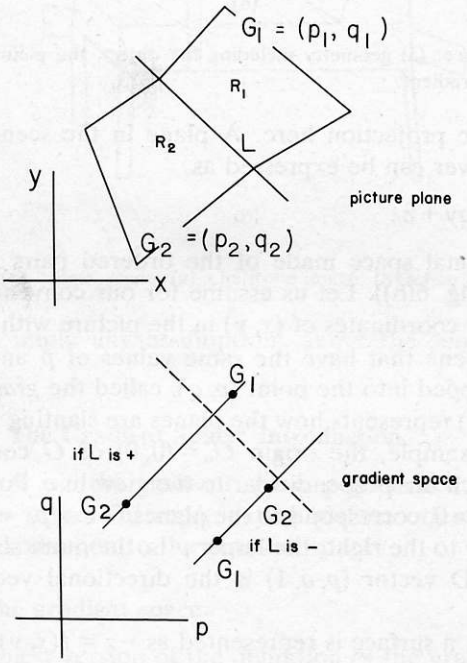


FIG. 7. Properties of dual lines. If two planes meet and the intersection line is projected to a picture line L, then the gradients of the two planes are on a gradient-space line which is perpendicular to L.

(the directions of lines) into constraints on the properties of a real object (the orientation of surfaces).

5. A theory of the Origami world

The Origami world [8] is a model for understanding line drawings in terms of plane surfaces, and finding their qualitative shape interpretations by assigning one of the three kinds of labels (+, −, ↑) to each line. The labels signify the physical meaning of the lines [6]: the label + stands for a convex edge, − for a concave edge, and ↑ for an occluding edge (the direction of the arrow is given so that the region to its right is occluding the region to its left).

The qualitative shape recovery by means of the Origami-world labeling is the first part of the shape recovery process in this paper. We find possible shapes which can produce the given line drawing by exploiting the basic assumptions on the properties of objects.

5.1. Surface-oriented assumption

An important feature of the Origami world is that it is surface oriented. That is, it assumes that plane surfaces themselves can be stand-alone objects, unlike solid-object oriented models, such as the trihedral world. This idea can be best illustrated by Fig. 4. Though it appears perfect, the Huffman–Clowes–Waltz labeling scheme for the trihedral world cannot handle it, and would call it an 'impossible' object. The reason for this failure is that the trihedral world assumes that the object is solid, and thus the line drawing of a box would need to be 'super' perfect, as in Fig. 8, in order for it to be handled.

The assumptions concerning the geometrical configuration around a vertex in the Origami world are as follows. No more than three planar surfaces of different orientations meet at a vertex, and no more than three edges of different directions are involved at a vertex. The combination of the three orientations are assumed to be 'general' in the sense that they span the three-dimensional space (i.e., each orientation has a vector component perpendicular to the other two). 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-

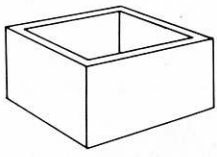
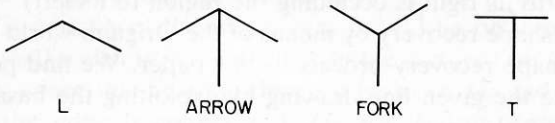


FIG. 8. A 'super-perfect' line drawing of the 'box' scene for the trihedral world.

TABLE 1. Comparison of the sizes of junction dictionaries for L, ARROW, FORK, and T junction types. The Origami world dictionary includes other junction types, such as K, X, and PSI

Junction type	Huffman–Clowes dictionary	Origami world dictionary
L	6	8
ARROW	3	15
FORK	3	9
T	4	16



and 2-surface vertices; that is, it allows free extending surfaces as stand-alone objects.

5.2. Junction labels for up-to-3-surface vertices

Once we recognize the basic assumptions of the Origami world, we can generate a junction dictionary as in the Huffman–Clowes–Waltz theory, which contains possible junction labels (i.e., possible line-label combinations) for each junction type. The size of the dictionary shown in Table 1 gives an idea of the degree of constraints imposed by the Origami world compared with the Huffman–Clowes trihedral world.

5.3. Augmented junction dictionary

A legal junction label represents a possible configuration of surfaces at a vertex. The consistent labeling of a line drawing, so that all the junctions are given legal junction labels, is nothing but a check on the consistency of surface interconnections: this is done by passing information by means of line labels from one junction to another. Waltz' filtering on junction labels is known to be a good method for doing this. However, labeling in the Origami world cannot simply rely on the filtering on junction labels. More thorough and global consistency checks concerning surface orientations are necessary. Because of the weaker restrictions at the vertices than the trihedral world, the anomalies caused by solely relying on junction labels show up as a more serious problem; i.e., the anomalous interpretations in which the labeling is consistent but the whole configuration is not possible [12, 14].

To remedy this, the junction dictionary for the Origami world can be augmented. To each junction label is attached the information as to what

constraints in the gradient space should be satisfied by the surfaces incident at the junction. As shown in Fig. 9(a), the constraints are represented by the links which connect a pair of related regions and which include information about the constraints on their gradients. The properties of dual lines explained in Section 4 are used here.

In such a junction label as shown in Fig. 9(b), the intersection line of surfaces is occluded. This junction label is typically a result of folding a sheet of a paper along BC: R_1, which is folded toward the viewer, occludes a part of R_2. But we can assume that this junction label represents slightly more general cases in which the intersection line of R_1 and R_2 lies within the angle ABC. (See the middle figure of Fig. 9(b)). That is, we assume that if we remove the right hand part of R_1 which is occluding R_2, then the rest of R_1 and R_2 will form a convex intersection line, and the line can be anywhere in the angle ABC. We call it an occluded intersection line and denote it by the label $\oplus$. Therefore, the associated link represents that 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.

Notice the crucial difference of our links from those Guzman [3] used. Guzman's links are for associating regions in the picture, whereas ours are for describing the relations between the corresponding surfaces in the scene.

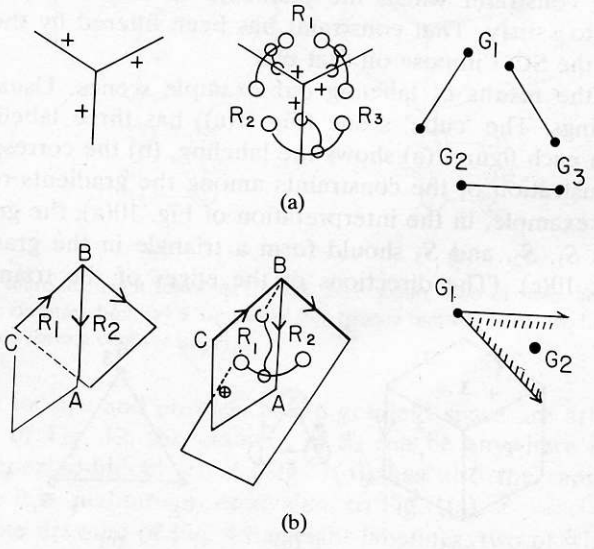


FIG. 9. The augmented junction dictionary for the Origami world. Two examples of entries are shown. For each, the first column is the junction label, the second associated links, and the third is the illustration of the relationships of gradients represented by the links.

5.4. Labeling procedure and surface connection graph

The labeling procedure of the Origami world uses the augmented dictionary. First, the Waltz filtering on junction labels is performed. Then the procedure begins to assign a junction label to each junction one by one. When a junction label is assigned to a junction, the constraints represented by the associated links are instantiated by using the directions of the lines at that particular junction.

The labeling procedure of the Origami world tests the consistency of surface orientations by using these instantiated constraints. In order to perform the test systematically, the labeling procedure constructs a graph called a *Surface Connection Graph* (SCG). The SCG is a labeled graph where a node represents a surface and an arc represents a constraint between the surfaces. It indicates what surfaces are connected with what surfaces through what constraints.

The test can be performed by using an iterative filtering operation in the gradient space defined on the SCG. One feature of this operation is that all the constraints are maintained symbolically in the SCG during the computation. The details can be found in [8].

5.5. The results of labeling the example scenes

As a result of the labeling procedure of the Origami world, we obtain not only a labeled line drawing, but also a filtered SCG. Each arc of the resultant SCG represents the constraint which the gradients of the corresponding pair of surfaces have to satisfy. That constraint has been filtered by those which the other parts of the SCG impose on that pair.

Let us see the results of labeling our example scenes. Usually we obtain multiple labelings. The 'cube' scene (Fig. 1(a)) has three labelings shown in Figs. 10–12. In each figure, (a) shows the labeling, (b) the corresponding SCG, and (c) the illustration of the constraints among the gradients represented by the SCG. For example, in the interpretation of Fig. 10(a), the gradients of the three surfaces S_1, S_2, and S_3 should form a triangle in the gradient space as shown in Fig. 10(c). (The directions of the edges of the triangle should be

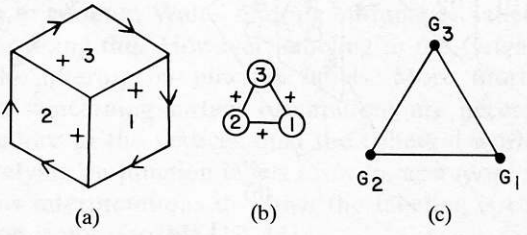


FIG. 10. The first labeling for the 'cube' scene: (a) labeling; (b) SCG; (c) illustration of the constraints on the gradients. This represents a convex corner.

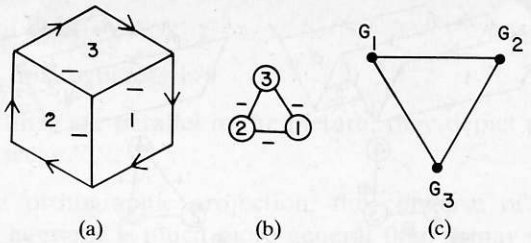


FIG. 11. The second labeling for the 'cube' scene. This represents a concave corner.

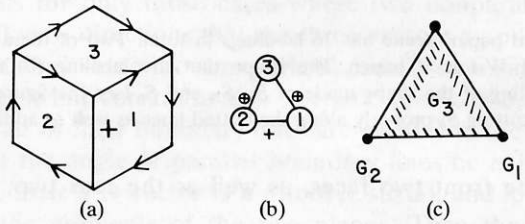


FIG. 12. The third labeling for the 'cube' scene. This represents a shape in which two surfaces S_1 and S_2 are connected along a convex edge, and occlude the third plane S_3 partially.

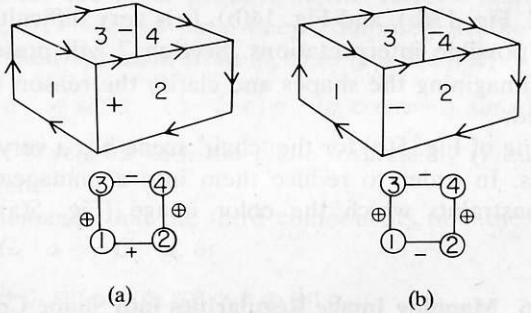


FIG. 13. In total there are eight labelings for the 'box' scene. Two of them are shown here: (a) corresponds to an ordinary box; (b) a 'squashed' box (notice that the front two faces, as well as the rear two, go in and form a concave edge).

as shown, but its size and position in the gradient space are arbitrary.) In the interpretation of Fig. 12, the gradient of S_3 can be anywhere in the hatched area. The 'trapezoid-block' scene (Fig. 1(b)) has also the same set of three labelings since it is qualitatively equivalent to Fig. 1(a).

The 'box' line drawing of Fig. 4 has eight labelings, two of which are shown in Fig. 13: The labeling in Fig. 13(a) corresponds to an 'ordinary box': the two front faces form a convex intersection and partially occlude the rear two faces which form a concave intersection. The labeling in Fig. 13(b) corresponds to a

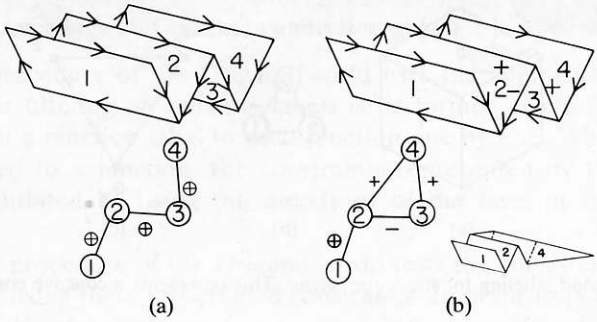


FIG. 14. The 'W-folded paper' scene has 16 labelings in total. Two of them are shown here (a) corresponds to a real W-folded paper. The shape that the labeling (b) represents might be imagined by first thinking of the shape made of S_1, S_2, and S_4 (see the figure, and imagine S_4 as horizontal), and then cutting S_4 properly along the dotted lines as well as adding S_3.

'squashed box': the front two faces, as well as the rear two, form a concave intersection.

The 'W-folded paper' scene of Fig. 3 has 16 labelings, two of which are shown in Fig. 14: (a) corresponds to a real W-folded paper, but (b) is a peculiar shape.

In interpreting these line drawings we do not usually think of such peculiar shapes as Fig. 12, Fig. 13(b), and Fig. 14(b). It is very difficult even to imagine those shapes as possible interpretations. Section 7 will present their rotated views to help in imagining the shapes and clarify the reason why they are not usually considered.

The line drawing of Fig. 5(b) for the 'chair' scene has a very large number of possible labelings. In order to reduce them into a manageable size, we will exploit other constraints which the color image (Fig. 5(a)) provides. (See Section 9.)

6. Mapping Image Regularities into Shape Constraints

The meta-heuristic of nonaccidental regularities exploited in this section is the following:

> "Regularities observable in the picture are not by accident, but are some projection of real regularities."

The technique for expointing this meta-heuristic is the mapping of the image regularities into the constraints in the gradient space. In this paper, parallelism of lines and skewed symmetry are particularly used as image regularities. This is a part of the theory and technique developed by Kanade and Kender [10]. The justification, generalization and other applications are found there.

6.1. Parallelism of lines

The heuristic for this regularity is

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

Since we assume orthographic projection, the converse of this heuristic is always true. The heuristic is much more general than it may sound. Referring to Fig. 15(a), imagine that two lines in the 3-D space, which are denoted by vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, are viewed from the direction which is denoted by a vector $\boldsymbol{e}$. The heuristic fails for only those cases where two nonparallel lines are seen from such special view directions that the three vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, and $\boldsymbol{e}$, if properly translated so that they share the same origin, can lie on a single plane.

