COMPUTER VISION AND IMAGE UNDERSTANDING
Vol. 64, No. 1, July, pp. 1–20, 1996
ARTICLE NO. 0043

Perceptual Completion of Occluded Surfaces

LANCE R. WILLIAMS*

NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540

AND

ALLEN R. HANSON†

Department of Computer Science, University of Massachusetts, Amherst, Massachusetts 01003

Received February 1, 1994; accepted June 20, 1995

Researchers in computer vision have primarily studied the problem of visual reconstruction of environmental structure that is plainly visible. In this paper, the conventional goals of visual reconstruction are generalized to include both visible and occluded forward facing surfaces. This larger fraction of the environment is termed the *anterior surfaces.* Because multiple anterior surface neighborhoods project onto a single image neighborhood wherever surfaces overlap, surface neighborhoods and image neighborhoods are not guaranteed to be in one-to-one correspondence, as conventional "shape-from" methods assume. The result is that the topology of three-dimensional scene structure can no longer be taken for granted, but must be inferred from evidence provided by image contours. In this paper, we show that the boundaries of the anterior surfaces can be represented in viewer-centered coordinates as a *labeled knot-diagram.* Where boundaries are not occluded and where surface reflectance is distinct from that of the background, boundaries will be marked by image contours. However, where boundaries are occluded, or where surface reflectance matches background reflectance, there will be no detectable luminance change in the image. Deducing the complete image trace of the boundaries of the anterior surfaces under these circumstances is called the *figural completion* problem. The second half of this paper describes a computational theory of figural completion. In more concrete terms, the problem of computing a labeled knot-diagram representing an anterior scene from a set of contour fragments representing image luminance boundaries is investigated. A working model is demonstrated on a variety of illusory contour displays. The experimental system employs a two-stage process of completion hypothesis and combinatorial optimization. The labeling scheme is enforced by a system of integer linear inequalities so that the final organization is the optimal feasible solution of an integer linear program.

* E-mail: williams@research.nj.nec.com.
† E-mail: hanson@cs.umass.edu.

1. INTRODUCTION

By considering only the visible surfaces, conventional models avoid the most difficult part of visual reconstruction—deducing the topology of 3D scene structure. Conventionally, the topology of the reconstructed scene is trivially determined by the assumption that the imaging process maps visible surface neighborhoods to image neighborhoods in one-to-one fashion. However, introspection suggests that the human visual system reconstructs a larger fraction of the environment. Imagine a pool ball sitting on the felt surface of a pool table. You would be quite surprised if upon picking up the pool ball, you discovered that a hole of its exact size and shape lay behind it. What can account for this surprise, other than the fact that new information contradicts previously held unconscious inferences about structure other than the visible surfaces?

In this paper, the goal of visual reconstruction is generalized to include the fraction of the environmental surfaces that are potentially visible (because they are forward facing) but are possibly occluded by intervening surfaces. Unfortunately, surface neighborhoods and image neighborhoods are no longer guaranteed to be in one-to-one correspondence. Occlusion confounds the visual mapping since multiple surface neighborhoods project onto a single image neighborhood wherever surfaces overlap. Consequently, the topology of 3D structure can no longer be taken for granted but must instead be inferred from the fragmentary evidence provided by image contours. This process is termed *figural completion*.

2. PAST WORK

The relevant literature can be divided into two categories: (1) computational theories and (2) algorithm and representation level theories. By computational theory, we mean a detailed analysis of the computational goal, natural

1077-3142/96 $18.00

constraints, and inherent ambiguities, leading to a definition of the function which maps the available input to a representation sufficient to explain the observed human competence. Ideally, this function is described independently and in advance of an algorithm and representation level theory [20].

Like us, Nitzberg and Mumford [21] have studied figural completion in illusory contour figures such as the Kanizsa triangle (see Fig. 1). They also describe a combinatorial optimization approach where potential completions are explicitly represented by curves of least energy. The objective function they minimize includes terms which implement a preference for organizations consisting of low-energy completions bounding regions of uniform brightness. Unlike us, the surface representation they compute (i.e., their computational goal) consists of sets of closed, nonself-intersecting plane curves (i.e., Jordan curves) of constant depth. Consequently, surfaces with boundaries which project as self-intersecting curves in the image plane cannot be represented.