Let us consider what constraint this heuristic provides on the gradients of two planes if a pair of their boundary lines are parallel in the picture, as shown in Fig. 15(b). Let the angle of parallel boundary lines be α in the x–y picture coordinates; i.e., their 2-D vector is $\boldsymbol{a} = (\cos\alpha, \sin\alpha)$, and let $\boldsymbol{G}_1 = (p_1, q_1)$ and $\boldsymbol{G}_2 = (p_2, q_2)$ be the gradients of the two planes. Then the 3-D vectors corresponding to these boundary lines are

$$(\cos\alpha, \sin\alpha, -\boldsymbol{G}_1 \cdot \boldsymbol{a}) \quad \text{and} \quad (\cos\alpha, \sin\alpha, -\boldsymbol{G}_2 \cdot \boldsymbol{a})$$

where $\langle \cdot \rangle$ stands for the inner product of the vectors. Here the third components of these 3-D vectors have been computed as the increment of z corresponding to the translation from $(0, 0)$ to $(\cos\alpha, \sin\alpha)$ in the picture:

$$(-p_i \cos\alpha - q_i \sin\alpha - c) - (-c) = -(p_i \cos\alpha + q_i \sin\alpha) = -\boldsymbol{G}_i \cdot \boldsymbol{a},$$

since the gradient is related with the plane formula by (1) and the boundary line lies on the plane.

The heuristic demands that the third components of these two 3-D vectors are equal, i.e., $-\boldsymbol{G}_1 \cdot \boldsymbol{a} = -\boldsymbol{G}_2 \cdot \boldsymbol{a}$, or

$$p_1 \cos\alpha + q_1 \sin\alpha = p_2 \cos\alpha + q_2 \sin\alpha . \tag{3}$$

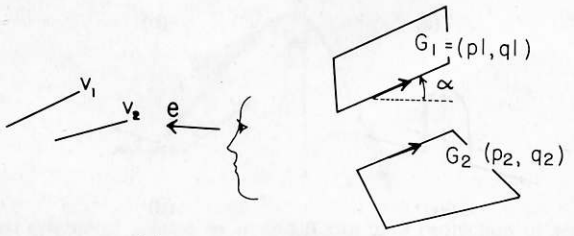


FIG. 15. Parallel-line heuristic: (a) If $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, and $\boldsymbol{e}$ can be on the same plane in the 3-D space by translating them, even when $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are not parallel, then the heuristic fails; (b) Two planes having parallel boundary lines.

This equation represents the fact that (p_1, q_1) and (p_2, q_2) are on a gradient-space line which is perpendicular to the direction $\boldsymbol{a} = (\cos\alpha, \sin\alpha)$. Alternatively, this condition can be intuitively understood as follows. If a pair of boundary lines is really parallel in the 3-D space, we can translate one of the planes toward the other, and make the two planes intersect along those boundary lines. Therefore, the gradients of the two planes should have the above relationship (3) because of the property of dual lines.

6.2. Skewed symmetry

Symmetry in a 2-D picture has an axis for which the opposite sides are reflective: in other words, the symmetrical property is found along the transverse lines perpendicular to the symmetry axis. The concept *skewed symmetry* relaxes this condition a little. It means a class of 2-D shapes in which the symmetry is found along lines not necessarily perpendicular to the axis, but at a fixed angle to it. Figs. 16(a–c)[1] show a few examples. Formally, such shapes are defined as 2-D linear (affine) transformations of real symmetries. A skewed symmetry defines two directions as shown in Figs. 16(a–c): let us call them the skewed-symmetry axis and the skewed transverse axis. Stevens [16] does not use the concept of the skewed symmetry, but he presents a good body of psychological experiments which suggest that human observers can perceive surface orientations from figures with this property.

A particular heuristic about this image regularity is

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

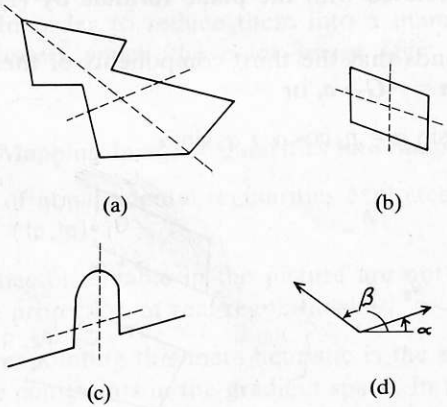


FIG. 16. Skewed symmetry. (a–c) are examples. (d) A skewed symmetry defines two directions: the skewed-symmetry axis and the skewed transverse axis.

[1]The mouse hole example (Fig. 16(c)) was given by K. Stevens.

Again, the converse of this heuristic is always true in the orthographic projection. We can transform this heuristic into constraints in the gradient space. As shown in Fig. 16(d), let α and β denote the directional angles of the skewed-symmetry axis and the skewed transverse axis, respectively. Let $\boldsymbol{G} = (p, q)$ be the gradient of the plane which includes the skewed symmetry. The 3-D vectors on the plane corresponding to the directions α and β are

$$\boldsymbol{A} = (\cos\alpha, \sin\alpha, -\boldsymbol{G}\cdot\boldsymbol{a}) \quad \text{and} \quad \boldsymbol{B} = (\cos\beta, \sin\beta, -\boldsymbol{G}\cdot\boldsymbol{b})$$

where $\boldsymbol{a} = (\cos\alpha, \sin\alpha)$ and $\boldsymbol{b} = (\cos\beta, \sin\beta)$. The heuristic demands these two 3-D vectors to be perpendicular, i.e., $\boldsymbol{A}\cdot\boldsymbol{B} = 0$, or

$$\cos(\alpha-\beta) + (\boldsymbol{G}\cdot\boldsymbol{a})(\boldsymbol{G}\cdot\boldsymbol{b}) = 0 \tag{4}$$

By rotating the p–q coordinates by the amount $\lambda = (\alpha+\beta)/2$ into the p'–q' coordinates, it is easy to show that

$$p'^2\cos^2((\alpha-\beta)/2) - q'^2\sin^2((\alpha-\beta)/2) = -\cos(\alpha-\beta) \tag{5}$$

where

$$p' = p\cos\lambda + q\sin\lambda$$

$$q' = -p\sin\lambda + q\cos\lambda \quad \text{and} \quad \lambda = (\alpha+\beta)/2\,.$$

Thus, (p, q)'s are on a hyperbola shown in Fig. 17. That is, the skewed

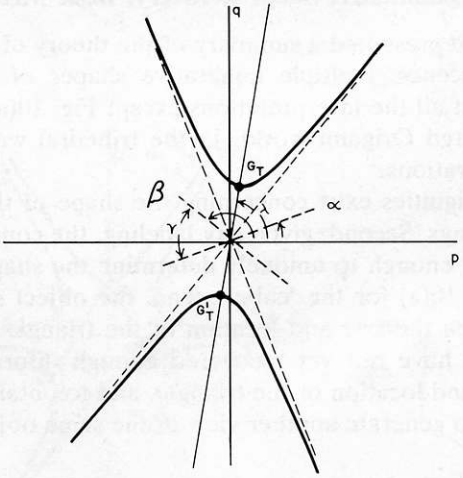


FIG. 17. A skewed symmetry defined by α and β can be a projection of a real symmetry if the gradient of the surface is on the hyperbola. The axis of the hyperbola is the bisector of the obtuse angle made by α and β. The asymptotes are perpendicular to the direction of α and β. The length from the origin to the tips is $(1-2d^2)^{1/2}/d$ where $d = \sin(\gamma/2)$ and γ = the acute angle made by α and β.

symmetry defined by α and β in the picture can be a projection of a real symmetry if and only if the gradient is on this hyperbola. An impression might be that if we assume the skewed symmetry in the picture to be a projection of a real symmetry, the surface orientation is uniquely determined, but actually we have still an infinite number of possible orientations represented by the gradients on the hyperbola. As a special case, if $|\alpha - \beta| = \pi/2$ (i.e., the skewed symmetry is now a usual symmetry), the hyperbola degenerates into two perpendicular lines passing through the origin: $\boldsymbol{G} \cdot \boldsymbol{a} = 0$ and $\boldsymbol{G} \cdot \boldsymbol{b} = 0$.

The tips or vertices G_T and G_T' of the hyperbola represent special orientations with interesting properties. Since they are closest to the origin of the gradient space, and since the distance from the origin to a gradient represents the magnitude of the surface slant, G_T and G_T' correspond to the least slanted orientations that can produce the skewed symmetry in the picture from a real symmetry in the scene. Also, since they are on the line (the axis of the hyperbola) which bisects the angle made by α and β, they correspond to the orientations for which the length metrics along the directions of α and β in the picture are equal; i.e., the ratio of lengths along them in the picture represents the real ratio. (See Appendix A for the proof.) In Fig. 17, G_T corresponds to looking down to the surface, and G_T' to looking up to it. If no other constraints are available, either one of these gradients may be the most reasonable selection as the surface orientation.

7. Quantitative Shape Recovery: Basic Method

In Section 5 we have presented a summary of the theory of the Origami world. For our example scenes, multiple qualitative shapes of objects have been recovered. Note that all the interpretations except Fig. 10(a) are obtained only in the surface-oriented Origami world: in the trihedral world they are called 'impossible' configurations.

Still a lot of ambiguities exist concerning the shape of the object. First, we have multiple labelings. Second, given any labeling, the constraints represented by the SCG are not enough to uniquely determine the shape. For example, in the labeling of Fig. 10(a) for the 'cube' scene, the object still can have many shapes, depending on the size and location of the triangle of Fig. 10(c) in the gradient space. We have not yet recovered enough information to uniquely determine the size and location of the triangle, and to obtain a 3-D description of the object so as to generate another view of the same object from a different angle.

Section 6 has described how some image regularities can be mapped into constraints in the gradient space. They provide additional constraints for determining the surface orientations.

In this section we will show that by putting these together we can recover quantitative shapes; i.e., we can assign a unique gradient to each surface.

7.1. Unique determination of gradients: Simple cases

7.1.1. *'Cube' scene*

Take the line drawing of Fig. 1(a). Fig. 18(a) shows the angles and the lengths. It has three labelings (Figs. 10–12). Consider the first one shown in Fig. 10(a) and reproduced in Fig. 18(b). The labeling indicates that there are three totally-visible surfaces S_1 $(= V_3V_4V_7V_2)$, S_2 $(= V_5V_6V_7V_4)$, and S_3 $(= V_1V_2V_7V_6)$, and that their gradients, G_1, G_2, and G_3, should form a triangle as shown in Fig. 10(c). On the other hand, S_1, S_2, and S_3 have skewed symmetries: their skewed-symmetry axes and skewed transverse axes are shown in Fig. 18(b) as dotted lines. If we assume these skewed symmetries to be projections of real symmetries, we can draw the hyperbola for each surface as shown in Fig. 18(c).

Now the problem is where we can locate the triangle of Fig. 10(c) in Fig. 18(c) so that each vertex of the triangle is on the corresponding hyperbola. It is

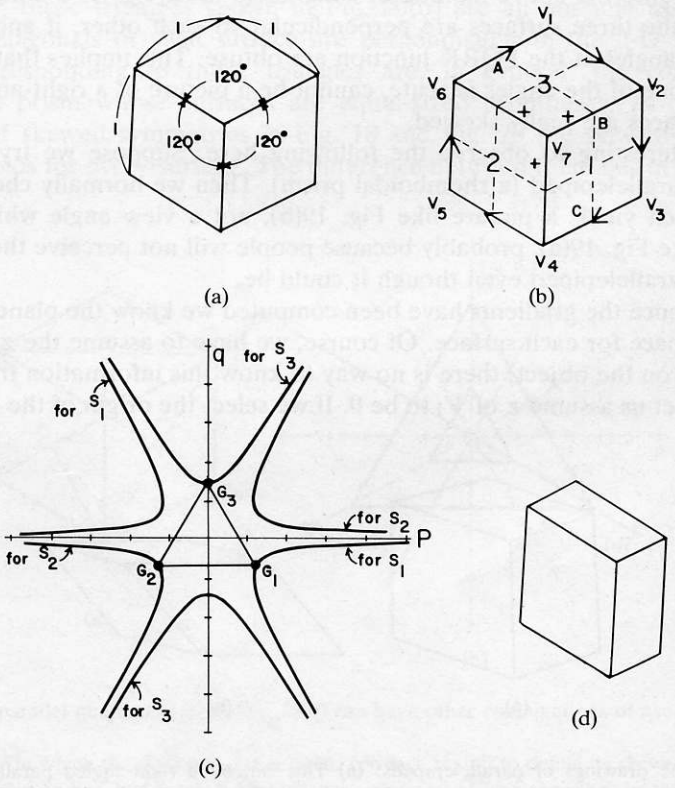


FIG. 18. Quantitative shape recovery for the 'cube' scene for the labeling of Fig. 10.

not difficult to prove that the location shown in Fig. 18(c) is the only possibility (see Appendix B for the proof). The gradients of S_1, S_2 and S_3 are thus uniquely determined as

$$G_1 = (\sqrt{3/2}, -\sqrt{1/2}), \quad G_2 = (-\sqrt{3/2}, -\sqrt{1/2}) \quad \text{and} \quad G_3 = (0, \sqrt{2}).$$

Four things should be noted about this assignment of gradients. First, the G_i's remain the same value when we change the sizes of the regions but keep the directions of lines the same. For example, Fig. 18(d) will result in the same combination of gradients. This seems reasonable.

Second, in this assignment the surfaces S_1, S_2, and S_3 are perpendicular to each other, because the inner products of their surface normals $(p_i, q_i, 1)$ vanish:

$$p_i p_j + q_i q_j + 1 = 0 \quad \text{for } i \neq j. \tag{6}$$

This is not a special result for this particular case in which the angles between lines aroung the FORK junction V_7 are all 120°. It can be generally shown (Appendix B) that in the case of a picture made of three parallel quadrilaterals like Fig. 19(a), the above method of assignment always gives a unique solution in which the three surfaces are perpendicular to each other, if and only if all the three angles at the FORK junction are obtuse. This implies that Fig. 19(b), in which one of the angles is acute, cannot be a picture of a right-angled block; some surfaces are really skewed.

It is interesting to observe the following here. Suppose we try to draw a skewed parallelepiped (a rhomboidal prism). Then we normally choose a view angle which yields a picture like Fig. 19(b), not a view angle which yields a picture like Fig. 19(a), probably because people will not perceive the latter as a skewed parallelepiped even though it could be.

Third, once the gradients have been computed we know the plane formula in the 3-D space for each surface. Of course, we have to assume the z position of one point on the object; there is no way to know this information from a single picture. Let us assume z of V_7 to be 0. If we select the origin of the x–y picture

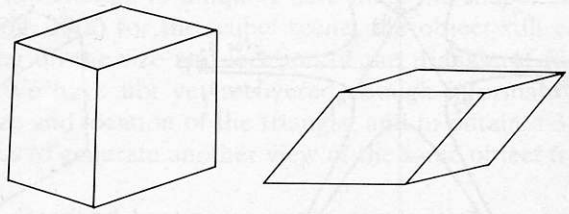


FIG. 19. Line drawings of parallelepipeds: (a) This can be a right-angled parallelepiped or a rhomboidal prism. (b) This must be a squashed rhomboidal prism (the shape of at least one surface must be really skewed).

plane at V_7, then we have

$$\begin{aligned} S_1&: \sqrt{3/2}x - \sqrt{1/2}y + z = 0, \\ S_2&: -\sqrt{3/2}x - \sqrt{1/2}y + z = 0, \\ S_3&: \sqrt{2}y + z = 0. \end{aligned} \tag{7}$$

The 3-D coordinates of the vertices are found to be

$$\begin{aligned} V_1&: (0, 1, -\sqrt{2}), & V_2&: (\sqrt{3}/2, 1/2, -\sqrt{1/2}), \\ V_3&: (\sqrt{3}/2, -1/2, -\sqrt{2}), & V_4&: (0, -1, -\sqrt{1/2}), \\ V_5&: (-\sqrt{3}/2, -1/2, -\sqrt{2}), & V_6&: (-\sqrt{3}/2, 1/2, -\sqrt{1/2}), \\ V_7&: (0, 0, 0). \end{aligned}$$

Fourth, for a parallel quadrilateral, another skewed symmetry can be defined by its diagonals. Therefore, other combinations of the axes, such as shown in Fig. 20(a) are possible. In the case of Fig. 20(a), the assignment of gradients is not unique; those triangles shown in Fig. 20(b) are all possible assignments. Since the diagonals of each surface are perpendicular in the 3-D space, the shapes corresponding to those triangles are, in general, the corner of a rhomboidal prism whose surfaces are equal-sized rhombuses. In both combinations of skewed symmetries in Fig. 18 and Fig. 20 the skewed-symmetry heuristic holds for every surface. The difference between them lies in whether a

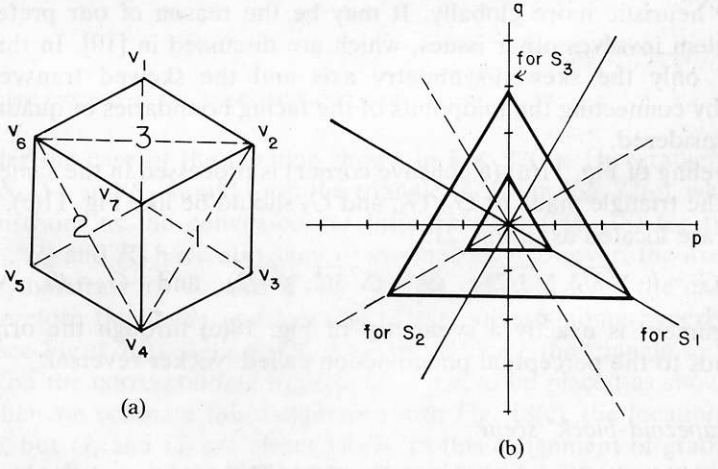


FIG. 20. The parallel quadrilaterals in Fig. 18(b) can have other combinations of axes for skewed symmetries.

(a) The case in which the diagonals are selected for each region to define its skewed symmetry.

(b) The corresponding loci (each hyperbola has degenerated into two lines in this case). The triangles are all possible assignments of gradients.

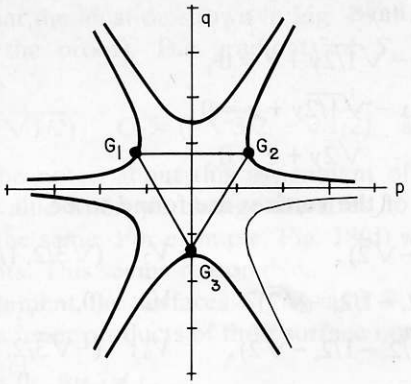


FIG. 21. Positions of G_1, G_2, and G_3 corresponding to the labeling of Fig. 11.

composite of skewed-symmetrical regions in the picture is interpreted as a symmetrical composite of surfaces in the scene. That is, in the interpretation of Fig. 18, the composite of S_1 and S_3 is symmetrical about the connected line segments in the 3-D space which correspond to the connected two axes (i.e., *ABC*) of the two skewed symmetrical regions in the picture. This is not the case for Fig. 20: here the composite of S_1 and S_3 is not symmetrical about the connected lines in the 3-D space which correspond to $V_1V_7V_3$ or $V_6V_2V_4$. In other words, the interpretation of Fig. 18 seems to satisfy the skewed-symmetry heuristic more globally. It may be the reason of our preferring it. This problem involves other issues, which are discussed in [10]. In this paper, hereafter, only the skewed-symmetry axis and the skewed transverse axis obtained by connecting the midpoints of the facing boundaries of quadrilaterals will be considered.

The labeling of Fig. 11(a) (a concave corner) is processed in the same way. In this case the triangle made of G_1, G_2, and G_3 should be like Fig. 11(c), thus the gradients are located as in Fig. 21:

$$G_1 = (-\sqrt{3/2}, \sqrt{1/2}), \quad G_2 = (\sqrt{3/2}, \sqrt{1/2}) \quad \text{and} \quad G_3 = (0, -\sqrt{2}).$$

This assignment is exactly a symmetry of Fig. 18(c) through the origin, and corresponds to the perceptual phenomenon called Necker reversal.

7.1.2. '*Trapezoid-block*' *scene*

Let us consider the line drawing of Fig. 1(b). The angles and the lengths are shown in Fig. 22(a). As we have noticed this line drawing has the same qualitative interpretations (labelings) as Fig. 1(a): i.e., three labelings which have been shown in Figs. 10–12. However, Fig. 22(a) seems to depict a quantitatively different shape. What makes the difference?

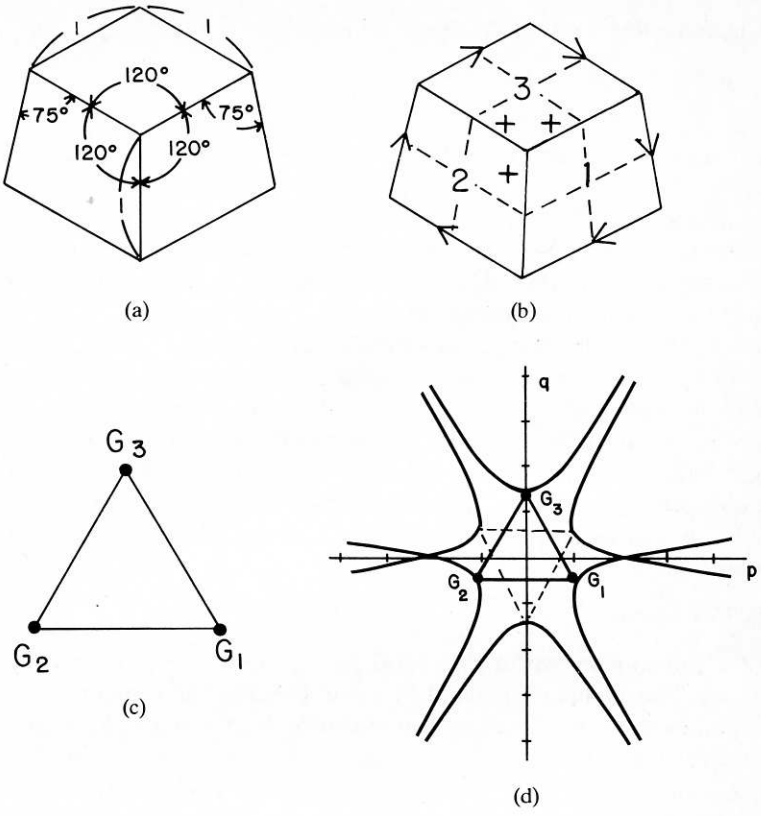


FIG. 22. Shape recovery of the 'trapezoid-block' scene.

Consider the case of the labeling shown in Fig. 22(b). The gradients of the surfaces S_1, S_2, and S_3 should form the triangle shown in Fig. 22(c), which is the same constraint as the convex-corner interpretation (Fig. 18) for the 'cube' scene. R_1, R_2 and R_3 have also skewed symmetries. However, the axes for the skewed symmetries for R_1 and R_2 are slightly different from the case of Fig. 18(b), therefore the shape and location of the corresponding hyperbolas also change (see Fig. 22(d)). As a result, if we try to locate the triangle so that each vertex is on the corresponding hyperbola, it has to be placed as shown in Fig. 22(d). When we compare this assignment with Fig. 18(c), the location of G_3 is the same, but G_1 and G_2 are closer to G_3. In this assignment of gradients, the angle made by S_1 and S_3 is equal to that made by S_2 and S_3, and is larger than 90°. Thus, the resultant shape is a trapezoid block.

As in the case of the 'cube' scene, we also have the combination of gradients shown as the dotted triangle in Fig. 22(d), corresponding to the labeling of Fig. 11(a); the shape is the reversal of Fig. 22(b).

7.2. Cases involving partially-visible surfaces

7.2.1. *Identification of surface boundaries*

In order to solve the cases which involve partially occluded surfaces, identification of surface boundaries and inference of the occluded parts is necessary. This problem needs research for a systematic method. Though we have a method which is sufficient for our examples, the topic is worth another complete paper. We do not pursue it further here, and show how the reasoning can be made in our particular example cases. The method may sound a little ad hoc, but the following is noteworthy.

For identifying the surface boundaries in a labeled drawing, the junction labels play a substantial role, because they represent how boundaries of surfaces 'appear' and 'hide' at vertices. Actually, notice that we can identify surfaces only after a line drawing has been labeled: for example, the labelings of Figs. 10(a) and 12(b) have a different set of surfaces. In fact, we can assign rules concerning boundary lines to each junction label and, when given a labeled line drawing, we can trace the boundaries of each surface according to the rules.

7.2.2. '*Box*' *scene*

Take the labeling shown in Fig. 13(a) for the 'box' scene. (It is reproduced in Fig. 23(a)). This is one of its eight interpretations in the Origami world [8]. The labeling indicates that there are four surfaces, S_1, S_2, S_3 and S_4. It also indicates that the relationships among their gradients should be as illustrated in Fig. 23(b): G_1 and G_2 are on a line perpendicular to V_5V_7 in the order as shown, because of the convex label given to V_5V_7. Similarly, G_3 and G_4 are on a line perpendicular to V_2V_8 in the order as shown, because of the concave label given to V_2V_8. G_3 should be within the angle G_2G_1a because of the junction label given to V_1, and G_4 within G_1G_2b. Again, these constraints that the labeling provides are not enough to uniquely determine the values of gradients. The heuristics of nonaccidental regularities provide additional constraints.

First, S_1 and S_2 are totally-visible surfaces, and both have skewed symmetries in the picture; their axes are shown in Fig. 23(a). Therefore, if we apply the skewed-symmetry heuristic, the gradients of S_1 and S_2 should be on the hyperbolas drawn as solid curves respectively in Fig. 23(c). Notice, however, the labeling indicates that S_3 and S_4 are only partially visible: the region $V_1V_2V_8V_7$, for example, depicts a part of S_3 because it is occluded along the lines V_1V_7 and V_7V_8. Therefore the shape of the region is meaningless.

Second, the line V_2V_8 is parallel in the picture with the line V_7V_5. Thus, if we apply the parallel-line heuristic, the surfaces which include these lines as boundaries (i.e., S_1, S_2, S_3, and S_4, due to the labeling) should have their gradients on a line perpendicular to V_5V_7. That is, G_1 through G_4 should be on a single line. Further, because V_1V_7 is parallel to V_2V_3, G_1 and G_4 (the gradients of the surfaces which include these lines) should be on a line

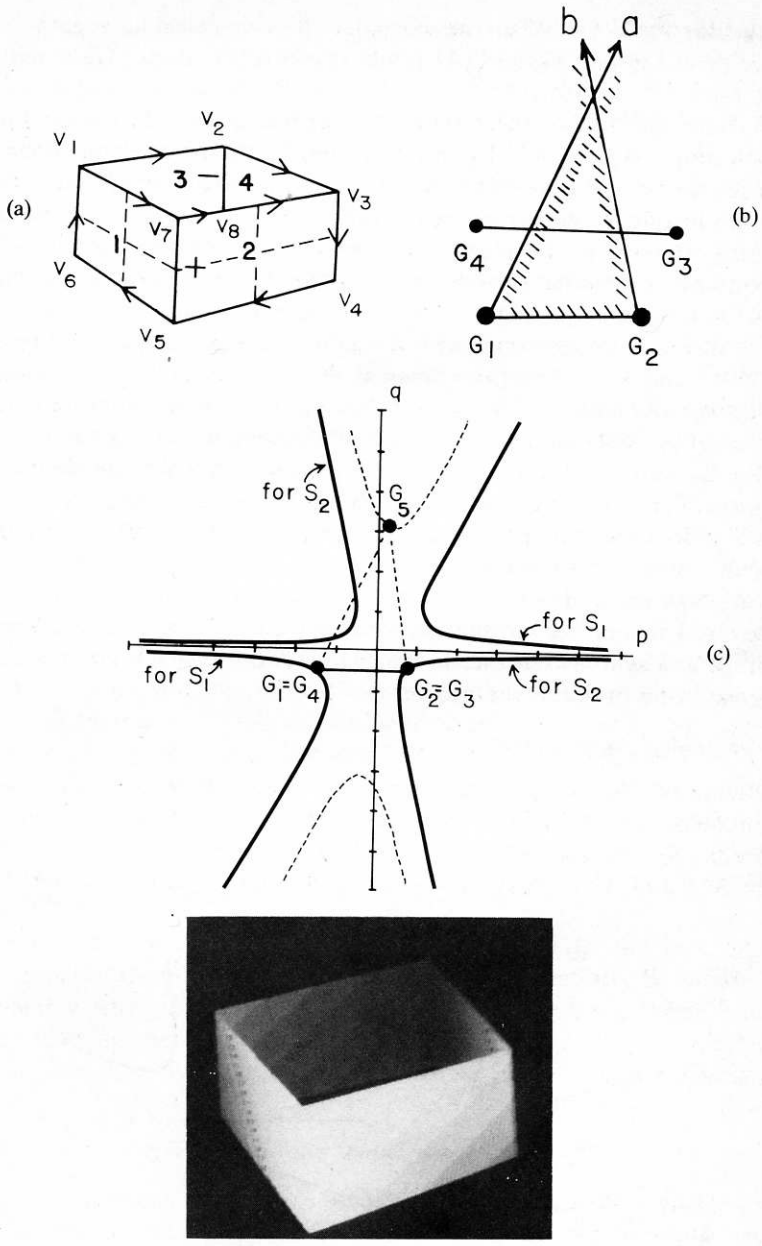


FIG. 23. Shape recovery of the labeling of Fig. 13(a) for the 'box' scene: (a) labeling; (b) constraints on the gradients; (c) assignment of the gradients; (d) view of the recovered shape from 10° to the right and 10° above the original view angle. This is a depth-coded image in which the depth of each point is coded by the intensity: the brighter, the nearer.

perpendicular to V_1V_7. When we consider these constraints together in Fig. 23(b), we conclude that G_1 and G_4 should coincide. Similarly, G_2 should be the same as G_3.