Related algorithm and representation level theories have been advanced by numerous researchers. The majority of these address only limited aspects of the figural completion problem, such as the shape of illusory contours [10, 11, 28, 30]. Algorithm and representation level theories of significantly broader scope are described by Grossberg and Mingolla [9] and Finkel and Sajda [7]. Both of these theories consist of descriptions of neural networks with multiple layers which solve distinct subproblems in cooperative fashion. For example, Grossberg and Mingolla divide the figural completion problem into two subsystems which they term the boundary contour system (BCS) and the feature contour system (FCS). The former deals with completing gaps in contours, the latter with completing surfaces within boundaries.[1]

3. COMPUTATIONAL GOAL

The stated goal of all "shape-from" methods is computing the depth and surface orientation of the *visible surfaces*. The visible surfaces are sometimes defined as the locus of surface points where the surface normal has a positive component in the viewing direction. This is precisely the subset of the environmental surfaces with orientations that can be represented as points in gradient space. Of course this definition only accounts for self-occlusion (i.e., the nonvisibility of backward facing surface patches) and does not account for nonlocal occlusion (i.e., nonvisibility due to intervening surfaces). Having a positive component in the viewing direction is therefore a necessary but not a sufficient condition for visibility. For this reason, in this paper, we distinguish between the *visible surfaces* and the *anterior surfaces:*

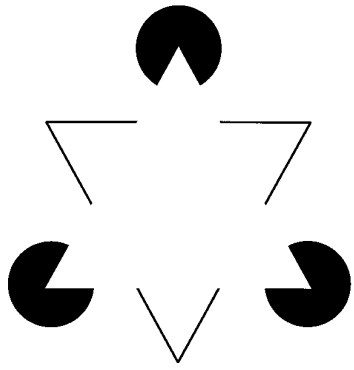

FIG. 1. The Kanizsa triangle.

DEFINITION. Visible surfaces: the locus of environmental surface points first incident along the lines of sight.

DEFINITION. Anterior surfaces: the locus of environmental surface points where the surface normal is defined and has a positive component in the viewing direction.

The difference between the visible surfaces and the anterior surfaces is illustrated through a series of figures beginning with Fig. 2a. This simple ray-traced image of a sphere and cone is illuminated by a point source coincident with the location of the viewer, so there are no visible shadows. Figure 2b depicts a sideview of the same scene with the illumination unchanged. Because of the location of the light source, surfaces not visible from the first viewpoint lie in shadow. The outlines of the visible and anterior surfaces (with respect to the first viewpoint) appear in sideview in Figs. 2c and 2d.

Perhaps it is not surprising that occluded surfaces are ignored by conventional models, after all, surface patches which are not visible do not contribute to image brightness. However, the most important reason is that the image irradiance equation [1] presupposes a continuous and invertible mapping between image neighborhoods and visible surface neighborhoods. Such a mapping is called a *homeomorphism*. Under homeomorphism, the visible surfaces are *embedded* in the image plane. This is why it is possible to speak of the depth and surface orientation of image point x, y without ambiguity: depth and surface orientation are assumed to be a function of image coordinates. This is also why shape-from methods can operate on image neighborhoods and reconstruct surface neighborhoods: They are assumed to be in one-to-one correspondence.

[1] It is useful to compare Grossberg and Mingolla's algorithm and representation level theory with the computational theory we describe here. While Grossberg and Mingolla describe mechanisms (i.e., the BCS and FCS), we describe corresponding computational goals (i.e., the *labeled knot-diagram* and *paneling*). For this reason, we believe the two theories are complementary.