Third, note that use of the parallel-line heuristic on both the pair of V_1V_7 and V_2V_3 and the pair of V_1V_2 and V_3V_7 implies that these four lines (in the 3-D space) are on the same plane and form a quadrilateral. Therefore, we can think of an imaginary planar surface segment S_5 ($= V_1V_2V_3V_7$). Because S_5 has a skewed symmetry in the picture, its (imaginary) gradient should be on the corresponding hyperbola shown as a dotted curve in Fig. 23(c). Since S_5 connects with S_1 and S_2 along V_1V_7 and V_3V_7, respectively, the gradients of S_1, S_2 and S_5 should form a triangle whose shape is determined by the directions of V_1V_7, V_3V_7, and V_5V_7 (the three lines at the FORK junction V_7). Remember that what we did here is not ad hoc but a result of considering the scene-domain meaning of labelings and the parallel-line heuristics.

Finally, by putting all the above constraints together we can determine the locations of G_1 ($= G_4$) and G_2 ($= G_3$) as shown in Fig. 23(c). As before, if we fix the origin of the x–y picture plane, and assume the z (depth) of any point on the object, the plane formulas for S_1 through S_4 can be determined, and we obtain the 3-D shape description of the object. Fig. 23(d) is a rotated view of the recovered shape. We see that an 'ordinary box' shape has been recovered. This image is a synthesized one in which the depth of each point is coded by the intensity: the brighter, the nearer.

7.2.3. *'W-folded paper' scene*

The labeling of Fig. 14(a) for the 'W-folded paper' scene is another example which involves partially visible surfaces (see Fig. 24(a)): S_1 is totally visible, but S_2 through S_4 are partially visible. Because of the parallelisms of lines ($V_1V_2 /\!/ V_9V_{10}$ and $V_1V_{11} /\!/ V_9V_{12}$) we conclude that S_1 and S_3 have the same

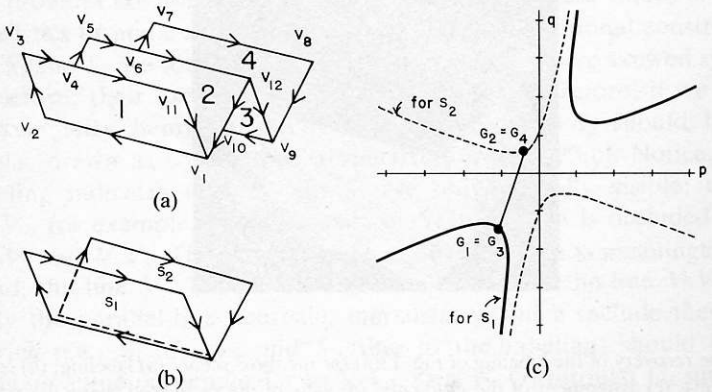


FIG. 24. Shape recovery of the 'W-folded paper' scene for the labeling of Fig. 14(a).

gradient ($G_1 = G_3$). Similarly, $G_2 = G_4$. Due to the parallelism of V_3V_{11} and V_5V_{12}, G_1 and G_2 are to be on a line perpendicular to them. The skewed symmetry of S_1 constrains G_1 to be on the hyperbola drawn as a solid curve in Fig. 24(c). The constraints we have obtained so far are not enough for the unique determination of G_1 and G_2. Even when we determine G_1 at one of the tips of the solid hyperbola, G_2 can still be anywhere on the half line extending upward from G_1 in the perpendicular direction to V_3V_{11}.

In order to extract more constraints, the shapes of partially visible surfaces need to be inferred. Let us show how the labeling and the constraints obtained so far can be used for this. Let us work with S_2. Imagine that we are tracing the boundaries of S_2 clockwise. The labeling tells that they show up along V_4V_5, V_5V_{12} and $V_{12}V_1$, but 'disappear' at V_1 and 'reappear' at V_4. The constraint that G_1 and G_2 are to be on a line perpendicular to V_3V_{11} or V_5V_{12} implies that if S_1 and S_2 actually intersect, the occluded intersection line (i.e., the occluded boundary line of S_2) should coincide with V_1V_2 (because $V_1V_2 /\!/ V_3V_{11} /\!/ V_5V_{12}$). How about the boundary line which reappeared at V_4? Since it is interrupted by S_1 at V_4 (notice that this is indicated by the junction label given to the T junction V_4), it is meaningful to extend V_5V_4 in the picture and we find that the extended line intersects at V_2 with the inferred occluded boundary from V_1. Thus the shape of S_2 could be the dotted line shown in Fig. 24(b).

Application of the skewed-symmetry heuristic to this region provides the dotted curve shown in Fig. 24(c). Since we have no more constraints, let us determine G_1 as the lower tip of the solid hyperbola, and G_2 on the dotted hyperbola so that the line connecting them is perpendicular to V_1V_2 (Fig. 24(c)). Note that the upper tip of the solid hyperbola is not appropriate for G_1 because then G_2 cannot be on the dotted hyperbola.

7.3. 'Strange' shapes violate the regularity heuristics

The labelings we have treated so far all correspond to the most 'natural' interpretations of the pictures. The qualitative shape recovery by the theory of the Origami world yields other labelings. This section will show that some shapes implied by the labelings are really 'possible' but violate some of the regularity heuristics. We conjecture that this is why they look 'strange' or 'unnatural'.

7.3.1. *Non-cube interpretation of the 'cube' scene*

Take the third labeling (Fig. 12(a)) for the 'cube' scene. It is duplicated in Fig. 25(a). It indicates that two totally-visible surfaces S_1 and S_2 which are connected and form a convex edge, occlude partially the third surface S_3. The relationships among the gradients G_1 through G_3 imposed by this labeling are illustrated in Fig. 25(b): G_3 should be inside of the triangle or on the base line connecting G_1 and G_2. We see that the parallel-line heuristic cannot be satisfied

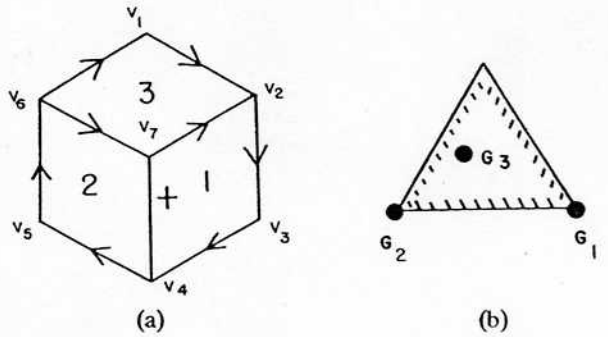


FIG. 25. The third labeling (Fig. 12) for the 'cube' scene does not satisfy the parallel-line heuristic: (a) labeling; (b) constraints in the gradient space. G_1 has to be inside of the triangle or on the base line, thus G_1 and G_2 cannot be placed so that the parallelism of V_1V_2 and V_6V_7 is not accidental.

in this case. The parallelism of lines V_1V_2 and V_6V_7 demands that G_3 and G_2 are on a line perpendicular to V_6V_7 (i.e., on the left leg of the triangle), and the parallelism of V_1V_6 and V_2V_7 demands that G_3 and G_1 are on a line perpendicular to V_2V_7 (i.e., on the right leg of the triangle), but both cannot be the case simultaneously.

7.3.2. *Phoney-box interpretations of the 'box' scene*

Let us go to the 'box' scene. The labeling of Fig. 13(b) represents a shape of a 'squashed' box with the front two faces going in. Again, Fig. 26 (the diagram illustrating the constraints in the gradient space) indicates that the parallel-line heuristic cannot be satisfied: G_1 and G_4 cannot be on a line perpendicular to V_1V_7 (or V_2V_3).

We have noted that there are eight labelings for the 'box' scene in the Origami world, but it can be found that the interpretations other than Fig. 13(a) which corresponds to an 'ordinary box' shape, do not satisfy the regularity heuristics.

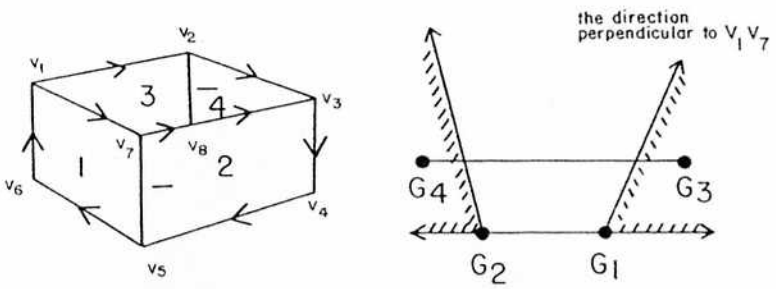


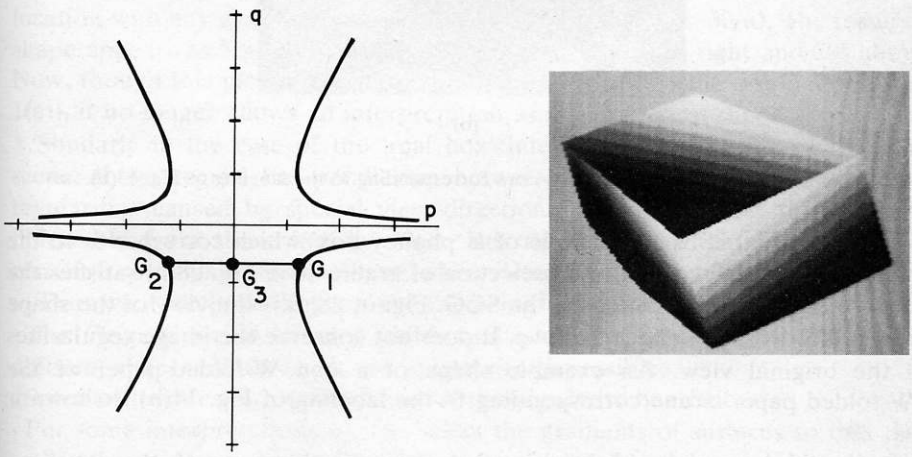
FIG. 26. Constraints on G_1 through G_4 imposed by the labeling of Fig. 13(b). G_1 and G_4 cannot be on a line perpendicular to V_1V_7.

7.3.3. *Non-W-folded-paper interpretations of the 'W-folded paper' scene*

In the case of the 'W-folded paper' scene, the labeling of Fig. 14(b) is an example which violates the heuristics. Again only Fig. 14(a), out of 16 possible labelings, can have the assignment of gradients without violating the regularity heuristics.

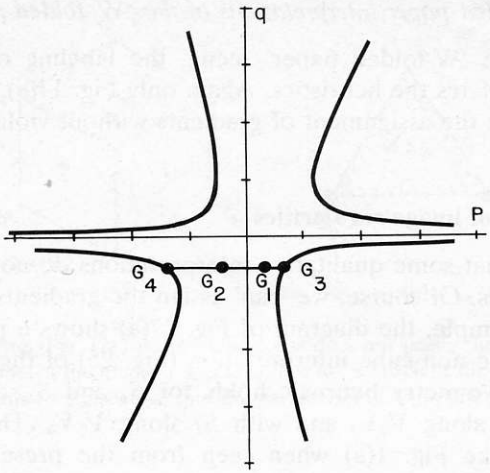
7.4. Conservation of image regularities

We have shown that some qualitative interpretations do not satisfy the image regularity heuristics. Of course, we 'can' assign the gradients, if we neglect the heuristics. For example, the diagram of Fig. 27(a) shows a particular selection of gradients for the non-cube interpretation (Fig. 25) of the 'cube' scene such that the skewed-symmetry heuristic holds for S_1 and S_2, and the surface S_3 intersects with S_1 along V_2V_3 and with S_2 along V_5V_6. The resultant object appears exactly like Fig. 1(a) when seen from the present view direction. However, when it is seen from other directions, say, from 15° to the right and 15° above, it looks like Fig. 27(b). We notice that the image regularity (parallelism) we observed in the original view (Fig. 1(a)) has disappeared. Recall that, in contrast, the labeling of Figs. 10 and 11 allows such a selection of gradients (Figs. 18 and 21) corresponding convex and concave corners, respectively, and that the resultant shapes produce pictures which conserve those regularities from whatever direction they are seen.

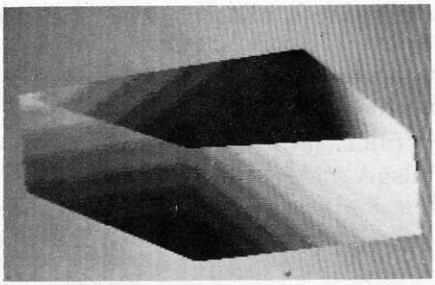


(a) (b)

FIG. 27. A 'strange' shape for the 'cube' scene: (a) a possible assignment of the gradients corresponding to the labeling of Fig. 12; (b) a view from another angle of the resultant shape. It does not preserve the parallelism of V_1V_2 and V_6V_7 observed in the original view.



(a)

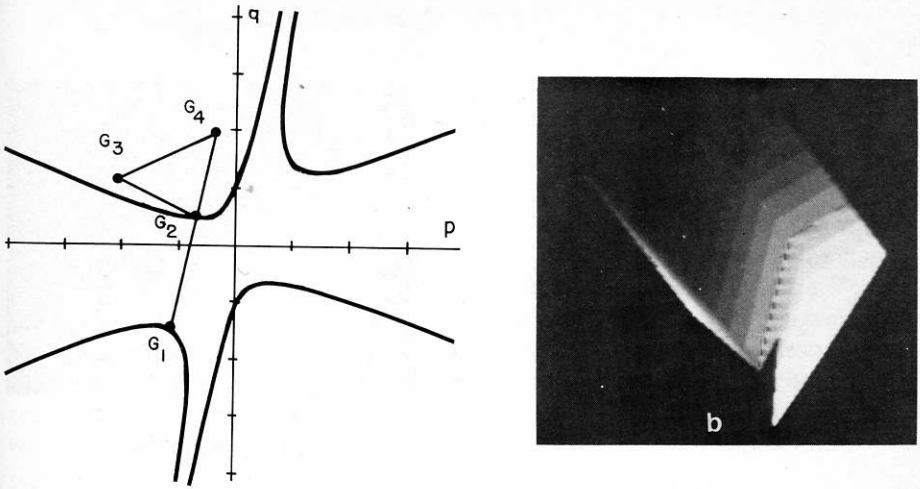


(b)

FIG. 28. A 'strange' shape for the 'box' scene corresponding to the labeling of Fig. 13(b).

Likewise, Fig. 28 is an example of a 'phoney box' which corresponds to the interpretation of Fig. 13(b). The selection of gradients in Fig. 28(a) satisfies the basic constraints represented by the SCG. Figure 28(a) is the view of the shape from 10° to the right and 10° above. It does not conserve the image regularities in the original view. An example shape of a non W-folded-paper of the 'W-folded paper' scene corresponding to the labeling of Fig. 14(b) is shown in Fig. 29.