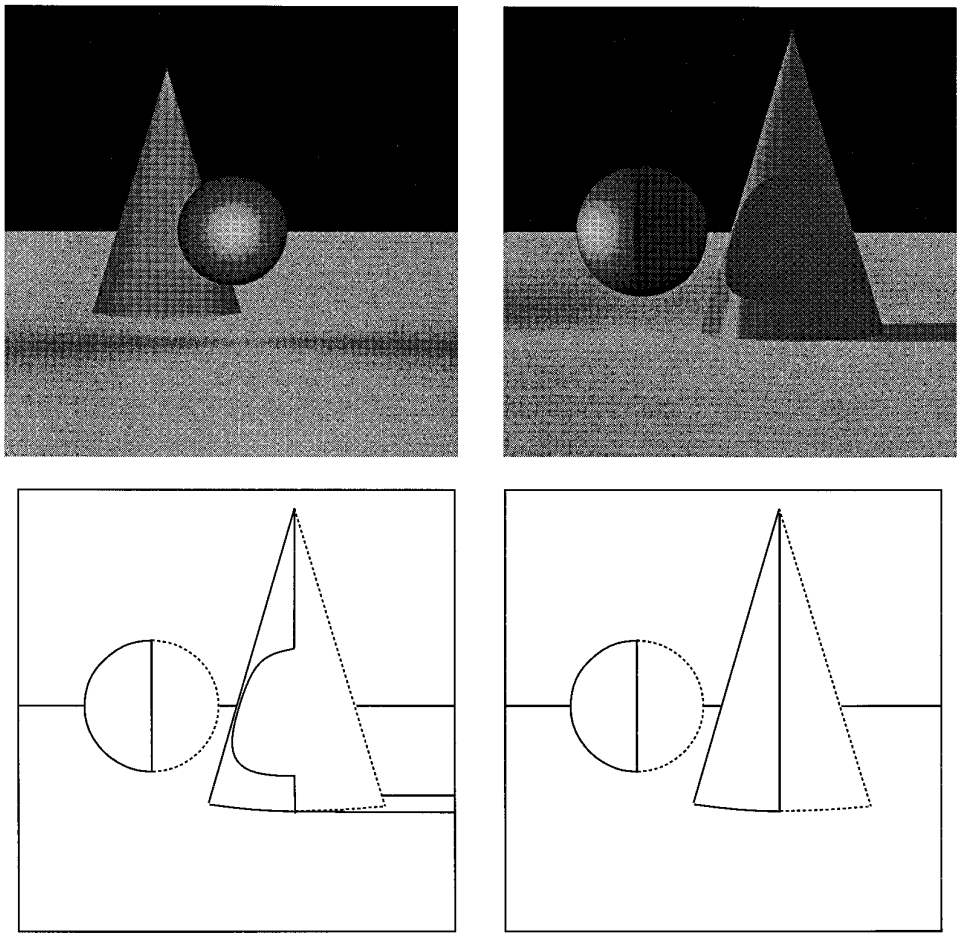

FIG. 2. Upper left: A simple ray-traced image of a sphere and a cone. This scene is illuminated by a point source coincident with the location of the viewer, so there are no visible shadows. Upper right: A sideview of the same scene with the illumination unchanged. Because of the location of the light source, surfaces not visible from the first viewpoint lie in shadow. Lower left: The *visible surfaces* are defined as the locus of surface points first incident along lines of sight. Lower right: The *anterior surfaces* are defined as the locus of surface points where the surface normal is defined and has a positive component in the viewing direction. The visible surfaces are a subset of the anterior surfaces. Conventional models only address the problem of reconstructing the visible surfaces.

Although homeomorphism may be an acceptable approximation of the mapping of visible surface neighborhoods onto the image plane, if the goal of visual reconstruction is expanded from reconstruction of the visible surfaces to reconstruction of the anterior surfaces (for example), then the assumption of embedding breaks down. Occlusion confounds the visual mapping of surface neighborhoods to image neighborhoods since multiple surface neighborhoods will project to one image neighborhood wherever surfaces overlap. The problem of inferring the neighborhood structure of some fraction of the environment (rather than simply taking it for granted) thus appears for the first time. This is the crux of the perceptual completion problem, and of perceptual organization problems in general: What are the neighborhoods? This is among the most difficult problems facing the field of computer vision today. Certainly, part of the appeal of the standard model is that it allows some aspects of the visual reconstruction problem to be studied without first having to solve the more difficult problem of deducing the topology of environmental structure.[2]

In this paper, it is hypothesized that the goal of early visual processing is to compute a viewer-centered representation of the anterior surfaces, of which the visible surfaces are merely a subset. The assumption of global homeomorphism between visible surface neighborhoods and image neighborhoods (i.e., *embedding*) is generalized[3] to the assumption of local homeomorphism between anterior sur-

[2] In a recent article, Barrow and Tenenbaum [2] confirm this. They state that "[o]ur interest in recovering scene characteristics arose in part through a belief that it was not possible to segment an image reliably into meaningful regions and boundaries on the basis of raw brightness. . . . However, the baby may have been thrown out with the bath water, and should perhaps be rescued: perceptual organization may play a much larger role. . ."

[3] This seems to be the simplest of the possible generalizations.

face neighborhoods and image neighborhoods (i.e., *immersion*). Since the visual mapping is locally singular only where a surface is tangent to the viewing direction, local singularities will never occur in images of smooth surfaces embedded in three-space so that the surfaces are nowhere tangent. This leads to the following definition:

DEFINITION. Anterior scene: a set of surfaces embedded in three space so that the surface normals everywhere are defined and have a positive component in the viewing direction.

By definition, no singularities can exist in the parallel projection of an anterior scene onto the image plane. It follows that the visual mapping can be modeled as an immersion (see [16] for a discussion of planar surface immersions).

A topological circle embedded in three-space is termed a knot. The projection of a knot onto a plane is called a knot-diagram, and in general, consists of a closed plane-curve which intersects itself at a finite number of points called crossings. In this paper, knot-diagrams are used to represent the image of surface boundaries in anterior scenes. Each of the closed plane curves which together comprise the projection of the boundary onto the image plane can be assigned an orientation which everywhere indicates which side of the curve the image of the surface lies. We adopt the usual convention that the surface lies to the right as the boundary is traversed in the direction of its orientation. Additionally, each boundary point can be assigned an integer value equal to the number of surfaces lying between the point and its projected image (i.e., a depth index).

If the view of the anterior scene is generic, then the crossings will be the only points of multiplicity two in the projection of the boundary onto the plane:

DEFINITION. Generic view: an image of an anterior scene where (1) the multiplicity of the image of the boundary is one everywhere except at a finite number of points where it is two and (2) the number of multiplicity two points is invariant to small changes in the viewing direction.

We observe that the depth index can change only at crossing points in a generic view of an anterior scene. In a knot-diagram, crossings are drawn in a manner which explicitly indicates the relative depth of the two overlapping strands. For our purposes, the upper and lower strands of the crossing in the knot-diagram will represent the overlapping image of the nearer and farther boundaries. The depth of the farther boundary changes by one as it is occluded by the surface defined by the nearer boundary. The depth of the nearer boundary, of course, remains unchanged.

The above observations constitute a set of necessary constraints on the appearance of surface boundaries in generic views of anterior scenes. These constraints have been incorporated into the labeling scheme illustrated in Fig. 3. It can be easily verified that the depth indices of the different edges in the labeling scheme accurately describe the effect of occlusion on the boundary depth at a crossing. The labeling scheme can therefore be considered necessary in the sense that the image of the boundary of any anterior scene satisfies the constraints. But does a set of closed contours satisfying the labeling scheme always define an anterior scene? Is the labeling scheme necessary *and* sufficient?

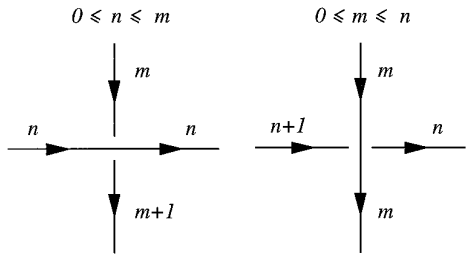

FIG. 3. A boundary labeling scheme. The image of the surface lies to the right when the image of its boundary is traversed in the direction of its orientation. Each boundary point is also assigned an integer value equal to the number of surfaces lying between the boundary and its projected image (i.e., its *depth index*). Finally, the depth of the boundary of the *occluding* surface must be less than or equal to the depth of the boundary of the *occluded* surface.

In his recent Ph.D. thesis, Williams [29] shows that this is in fact the case. Specifically, he shows that every knot-diagram satisfying the labeling scheme illustrated in Fig. 3 represents a generic view of an anterior scene. The proof is based upon the existence of a procedure for building a combinatorial model (i.e., a *paneling* [8]) of an anterior scene from a set of closed plane-curves satisfying the labeling scheme (see Fig. 4). We conclude that labeled knot-diagrams and generic views of anterior scenes are in one-to-one correspondence. Consequently, by enforcing the labeling scheme during figural completion, it is possible to ensure the topological validity and genericness of the reconstructed scene.

4. FIGURAL COMPLETION: A COMPUTATIONAL THEORY

In this section, a computational theory of figural completion is described. In more concrete terms, the problem of computing a labeled knot-diagram representing an anterior scene from a set of contour fragments representing image luminance boundaries is investigated.