It should be emphasized again that even in the case of the labelings corresponding to 'natural' shapes there is no physical reason that the objects have to be of such particular shapes. In fact, in the convex-corner interpretation of the 'cube' scene (Fig. 18), if we do not use the skewed-symmetry heuristic, we could place the triangle of Fig. 10(c) in the gradient space at any



(a) (b)

FIG. 29. A 'strange' shape for the 'W-folded paper' scene corresponding to the labeling of Fig. 14(b): (a) assignments of gradients; (b) a rotated view of the shape from 25° to the right and 25° above. This shape produces the line drawing of the 'W-folded paper' scene when seen from the original direction.

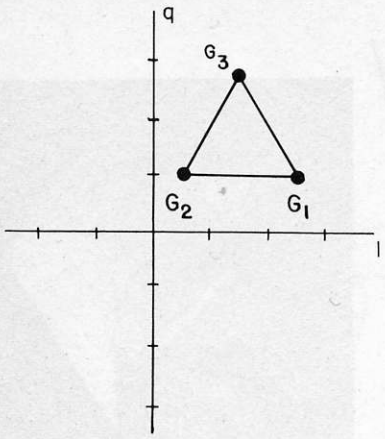
location with any size. Suppose we place it as shown in Fig. 30(a). The resultant shape appears as Fig. 30(b) when viewed from 20° to the right and 20° above. Now, though this picture is still qualitatively the same as the original one (Fig. 1(a)), it no longer allows an interpretation as a symmetrical shape.

Similarly in the case of the 'real box' interpretation (Fig. 23) of the 'box' scene, if we regard the parallelism between the picture lines as accidental regularities caused by special view directions, we could select the gradients differently from those in Fig. 23(c). But then the resultant shape will not conserve some parallelism of lines when viewed from other directions.

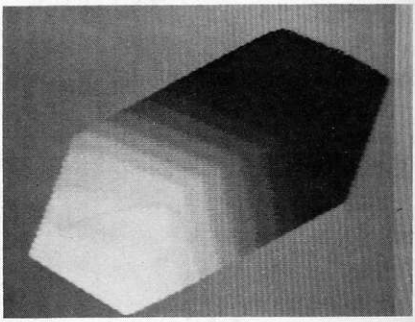
Thus what we have shown is the following.

- The Origami world labeling yields multiple interpretations of qualitatively different shapes, all of which are geometrically possible.
- Some of them cannot satisfy some of the regularity heuristics.
- For some interpretations we can select the gradients of surfaces so that they satisfy the regularity heuristics. Then the resultant shapes have real regularities and, therefore, conserve the image regularities observed in the original view, even when seen from other view angles.

We feel that this conservation of image regularities is strongly related to whether the interpreted shapes seem 'natural' to a human observer.



(a)



(b)

FIG. 30. A placement of the triangle of Fig. 10(c) in the gradient space different from that in Fig. 18(c) results in a skewed parallelepiped.

8. Shape Recovery from Real Images: More Complexity and More Constraints

In the previous sections, we have presented the basic theory and technique to extract and to exploit geometric constraints for quantitative shape recovery. The three-dimensional shapes of our simple example scenes have been recovered. When it comes to the shape recovery from real images, such as the 'chair' scene (Fig. 5), more complexities are involved and more constraints are available. The succeeding two sections will present the techniques to be used for the shape recovery from real images. The 'chair' scene example will be demonstratively processed.

Those simple line drawings we have treated so far (Figs. 1, 3, and 4) consist of a small number of lines and regions, and the number of legal labelings in the Origami world is also small. Thus we could consider each labeling and test whether the regularity heuristics are applicable in order to select 'natural' or 'probable' shapes. However, line drawings with many lines and regions will have too many labelings (i.e., geometrically possible shapes) to consider them all individually. The next section shows how we can extract the constraints that color images provide on line labels, and how we can exploit them to order multiple labelings according to their degree of match with image properties.

The next problem is related with noise and distortions which the real pictures include. Due to them, positioning of the gradients may not proceed in such an idealistic manner as shown in Section 7. Section 10 presents a solution to this problem.

9. Use of Color Edge Profiles

9.1. Edge profiles and line labels

An edge profile is a curve showing how the intensity (or more generally, a certain image property) changes across an edge. Historically, in the scene analysis of the block world, it has been known that a few typical types of intensity profiles exist (Herskovits and Binford [4]), and Horn [5] showed that they can suggest different types of edges: for instance, a peak-type profile suggests a convex edge, a roof type a concave, and a step type an occluding edge. Thus the edge profiles might be used to provide constraints on line labels. However, this absolute method, which tries to associate properties of edge profiles with label identities, is not usually very reliable. It strongly depends on the lighting conditions and the physical composition of objects. Further, for this method to work, particular properties of edge profiles, such as roof, peak, etc., need to be recognized.

There is another way of exploiting edge profiles which is employed in this paper; that is the relative method. It is based on the observation that if two lines have 'similar' edge profiles, it suggests that they will likely take the same label, even though the label identity itself is not known. It can be noted that this relative method of using color edge profiles is an instance, applied to their similarities, of the meta-heuristic of non-accidental regularities described in Section 6.

If the lines with similar edge profiles form a certain special geometrical configurations, the likelihood for a similar labeling will be higher. The classical matched T configuration, first described by Guzman [3], is such an example. In Fig. 31 three pairs of lines (L_1 and L_2, L_3 and L_4, and L_5 and L_6) are collinear. Guzman used this configuration as a clue suggesting that the regions R_1 and R_2 belong to the same object, and similarly R_3 and R_4.

What the matched T's mean in the context of this paper is the following. If the edge profiles of L_1 and L_2 are similar, then the labels of L_1 and L_2 are likely to be the same, and the lines L_3 through L_6 will take occluding-edge labels in such a way that the middle region R_0 interrupts an edge which is projected into L_1 and L_2 in the picture. In this section we formalize these kinds of ideas.

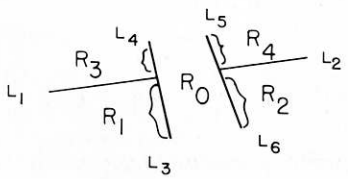


FIG. 31. Guzman's matched T's.

9.2. Extraction of color edge profiles

Color edge profiles for a line consist of three profile curves corresponding to three color components: R (red), G (green), and B (blue). As shown in Fig. 32, for each color component, a set of profiles is first taken across the line; the margin (m), the interval (i), and the length (l) are specifiable parameters. Profile curves in each set are registered and then averaged to obtain the mean profile. In this way for each line L_i we obtain three profile curves $P_i^c(t)$, $(-l \leqq t \leqq l)$ where $c \in \{\mathrm{R}, \mathrm{G}, \mathrm{B}\}$.

Notice that lines (and thus profiles) have to be defined with directionality. That is, a line L_i is to be defined as connecting *from* V_{i1} *to* V_{i2}, rather than connecting V_{i1} *and* V_{i2}. Let us denote this line as $L_i = (V_{i1}\ V_{i2})$. Then $-L_i = (V_{i2}\ V_{i1})$ denotes the same line traced in the opposite direction. If L_i has a profile $P_i^c(t)$, $-L_i$ has a profile $P_{-i}^c(t) = P_i^c(-t)$. In the following discussions the directions of lines and profiles always have to be considered.

Since we can compare three profile curves the relative method is more reliable and useful when exploiting color edge profiles than when using only monochrome intensity profiles.

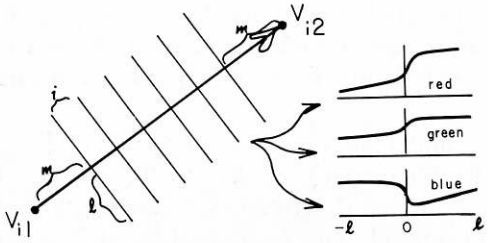


FIG. 32. Extraction of color edge profiles. A set of profile curves are taken across the line for each color component. Each set is registered and averaged.

9.3. Comparison of color edge profiles

For the purpose of comparing the edge profiles let us define the amplitude of $P_i^c(t)$ and the distance between $P_i^c(t)$ and $P_j^c(t)$ as follows.

$$A(i, c) = \max_t P_i^c(t) - \min_t P_i^c(t),$$

$$d(i, j, c) = \min_s \left(\int |P_i^c(t) - P_j^c(t+s)|^2 \, \mathrm{d}t \right)^{1/2}. \tag{8}$$

The amplitude $A(i, c)$ indicates roughly how much the property of the profile changes across the line. The distance $d(i, j, c)$ is the minimum mean square

difference between the two curves; the minimum is taken over various shifts of the curves in order to compensate possible dislocations of the origins of the two profiles. It is easy to see the following relations.

$$A(i, c) = A(-i, c), \qquad d(-i, -j, c) = d(i, j, c) \tag{9a}$$

but, in general,

$$d(-i, j, c) \neq d(i, j, c), \qquad d(i, -j, c) \neq d(i, j, c). \tag{9b}$$

Let us now define $s(i, j)$, the similarity of lines, for a pair of lines L_i and L_j. There is not a definite reason to employ the following definition but it can be thought that the larger the amplitude, the more meaningful is the distance between the profile curves. Thus we will define $s(i, j)$ as the sum of the ratios of the amplitude to the distance for each color component: that is, if

$$(A(i, c) + A(j, c))/d(i, j, c) > 1 \quad \text{for all } c,$$

$$s(i, j) = \sum_{c \in \{R,G,B\}} \frac{A(i, c) + A(j, c)}{d(i, j, c)}, \tag{10a}$$

otherwise (i.e., if the ratio is small, the profiles may not be reliable)

$$s(i, j) = 0. \tag{10b}$$

Because of the property of $d(i, j, c)$ and $A(i, c)$, $s(i, j) = s(-i, -j)$ but usually $s(i, j) \neq s(-i, j)$.

9.4. Constraint expressions from edge profile analysis

The similarity in edge profiles, together with certain geometrical configurations of lines, can be converted to the constraint expressions on line labels.

9.4.1. *Similar edge profiles*

For a pair of lines L_i and L_j whose similarity $s(i, j) > 0$ we generate a constraint expression,

$$(\text{SAME } (L_i\, L_j)\, s(i, j))$$

which means that L_i and L_j may have the same label with a confidence value (weight) $s(i, j)$. All the pairs of L_i's and $-L_i$'s have to be considered for generating this type of constraint.

9.4.2. *Matched T configuration*

Suppose that L_1 through L_6 form a matched T configuration, as shown in Fig. 33: it is a pair of opposing T junctions whose vertical bars are collinear. Define the directions of lines as shown by the big arrows in Fig. 33. Let $s(1, 2)$, $s(4, 5)$, and $s(3, 6)$ be the similarities between the corresponding pairs of lines, respec-

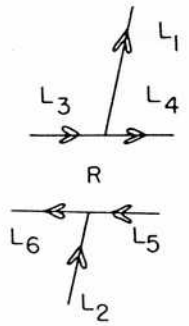


FIG. 33. Matched T configuration and the specification of the line directions.

tively. Then the matched T configuration generates two constraint expressions:

$$(\text{SAME}\ (L_1\ L_2)\ w_T) \quad \text{and} \quad (\text{IDENT}\ (L_3\ L_4\ L_5\ L_6)\ (\rightarrow\rightarrow\leftarrow\leftarrow)\ w_T)$$

where $w_T = (s(1, 2) \cdot s(4, 5) \cdot s(3, 6))^{1/3}$.

The first expression constrains L_1 and L_2 to take the same label. The second one means that L_3 through L_6 take such a combination of labels (occluding edges in the proper directions), which corresponds to the case that the middle region R sandwiched by the two transverse bars of T's occludes the collinear vertical bars. Since the first expression about L_1 and L_2 must have been generated from their own similarity, we can regard the matched T configuration as amplifying the constraint. The definition of w_T is rather arbitrary, but the reason for the above definition is just for giving it the same dimension as $s(i, j)$'s.

9.4.3. *Matched π configuration*

A matched π configuration shown in Fig. 34 is another geometrical configuration which can amplify constraints. It is made of two neighboring T junctions which share the same line (L_4) as their right and left transverse bars, and which have parallel vertical bars (L_1 and L_2). Let us define the directions of lines as shown, and let $s(1, -2)$ and $s(3, 5)$ denote the similarities between L_1 and $-L_2$ and between L_3 and L_5, respectively. The matched π configuration corresponds most often to the case where the same physical configuration occurs in the right (L_2 and L_5) and left (L_1 and L_3) halves. Therefore we generate a

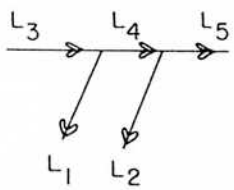


FIG. 34. Matched π configuration and the specification of the line directions.

pair of constraint expressions:

$$(\text{SAME}\ (L_3L_5)\ w_P) \quad \text{and} \quad (\text{SAME}\ (L_1 - L_2)\ w_P)$$

where $w_P = (s(1, -2) \cdot s(3, 5))^{1/2}$.

9.5. Search for interpretations which most satisfy the constraints

Now that we have obtained a set of constraints which the line labels should satisfy, we can search for the 'best' interpretations in the sense that the constraints are most satisfied. In evaluating an interpretation, if it does not satisfy a constraint, a penalty (the associated weight) is given to it. Thus, the best interpretation is the one with the least penalty.

In the present implementation, the search is combined with the Origami-world labeling procedure. The interpretation of least penalty is searched for in a depth-first manner, along with the search for consistent assignments of junction labels: each time a junction label is assigned to a junction, not only does the partial interpretation undergo the consistency check of surface orientations, but also the penalty score is renewed if any constraint expression is available (i.e., all the lines involved in it have been given labels).

If the partial interpretation turns out to be inconsistent in surface orientations, or if the penalty score exceeds the best (smallest) score so far obtained, then the search backtracks and the leaves below the present node in the search tree are pruned.

9.6. Labeling the 'chair' scene

9.6.1. *Color edge profile analysis*

For the 'chair'-scene image of Fig. 5(a), color edge profiles were computed for each line. Analysis of them yielded 85 constraint expressions. The total sum of the weights of all the constraints is 172.26.

9.6.2. *'Positive' and 'negative' chairs*

In labeling Fig. 5(b) with the constraints obtained above, it was found that the labelings shown in Figs. 35 and 36 are the two best interpretations: in each figure, (a) is the labeling and (b) is the SCG. Their penalty score is 13.17 out of 172.26.

Though these labelings still have ambiguities in their exact quantitative shape, it is interesting to think what kind of shapes they represent. Fig. 35(a) corresponds to the shape of an 'ordinary chair': the regions R_{14} and R_{15} (the wainscot in the wall) are occluded by others; the region R_9 (the arm nearer to the viewer) occludes other regions; the regions R_1 through R_8 are connected at the proper convex and concave intersections to form the shape of a seat and a back.

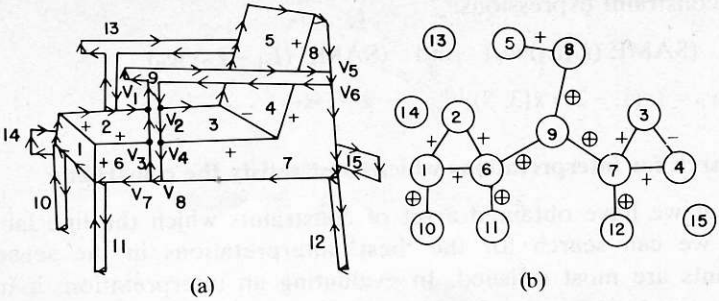


FIG. 35. One of the two 'best' interpretations for the 'chair' scene. We call this a 'positive' chair.

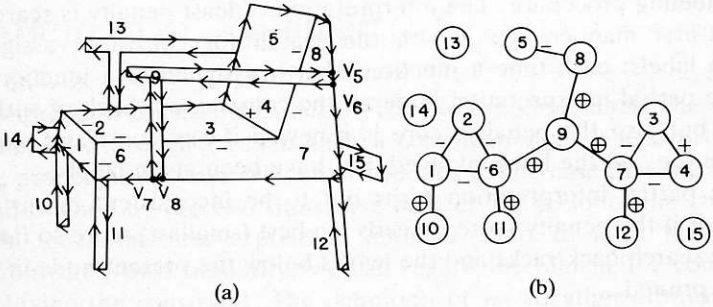
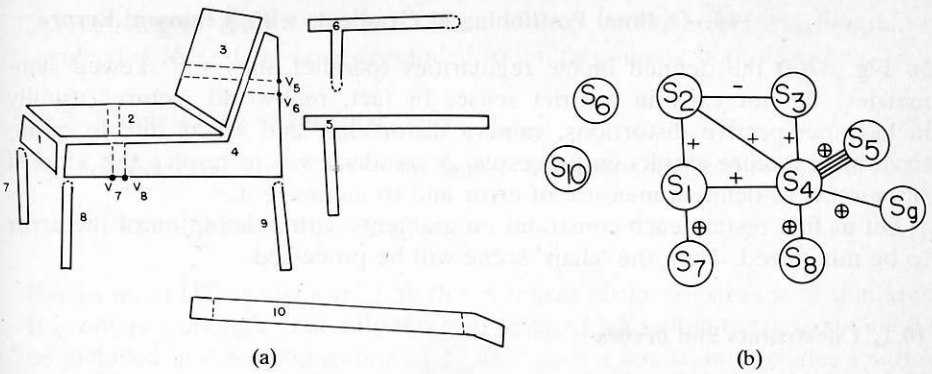


FIG. 36. The other of the two 'best' interpretations for the 'chair' scene. We call this a 'negative' chair.

Comparing the labeling of Fig. 35(a) with Fig. 36(a), we notice that the lines labeled as convex (+) in Fig. 35 are changed to concave (−) in Fig. 36, and concave to convex. The occlusion labels (↑) remain the same. These two labelings correspond to the phenomenon known as reversal in perceptual psychology, such as the Necker cube and the Shroeder stairs. Let us name the shape of Fig. 35 a 'positive' chair, and that of Fig. 36 a 'negative' chair. It might be difficult to imagine the shape of the 'negative' chair. Descriptively, the shape is obtained by first reversing the structure corresponding to the seat and back in the 'positive' chair and then attaching the arms and legs with appropriate relationships to the reversed structure.

9.6.3. *Identification of surfaces*

Now let us identify surfaces involved in the labeled 'chair' scene. Consider the labeling shown in Fig. 35(a) (the 'positive' chair). Several matched T's exist (e.g., V_1 and V_2, V_3 and V_4, etc.), where the collinear lines showing strong matches in color edge profiles are given identical labels and other lines are given occluding-edge labels of the appropriate directions. That is, these mat-



(a) (b)

FIG. 37. Identification of surfaces in the 'chair' scene corresponding to the labeling of Fig. 35 ('positive' chair): (a) identified surfaces; (b) merged SCG.

ched T configurations have been interpreted as the result of the interruption of a single 3-D space edge by another surface. Therefore, in tracing the boundary lines of surfaces, we can group a few regions into one region. For example, R_2 and R_3 are regarded as corresponding to a single surface. Ten surfaces are identified in this way (see Fig. 37(a)):

S_1: (R_1), S_2: $(R_2 R_3)$, S_3: $(R_4 R_5)$,

S_4: $(R_6 R_7 R_8)$, S_5: (R_9), S_6: (R_{13}),

S_7: (R_{10}), S_8: (R_{11}), S_9: (R_{12}),

S_{10}: $(R_{14} R_{15})$.

Corresponding to the merges of regions, the nodes in the SCG (Fig. 35(b)) are also merged. Fig. 37(b) shows the revised SCG which represents the interconnection relationships among the ten identified surfaces.

The 'negative' chair (Fig. 36) can be treated in the same way. We identify the same set of surfaces, and Fig. 38 shows the merged SCG. The difference between the 'positive' and the 'negative' chairs is only that the labels on the arcs connecting S_1, S_2, S_3, and S_4 in the SCG's of Figs. 37(b) and 38 are interchanged from + to − and vice versa.

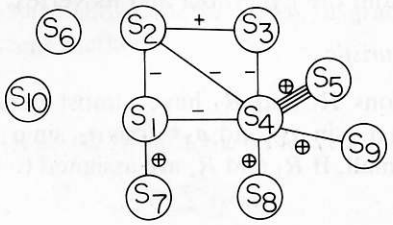


FIG. 38. The merged SCG for the labeling of Fig. 36 ('negative' chair).

10. Optimal Positioning of Gradients with Minimum Errors

In Fig. 37(a) the defined image regularities (parallel lines and skewed symmetries) do not exist in a strict sense. In fact, real-world pictures usually include perspective distortions, camera distortions, and errors due to quantization or feature extraction processes. A standard way to resolve this kind of situation is to define a measure of error and to minimize it.

Let us first restate each constraint on gradients with a definition of the error to be minimized. Then the 'chair' scene will be processed.

10.1. Constraints and errors

10.1.1. *Intersection of surfaces*

Suppose that two regions are connected along a convex or concave line with a direction $\boldsymbol{a} = (\cos\alpha, \sin\alpha)$ in the picture, and that they are given the gradients $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$. We will define the error for this case as,

$$e_I = |\boldsymbol{G}_1 \cdot \boldsymbol{a} - \boldsymbol{G}_2 \cdot \boldsymbol{a}|^2 = |(\boldsymbol{G}_1 - \boldsymbol{G}_2) \cdot \boldsymbol{a}|^2 \tag{11}$$

where e_I is the projection of the vector $\boldsymbol{G}_1 - \boldsymbol{G}_2$ onto a gradient-space line perpendicular to the picture line. It should vanish when $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ strictly satisfy the property of dual lines.

10.1.2. *Skewed-symmetry heuristic*

Suppose that a region has a skewed symmetry defined by the two directions $\boldsymbol{a} = (\cos\alpha, \sin\alpha)$ and $\boldsymbol{b} = (\cos\beta, \sin\beta)$, and that it is given a gradient $\boldsymbol{G} = (p, q)$. Since ideally $\boldsymbol{G}$ satisfies (4), we will define the error e_S for this case as,

$$e_S = |(\boldsymbol{G} \cdot \boldsymbol{a})(\boldsymbol{G} \cdot \boldsymbol{b}) + \cos(\alpha - \beta)|^2 . \tag{12}$$

Since the vertex of the hyperbola represents a special orientation, as described in Section 6, sometimes it is desirable to include the distance from the gradient to the vertex in the error as

$$e_{S'} = e_S + |\boldsymbol{G} - \boldsymbol{G}_T|^2 . \tag{13}$$

$\boldsymbol{G}_T$ is that one of the two vertices of the hyperbola nearer to $\boldsymbol{G}$. $e_{S'}$ is roughly the discrepancy of $\boldsymbol{G}$ from the hyperbola and its vertex.

10.1.3. *Parallel-line heuristic*

Suppose that two regions R_1 and R_2 have almost parallel boundaries whose directions are $\boldsymbol{a}_1 = (\cos\alpha_1, \sin\alpha_1)$ and $\boldsymbol{a}_2 = (\cos\alpha_2, \sin\alpha_2)$, respectively. Here it is assumed $|\alpha_1 - \alpha_2|$ is small. If R_1 and R_2 are assigned to $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ respectively, then

$$e_P = 1/2(|(\boldsymbol{G}_1 - \boldsymbol{G}_2) \cdot \boldsymbol{a}_1|^2 + |(\boldsymbol{G}_1 - \boldsymbol{G}_2) \cdot \boldsymbol{a}_2|^2) . \tag{14}$$

By referring to (11) we see that e_P is the mean of the errors which would be involved if R_1 and R_2 were demanded to intersect along both $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$.

10.2. Minimization of the total error

The total error E is defined as the sum of the individual errors defined above:

$$E = \sum e_I + \sum e_{S'} + \sum e_P . \tag{15}$$

Each sum in (15) is taken over all the instances of the constraints of that kind.[2] It is obvious, though, that only those surfaces which mutually constrain need to be included in the minimization of E. That is, if a constraint includes a surface whose gradient does not appear in the other constraints in E, the constraint can be removed from E because that gradient can be determined independently after all others have been determined.

The minimization of E can be obtained by the steepest descent method. The initial values of the gradients can be selected, so that as many constraints as possible for the basic interconnection relationships (the constraints for e_I, or the constraints in the SCG) are satisfied. For a skewed-symmetrical region one of the vertices of the corresponding hyperbola may be a good initial point.

10.3. Shape recovery of the 'chair' scene

10.3.1. *Determination of the gradients*

Let us consider the case of Fig. 37 ('positive' chair). The minimization of E involves the four surfaces (S_1, S_2, S_3, and S_4) which mutually constrain. The total error $E = E(p_1, q_1, p_2, q_2, p_3, q_3, p_4, q_4)$ consists of five terms for their surface interconnection (corresponding to the five arcs in the SCG), three terms for the skewed-symmetry heuristic (S_1, S_2, and S_3), and one term for the parallel-line heuristic (S_1 and S_3 have almost parallel boundaries as shown in Fig. 37(a); others were omitted for simplicity).

The initial locations are selected as shown in Fig. 39(a): G_1, G_2 and G_3 are at the vertices of the corresponding hyperbolas, and G_4 is at the intersection of two lines extending from G_1 and G_2 because of the constraints due to the intersections of S_4 with S_1 and S_2. Fig. 39(b) shows the locations of gradients determined by the iterative steepest descent method.

[2]If it is better to weigh the errors, we can define the total error as,

$$E_2 = W_I \sum w_i e_{I_i} + W_S \sum w_j e_{S_j} + W_P \sum w_k e_{P_k}$$

with appropriate selection of W's and w's.

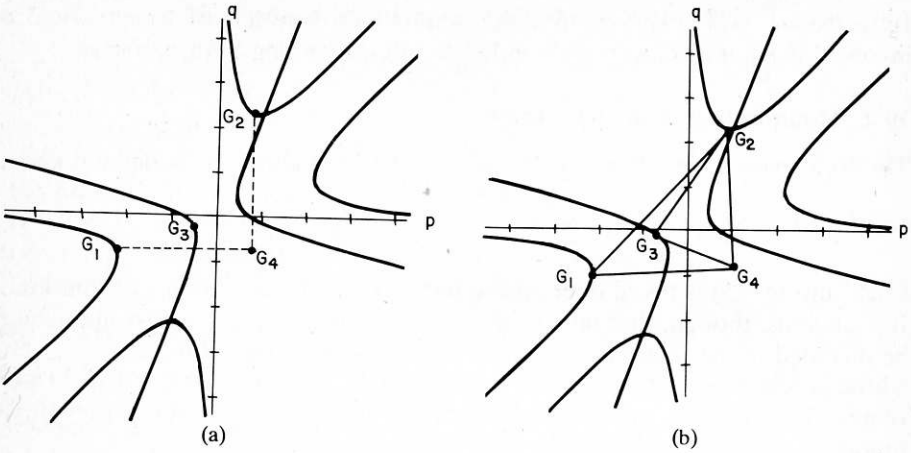


FIG. 39. Positioning of the gradients of S_1 through S_4 for the 'positive' chair by the iterative minimization of errors: (a) initial positions; (b) final positions.

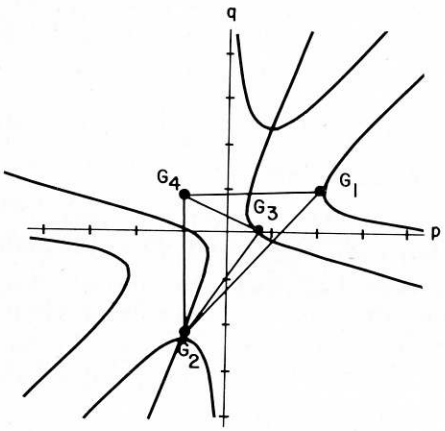


FIG. 40. Gradients of S_1 through S_4 for the 'negative' chair.

If we begin the iterative minimization for the interpretation of Fig. 38 with the corresponding initial locations, it ends up with the assignments shown in Fig. 40. This is exactly the reversal of Fig. 39; i.e., $G_i = (p, q)$ in Fig. 39(b) corresponds to $G_i = (-p, -q)$ in Fig. 40.

10.3.2. *Completion of the shape recovery*

So far we have recovered the shape of the substructure made of the surfaces S_1, S_2, S_3 and S_4 (which happens to be the seat and the back) in the 'chair' scene,

for both cases of the 'positive' and the 'negative' chairs. The scene includes more surfaces. We will determine their orientations and relative depths. However, as opposed to the foregoing subsections, the reasoning in this subsection admittedly lack firm theoretical bases. One big reason for this weakness is that at this point we cannot do much without semantics or, more directly, a model of 'chairs'. Discussion concerning the role of object models is found in Section 11. For the time being let us proceed and complete the shape recovery of the 'chair' scene without explicitly mentioning semantics.

Let us refer to Fig. 37. First of all, S_5 (which happens to be one of the arms of the chair) has to be given the same orientation as S_4: the junction labels given to the four T junctions (V_5, V_6, V_7, and V_8: see Fig. 35), where S_4 is connected with S_5, together result in this requirement. Then we can see that the 'negative' chair (Fig. 36 or Fig. 38) is not realizable by planes, because if S_4 and S_5 are in the same orientation, S_5 cannot cross and occlude S_2 or S_3, each of which connects with S_4 along a concave edge. However, if we relax this requirement by 'loosening' the connection of S_4 and S_5 at the four T junctions (we can make S_4, say, be connected with S_5 at two points V_6 and V_8 so that it can come in front of S_2 and S_3), then we shall have the 3-D shape of a 'negative' chair. Appendix C shows pictures of a model of a 'negative' chair.