The natural constraints which apply to this problem are few in number and not nearly sufficient to determine a unique solution. These constraints have two sources. The first is the requirement that the organization be a labeled knot-diagram. This can be termed the *topological validity*

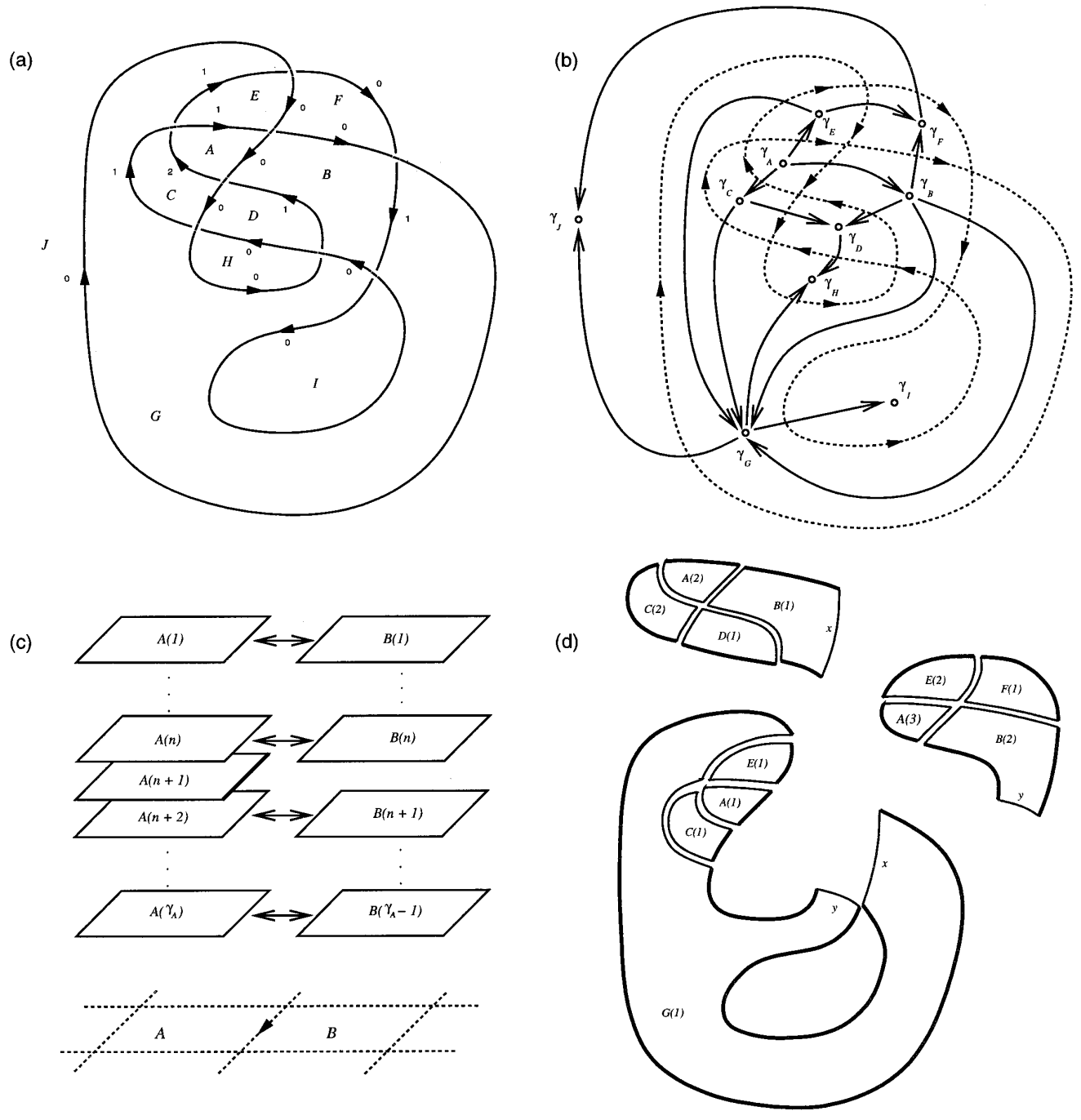

FIG. 4. (a) A *labeled knot-diagram.* (b) A network representing difference constraints on the number of surface points which project onto adjacent regions of the labeled figure. The weights of edges in the network are 1 when traversed in the direction of the arrows and -1 when traversed in the opposite direction. (c) Paper panels stacked above regions A and B in the plane. Following the *identification scheme,* all copies of regions A and B except $A(n + 1)$ are glued along their adjacent sides. The free side of $A(n + 1)$ becomes part of the boundary of the surface. (d) The *paneling* resulting from the construction. Bold edges remain free and form the boundary. Additional identifications are indicated by x and y.