Let us proceed with the 'positive' chair. S_7, S_8, and S_9 (which happen to be the legs of the chair) have individually a single constraint with the substructure (S_1, S_2, S_3, and S_4) whose shape has been determined. S_7, for instance, is connected to S_1 at its left lower corner but it is not enough of a constraint to uniquely determine the orientation of S_7. Tentatively (without a definite reason), S_7 is given the same gradient as S_1. Similarly we give S_8 and S_9 the same gradient as S_4 with which they are connected.

The surfaces whose gradients have been determined so far (S_1 through S_5, and S_7 through S_9) are all connected. Therefore, if we assume the z of any one point on the structure, we can determine their plane formulas.

What can we do about the remaining two surfaces, S_6 and S_{10}, which are separate nodes in the SCG? The visible part of S_6 (which happens to be the other arm) strongly matches S_5 with respect to both color and shape, plus several pairs of parallel boundaries. This may provide a good reason to give S_6 the same gradient as S_5. How about its relative depth (position)? Since S_5 is touching S_2 and S_3 on this side, and the visible part of S_6 is the same in the image as a part of S_5, we may conclude that S_6 is also touching S_2 and S_3 on their other side.

Finally, as for S_{10} (which happens to be a wainscot), we have no clue except that it is behind the object. Thus all we can do is to assign the gradient $(0, 0)$ (the orientation perpendicular to the present view direction) as a default, and assign a big enough depth so that it is behind the structure made of S_1 through S_9.

In this way we have determined the plane formulas of all the surfaces. The

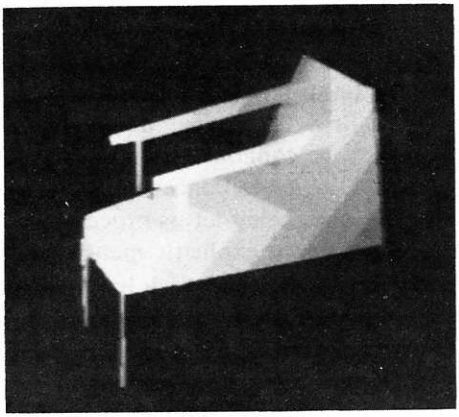


FIG. 41. A rotated view of the 'positive' chair interpretation; 10° to the left and 10° above the original.

resultant 3-D shape description can be supplied to a display program which generates other views of the object. Fig. 41 shows a rotated view of the 'positive' chair interpretation of the 'chair' scene.

11. Concluding Remarks

11.1. Summary of the results

We have demonstrated how the 3-D descriptions for simple scenes can be achieved systematically from a single view by exploiting a few assumptions. The assumptions we have used are:

(A1) Objects are planar-surfaced ones with restricted configurations at vertices.

(A2) The meta-heuristic concerns nonaccidental regularities in the picture; in particular,

(A2.1) similarity of color edge profiles,

(A2.2) parallelism of lines, and

(A2.3) skewed symmetry.

We have developed the theories and techniques for exploiting these assumptions. The theory of the Origami world for (A1) provides a labeling procedure which can recover qualitative shapes of line drawings together with constraints on surface orientations. The analysis of color edge profiles for (A2.1) provides constraints on line label combinations with weights, and thus can order the interpretations which are all geometrically possible. The parallel-line heuristic for (A2.2) and the skewed-symmetry heuristic for (A2.3) map the picture properties into constraints in the gradient space. By putting these all together

we can uniquely determine the surface orientations and obtain the 3-D quantitative shape descriptions of the object.

Notice that the up-to-3-surface assumption of the Origami world for (A1) is independent of the image-regularity assumption (A2). We could have selected the trihedral assumption for (A1). The Origami world, which is less restricted than the trihedral world, allowed to interpret those realistic line drawings such as the 'box' scene. In addition, because of its diversity, it helped in emphasizing the importance of other kinds of assumptions of (A2). One additional comment on the skewed-symmetry heuristic: it does not assume symmetrical objects but assumes local symmetry of the surfaces whose projection (picture) is skewed-symmetrical.

We have used only similarity of edge profiles, parallelism of lines, and skewed symmetry in this paper. There are other instances of image regularities which are usable in the same way by formalizing them under the meta-heuristic of non-accidental regularities. For example, equal-length lines, nearly right angles, etc. This topic will be discussed in [10], and further developed by Kender [11].

The essential issue pursued in this paper is how the properties in the picture (picture-domain cues) can be related to the properties in the scene (scene-domain cues). The importance of this distinction is emphasized by Kanade [8]. In our discussion, the scene-domain cues are the physical meaning of lines signified by the labels and the surface orientations represented by the gradients. The picture-domain cues are the junction types, the direction of lines, the shapes of regions, and the paralellisms of lines in the picture.

How the two different levels of cues interact mutually for the shape recovery is noteworthy. From the junction types and directions of individual lines, we have obtained the labelings, which can tell the shapes of which regions and the parallelisms of which lines are usable. This in turn can be converted into constraints on the gradients.

11.2. 'Natural' interpretations and the roles of object models

When presented the multiple interpretations which the theory of the Origami world yields, one of the common impressions we tend to have is that most of them are 'unnatural'. For example, the labeling of Fig. 13(b) which corresponds to a 'squashed' box may look 'unnatural' for a human observer. However, any labeling is geometrically no more 'natural' or 'unnatural' than others: they are equally possible. The feeling of 'naturalness' is due to our other presumptions.

This paper has shown that the exploitation of the heuristics which prefer the conservation of regularities allows us to select the 'natural' interpretations for our simple example scenes. The core of our method is still purely geometrical. We have tried to keep from using specific object models in order to emphasize the geometric aspects in the shape recovery.

Of course, though, I do not intend to claim that object models are unnecessary. As we have seen in the shape recovery of the 'chair' scene, the 'arms' and 'legs' can be attached to the main structure ('back' and 'seat') at any angle. We might be able to do a little more by pursuing some higher-level regularities, say the 3-D symmetries in this case, but eventually the determination of their particular shape should rely on our understanding of the functional shape that 'chairs' usually take. However, who can know that the object is a 'chair' before knowing that it could be of a 'chair' shape?

'Chairs' would have no particular predefined shapes but usually are defined by the descriptions of their functional shapes: say, an L-shaped main structure made of a 'seat' and a 'back', both usually flat; often four legs attached to the lower corners of the seat; optional two 'arms' attached symmetrically to the main structure, etc. Like this, the generic models of objects are described in terms of general 3-D shapes and relations (i.e., scene-domain cues). Therefore, in order to access appropriate models for the top–down use of semantic information, we have to first reach certain shape descriptions, either qualitative or quantitative, from the picture in a data-driven manner. Once the appropriate model is found, the general hypothesis-and-test mechanism begins to work. The theory and the technique in this paper have demonstrated the crucial point in interfacing between the model-driven part and the data-driven part of image understanding.

ACKNOWLEDGMENT

I thank John Kender, Allen Newell, Raj Reddy, and Steven Shafer for many stimulating discussions. John Zsarnay provided excellent programming support for the color edge profile analysis and the display program for generating rotated views of objects.

Appendix A

Let us take two gradient-space lines passing through the origin in the directions of α and β, as shown in Fig. A.1(a). We will prove that the gradients (such as $\boldsymbol{G}_e$ or $\boldsymbol{G}_{e'}$) which are on the bisectors of the two lines correspond to the orientation of the planar surfaces for which the length metrics along the directions of α and β in the picture are equal; i.e., the ratio of lengths along them in the picture represents the real ratio.

In general, for a surface with gradient $\boldsymbol{G} = (p, q)$, the change in depth (z) corresponding to the transposition $\Delta\boldsymbol{v} = (\Delta x, \Delta y)$ in the picture is

$$\Delta z = -p\Delta x - q\Delta y = -\boldsymbol{G} \cdot \Delta\boldsymbol{v}\,. \tag{A.1}$$

We can generally express the gradients on one of the bisectors l_1 as

$$\boldsymbol{G}_e = (\sigma \cos((\alpha + \beta)/2), \sigma \sin((\alpha + \beta)/2))\,. \tag{A.2}$$

Similarly the gradients on the other bisector l_2 are

$$\boldsymbol{G}_{e'} = (-\sigma \sin((\alpha + \beta)/2), \sigma \cos((\alpha + \beta)/2))\,. \tag{A.3}$$

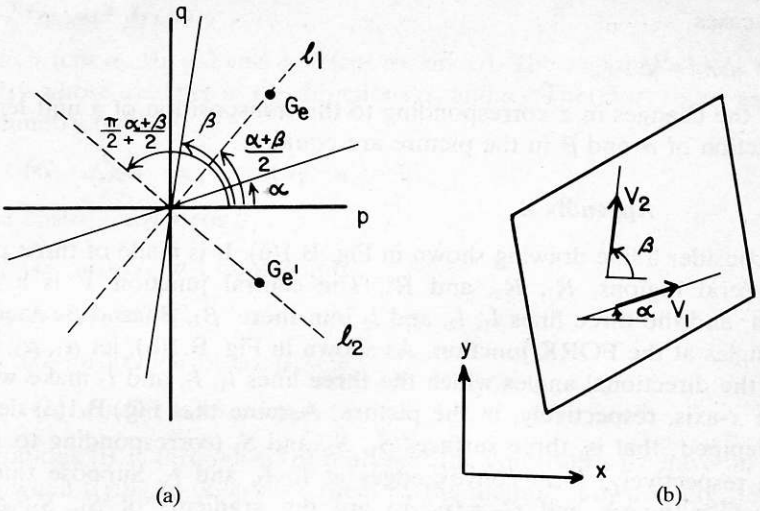


FIG. A.1. (a) Two gradient-space lines passing through the origin in the direction of α and β. G_e and $G_{e'}$ are the gradients lying on their bisectors. (b) The projection of a surface and two picture lines in the direction of α and β.

Now suppose that surface with a gradient G_e or $G_{e'}$ is projected onto a picture plane, and consider two picture lines in the direction of α and β as in Fig. A.1(b). What we have to show is that the same length along these two picture lines depict the same length on the surface in the 3-D space, or, equivalently, that the rate of change in depth (z) to S is equal along these two lines.

We can represent unit-length vectors along these picture lines as follows, respectively:

$$v_1 = (\cos\alpha, \sin\alpha) \quad \text{and} \quad v_2 = (\cos\beta, \sin\beta)\,. \tag{A.4}$$

The changes in z corresponding to v_1 and v_2 are, if S has the gradient G_e,

$$\begin{aligned}\Delta z_1 &= -G_e \cdot v_1 = -\sigma\cos((\alpha+\beta)/2)\cos\alpha - \sigma\sin((\alpha+\beta)/2)\sin\alpha \\ &= -\sigma\cos((\alpha-\beta)/2)\,, \end{aligned} \tag{A.5}$$

$$\begin{aligned}\Delta z_2 &= -G_e \cdot v_2 = -\sigma\cos((\alpha+\beta)/2)\cos\beta - \sigma\sin((\alpha+\beta)/2)\sin\beta \\ &= -\sigma\cos((\alpha-\beta)/2)\,. \end{aligned} \tag{A.6}$$

and if S has the gradient $G_{e'}$

$$\Delta z_1 = -G_{e'} \cdot v_1 = -\sigma\sin((\alpha-\beta)/2)\,,$$

and

$$\Delta z_2 = -G_{e'} \cdot v_2 = \sigma\sin((\alpha-\beta)/2)\,. \tag{A.7}$$

In both cases

$$|\Delta z_1| = |\Delta z_2| . \tag{A.8}$$

That is, the changes in z corresponding to the transposition of a unit length in the direction of α and β in the picture are equal.

Appendix B

Let us consider a line drawing shown in Fig. B.1(a). It is made of three parallel quadrilateral regions, R_1, R_2, and R_3. The central junction V is a FORK junction, and the three lines l_1, l_2, and l_3 join there. β_1, β_2, and β_3 denote the three angles at the FORK junction. As shown in Fig. B.1(b), let α_1, α_2, and α_3 denote the directional angles which the three lines l_1, l_2, and l_3 make with the positive x-axis, respectively, in the picture. Assume that Fig. B.1(a) depicts a parallelepiped; that is, three surfaces S_1, S_2, and S_3 (corresponding to R_1, R_2, and R_3 respectively) form convex edges at l_1, l_2, and l_3. Suppose that $\boldsymbol{G}_1 = (p_1, q_1)$, $\boldsymbol{G}_2 = (p_2, q_2)$, and $\boldsymbol{G}_3 = (p_3, q_3)$ are the gradients of S_1, S_2, and S_3, respectively.

Since S_1 and S_2 intersect along l_2,

$$p_1 \cos \alpha_2 + q_1 \sin \alpha_2 = p_2 \cos \alpha_2 + q_2 \sin \alpha_2 ,$$

or, in brief,

$$\boldsymbol{G}_1 \cdot \boldsymbol{a}_2 = \boldsymbol{G}_2 \cdot \boldsymbol{a}_2 \tag{B.1}$$

where $\boldsymbol{a}_2 = (\cos \alpha_2, \sin \alpha_2)$. Similarly, we have

$$\boldsymbol{G}_2 \cdot \boldsymbol{a}_3 = \boldsymbol{G}_3 \cdot \boldsymbol{a}_3 \tag{B.2}$$

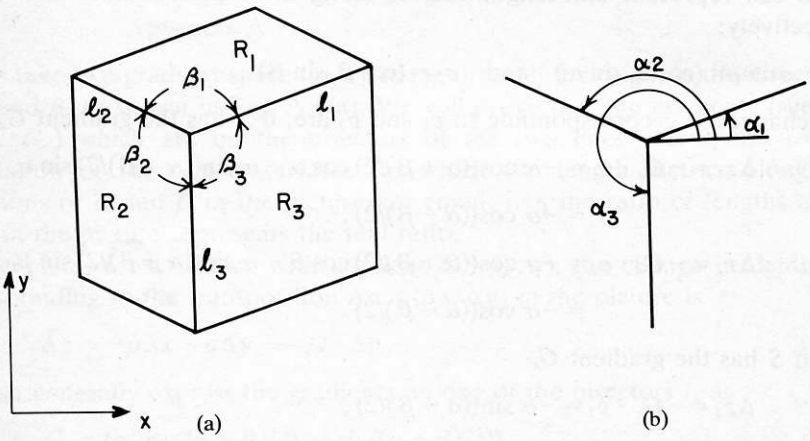


FIG. B.1. (a) A line drawing of a parallelepiped. (b) The directional angles α_1, α_2, and α_3 of the three lines joining at the FORK junction.

$$\boldsymbol{G}_3 \cdot \boldsymbol{a}_1 = \boldsymbol{G}_1 \cdot \boldsymbol{a}_1 , \tag{B.3}$$

where $\boldsymbol{a}_3 = (\cos\alpha_3, \sin\alpha_3)$ and $\boldsymbol{a}_1 = (\cos\alpha_1, \sin\alpha_1)$. The region R_1 has a skewed symmetry whose axes are in the directions α_1 and α_2. Therefore, if we assume a real symmetry for S_1,

$$(\boldsymbol{G}_1 \cdot \boldsymbol{a}_1)(\boldsymbol{G}_1 \cdot \boldsymbol{a}_2) + \cos(\alpha_1 - \alpha_2) = 0$$

or, since $\cos(\alpha_1 - \alpha_2) = \cos\beta_1$,

$$(\boldsymbol{G}_1 \cdot \boldsymbol{a}_1)(\boldsymbol{G}_1 \cdot \boldsymbol{a}_2) + \cos\beta_1 = 0 . \tag{B.4}$$

Similarly, for R_2 and R_3,

$$(\boldsymbol{G}_2 \cdot \boldsymbol{a}_2)(\boldsymbol{G}_2 \cdot \boldsymbol{a}_3) + \cos\beta_2 = 0 , \tag{B.5}$$

$$(\boldsymbol{G}_3 \cdot \boldsymbol{a}_3)(\boldsymbol{G}_3 \cdot \boldsymbol{a}_1) + \cos\beta_3 = 0 . \tag{B.6}$$

The equations (B.1) through (B.6) represent the constraints we have on $\boldsymbol{G}_1$, $\boldsymbol{G}_2$ and $\boldsymbol{G}_3$, when trying to determine them by the method shown in Fig. 18.

First, let us see that none of β_i's can be $\pi/2$. For example, suppose $\beta_1 = \pi/2$. Then $\cos\beta_1 = 0$. From (B.4), either $\boldsymbol{G}_1 \cdot \boldsymbol{a}_1$ or $\boldsymbol{G}_1 \cdot \boldsymbol{a}_2$ is zero. Suppose $\boldsymbol{G}_1 \cdot \boldsymbol{a}_1 = 0$. Then from (B.3) and (B.6) we have $\cos\beta_3 = 0$; i.e., $\beta_3 = \pi/2$. Now $\beta_1 = \beta_3 = \pi/2$ is contradictory to the assumption that the junction V is a FORK junction. In this way we know that $\beta_i \neq \pi/2$, or $\cos\beta_i \neq 0$ for $i = 1, 2, 3$.

From (B.1) through (B.6) we can derive

$$(\boldsymbol{G}_1 \cdot \boldsymbol{a}_1)^2 = -\cos\beta_1 \cos\beta_3 / \cos\beta_2 > 0 . \tag{B.7}$$

Since $\pi > \beta_i > 0$ and $\beta_1 + \beta_2 + \beta_3 = 2\pi$, we see that $\pi > \beta_i > \pi/2$ (obtuse angle) for $i = 1, 2, 3$ is necessary for the right side of (B.7) to be positive.

Now assume β_i's are all obtuse. Since $\cos\beta_i < 0$, let us put

$$\cos\beta_i = -c_i , \quad c_i > 0 \quad \text{for } i = 1, 2, 3 .$$

Then (B.7) becomes

$$\boldsymbol{G}_1 \cdot \boldsymbol{a}_1 = \pm(c_1 c_3 / c_2)^{1/2} . \tag{B.8}$$

Substitution of (B.8) into (B.4) gives

$$\boldsymbol{G}_1 \cdot \boldsymbol{a}_2 = -\cos\beta_1 / (\boldsymbol{G}_1 \cdot \boldsymbol{a}_1) = \pm(c_1 c_2 / c_3)^{1/2} . \tag{B.9}$$

We can solve (B.8) and (B.9) for p_1 and q_1.

$$\begin{pmatrix} p_1 \\ q_1 \end{pmatrix} = \pm \frac{(c_1/c_2c_3)^{1/2}}{\sin(\alpha_2 - \alpha_1)} \begin{pmatrix} c_3 \sin\alpha_2 - c_2 \sin\alpha_1 \\ -c_3 \cos\alpha_2 + c_2 \cos\alpha_1 \end{pmatrix} . \tag{B.10}$$

Similarly we obtain the following. (The multiple signs are to be read correspondingly in (B.10), (B.11), and (B.12).)

$$\begin{pmatrix} p_2 \\ q_2 \end{pmatrix} = \pm \frac{(c_2/c_1c_3)^{1/2}}{\sin(\alpha_3 - \alpha_2)} \begin{pmatrix} c_1 \sin\alpha_3 - c_3 \sin\alpha_2 \\ -c_1 \cos\alpha_3 + c_3 \cos\alpha_2 \end{pmatrix} , \tag{B.11}$$

$$\begin{pmatrix} p_3 \\ q_3 \end{pmatrix} = \pm \frac{(c_3/c_1c_2)^{1/2}}{\sin(\alpha_1 - \alpha_3)} \begin{pmatrix} c_2 \sin\alpha_1 - c_1 \sin\alpha_3 \\ -c_2 \cos\alpha_1 + c_1 \cos\alpha_3 \end{pmatrix}. \tag{B.12}$$

Thus there are two combinations of G_i's which satisfy (B.1) through (B.6). Apparently they are symmetrical through the gradient-space origin, and correspond to the two labelings of Fig. B.1(a): the convex-corner (three lines at the FORK junction are all convex lines) and the concave-corner (they are all concave). Therefore we have shown that for the convex-corner interpretation we always obtain the unique assignment of G_i's which satisfies the skewed-symmetry heuristic if and only if β_i's are all obtuse.

Next we will prove that the parallelepiped is right-angled in the above assignment of gradients. Let us consider S_1 and S_2, for instance. From (B.10) and (B.11), we can compute

$$\begin{aligned} p_1p_2 + q_1q_2 &= [(c_3 \sin\alpha_2 - c_2 \sin\alpha_1)(c_1 \sin\alpha_3 - c_3 \sin\alpha_2) \\ &\quad + (-c_3 \cos\alpha_2 + c_2 \cos\alpha_1)(-c_1 \cos\alpha_3 + c_3 \cos\alpha_2)] \\ &\quad \times [c_3 \sin(\alpha_2 - \alpha_1) \sin(\alpha_3 - \alpha_2)]^{-1} \\ &= [c_1c_3 \cos(\alpha_3 - \alpha_2) - c_1c_2 \cos(\alpha_1 - \alpha_3) - c_3^2 \\ &\quad + c_2c_3 \cos(\alpha_2 - \alpha_1)]/[c_3 \sin(\alpha_2 - \alpha_1) \sin(\alpha_3 - \alpha_2)] \\ &= -(c_3 + c_1c_2)/\sin(\alpha_2 - \alpha_1) \sin(\alpha_3 - \alpha_2) \\ &= -[-\cos(\alpha_3 - \alpha_1) + \cos(\alpha_2 - \alpha_1) \cos(\alpha_3 - \alpha_2)] \\ &\quad \times [\sin(\alpha_2 - \alpha_1) \sin(\alpha_3 - \alpha_2)]^{-1} \\ &= -1 . \end{aligned}$$

That is, $p_1p_2 + q_1q_2 + 1 = 0$. Since $(p_i, q_i, 1)$ is the surface normal, we have proven that S_1 and S_2 intersect perpendicularly. We can prove the same for the other pairs of surfaces.

Appendix C. Pictures of a 'negative' chair

A wooden model of a 'negative' chair (only the main structure including the seat, back and arms) was actually made. The picture (1) is taken from the angle such that the object looks like a positive (real) chair. The sequence of pictures above, below and right show how it appears as we move our eye position upward, downward and to the right, correspondingly.

An interesting phenomenon happens when we hung the negative chair in the air by string and swing it a little as our eye position is fixed where we see the picture (1) first. Then what we perceive is that a real (positive) chair is flexing its arms back and forth, rather than a rigid object of funny shape is swinging. The moving parallax does not help much in perceiving the real shape. This is probably because once we fix the shape description as a real chair, we try to interpret the sequence of pictures consistently and one way for that is to regard it flexing.

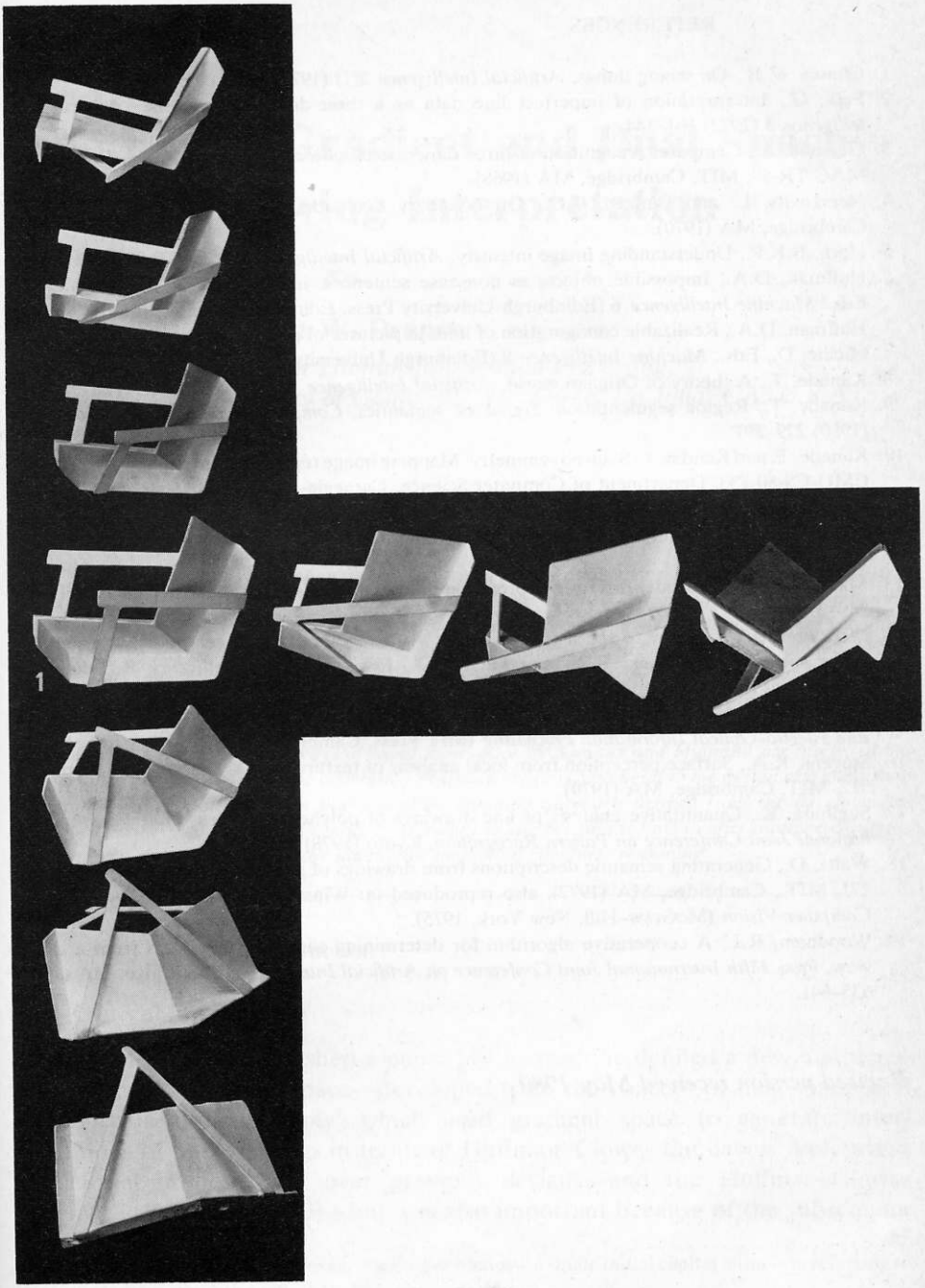


FIG. C.1. Pictures of a 'negative' chair.

REFERENCES

1. Clowes, M.B., On seeing things, *Artificial Intelligence* **2**(1) (1971) 79–116.
2. Falk, G., Interpretation of imperfect line data as a three-dimensional scene, *Artificial Intelligence* **3** (1972) 101–144.
3. Guzman, A., Computer recognition of three dimensional objects in a visual scene, Tech. Rept. MAC-TR-59, MIT, Cambridge, MA (1968).
4. Herskovits, L. and Binford, T.O., On boundary detection, MAC. AI Memo 183, MIT, Cambridge, MA (1970).
5. Horn, B.K.P., Understanding image intensity, *Artificial Intelligence* **8**(2) (1977) 201–231.
6. Huffman, D.A., Impossible objects as nonsense sentences, in: Meltzer, B. and Michie, D., Eds., *Machine Intelligence* **6** (Edinburgh University Press, Edinburgh, 1971).
7. Huffman, D.A., Realizable configuration of lines in pictures of polyhedra, in: Elcock, E.W. and Michie, D., Eds., *Machine Intelligence* **8** (Edinburgh University Press, Edinburgh, 1977).
8. Kanade, T., A theory of Origami world, *Artificial Intelligence* **13**(1) (1980) 279–311.
9. Kanade, T., Region segmentation: Signal vs. semantics, *Comput. Graphics Image Process* **13** (1980) 279–297.
10. Kanade, T. and Kender, J., Skewed symmetry: Mapping image regularities into shape, Tech. Rept. CMU-CS-80-133, Department of Computer Science, Carnegie–Mellon University (1980).
11. Kender, J., Ph.D. Thesis, Carnegie–Mellon University, Pittsburgh (1980).
12. Mackworth, A.K., Interpreting pictures of polyhedral scenes, *Artificial Intelligence* **4**(2) (1974) 121–137.
13. Mackworth, A.K., Model-driven interpretation in intelligent vision systems, *Perception* **5** (1976) 349–370.
14. Mackworth, A.K., How to see a simple world: An exegesis of some computer programs for scene analysis, in: Elcock, E.W. and Michie, D., Eds., *Machine Intelligence* **8** (Edinburgh University Press, Edinburgh, 1977).
15. Roberts, L.G., Machine perception of three-dimensional solids, in: Tippett et al., Eds., *Optical and Electro-Optical Information Processing* (MIT Press, Cambridge, 1965).
16. Stevens, K.A., Surface perception from local analysis of texture, Ph.D. Thesis, Tech. Rept. TR 512, MIT, Cambridge, MA (1979).
17. Sugihara, K., Quantitative analysis of line drawings of polyhedral scenes, *Proc. Fourth International Joint Conference on Pattern Recognition*, Kyoto (1978) 771–773.
18. Waltz, D., Generating semantic descriptions from drawings of scenes with shadows, MAC-TR-271, MIT., Cambridge, MA (1972), also reproduced in: Winston, P., Ed., *The Psychology of Computer Vision* (McGraw–Hill, New York, 1975).
19. Woodham, R.J., A cooperative algorithm for determining surface orientations from a single view, *Proc. Fifth International Joint Conference on Artificial Intelligence*, Cambridge, MA (1977) 635–641.

Revised version received May 1980