requirement. Only plane-curves satisfying the labeling scheme define topologically valid anterior scenes. The second source of constraints has been stressed by Irvin Rock [25], and is termed the *stimulus conformity* requirement. Rock observed that ''the [perceptual] solution must not contradict the stimulus'' and ''must contain everything implied by the stimulus.'' Regarding illusory contour displays, Rock hypothesized that the depth of visible boundaries must be zero and that light surfaces must be visible against dark surfaces and vice versa.

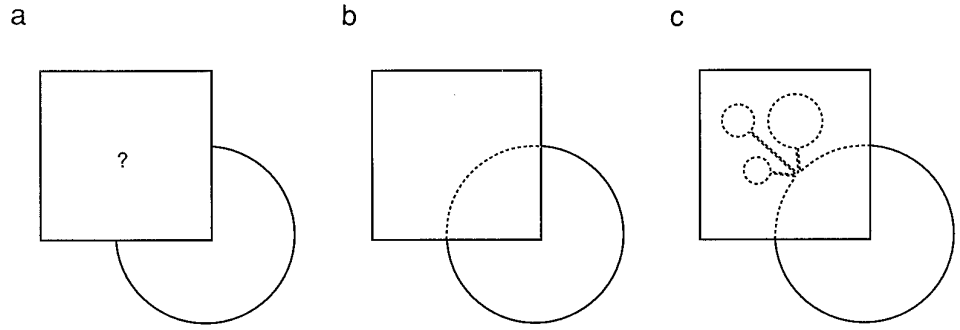

FIG. 5. Shape ambiguity. (a) A square opaque surface occluding a second surface of indeterminant shape. (b) The shape which properly functioning human visual systems infer. (c) This completion can be imagined but is not perceived preattentively.

Given only these constraints, the problem of computing a labeled knot-diagram from an image of an anterior scene remains underconstrained in three qualitatively different ways. The first kind of ambiguity can be termed *shape ambiguity*. The essence of shape ambiguity is illustrated in Fig. 5. Figure 5a apparently depicts a square opaque surface occluding a second surface of indeterminate shape. Without ''X-ray'' vision, this problem cannot be solved in any absolute sense. Even if smoothness is assumed (with no real justification), an infinitude of completion shapes which can be transformed one into another by smooth deformations in the plane still remain. Yet humans experience a particular shape, which is depicted in Fig. 5b. The other completion can be imagined but is not perceived preattentively.

Unlike shape ambiguity, the other two forms of ambiguity are combinatorial and therefore finite. The first of these is *unit ambiguity,* which is ambiguity in identifying which contour fragments match which to form boundaries (see Fig. 6). Second, there is *depth ambiguity,* which is ambiguity in the signs of occlusion of different boundary components and their relative depths at crossings (See Fig. 7).

Since the applicable physical constraints are insufficient to overcome these ambiguities, an additional assumption must be introduced. Broadly speaking, this assumption is: *the shape of a perceptual completion is independent of the role it plays in the organization*. This allows the figural completion problem to be decomposed into two independent subproblems, the first devoted to shape, the second devoted jointly to unit and depth. Many researchers have proposed that the shape of an illusory contour is described by the curve of least energy [12, 15, 21, 28]. However, Witkin and Tenenbaum [31] offer the most convincing explanation of why this should be so. They point out that while minimum energy solutions are not significantly more likely than other interpolating curves, the simple fact that a smooth low-energy interpolating curve exists is itself a reliable indicator of a *nonaccidental* relationship. This has important implications for the task at hand: the likelihood that two contour fragments form consecutive segments of the same boundary is assumed to be a function of the shape of the smooth interpolating curve of least energy joining them. We hypothesize that the human visual system resolves the ambiguity in completion shape somewhat arbitrarily, but in doing so, it gains information useful for resolving the unit ambiguity.

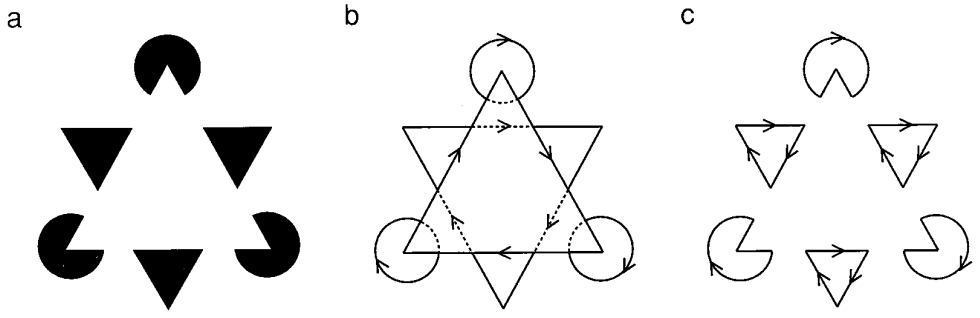

FIG. 6. Unit ambiguity. (a) A variation of the Kanizsa triangle. (b) The organization experienced by most observers. (c) The organization Rock calls the ''literal solution.'' Although both are topologically valid anterior scenes conforming to the image evidence, they contain different sets of closed boundaries.

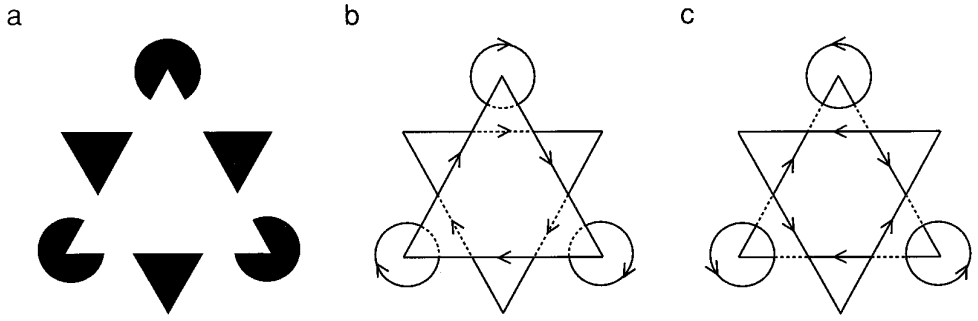

FIG. 7. Depth ambiguity. (a) A variation of the Kanizsa triangle. (b) The organization experienced by most observers. (c) A second topologically valid anterior scene conforming to the image evidence. Although each organization contains the same set of closed boundaries, individual boundaries differ in sign of occlusion and/or relative depth.

4.1. *Surface Organization*

Regardless of whether or not the shape of the curve of least energy provides inferential leverage useful for resolving unit ambiguity, significant computational gains are achieved simply by committing to a set of completions of fixed plausible shape. Since the image traces of potential completions are determined solely by the tangents and curvatures of the boundary fragments which they join, the locations of points of contour intersection (whether of completions or of completions and boundary fragments) are independent of a specific surface organization. By committing to a set of potential completions of fixed shape, the problem of constructing a labeled knot-diagram representing the surfaces in a scene becomes purely combinatorial:

1. Selecting an optimal subset of unique completions of fixed shape.
2. Enforcing the crossing labeling scheme at fixed points of contour intersection.
3. Ensuring that the depth of every contour conforms to the stimulus.

These three tasks can be combined into a single graph labeling problem. We maintain that this graph labeling problem is intrinsic to figural completion, not to a specific method of solution, and is therefore an essential part of the computational theory. To proceed with this characterization, it will be necessary to define the graph upon which the labeling problem operates.

We begin with a set of simple closed plane curves which define regions of roughly uniform brightness. These closed plane curves are segmented at tangent discontinuities to create a set of contours which will be called *boundary fragments*. The boundary fragments form the edges of a graph, $G_{\text{input}} = (V_{\text{endpoints}}, E_{\text{fragments}})$. Every vertex (representing a fragment endpoint) is located at a point in the plane, and every edge is a C^1 smooth contour joining two vertices. Boundary fragments may or may not be oriented. If they are oriented, then the direction of the edge indicates the sign of the brightness gradient. We adopt the convention that the darker region lies to the right as the boundary fragment is traversed in the direction of its orientation.

G_{input} is augmented to form $G_{\text{nonplanar}} = (V_{\text{endpoints}}, E_{\text{fragments}} \cup E_{\text{completions}})$ by adding edges representing potential completions. As with each element of $E_{\text{fragments}}$, each element of $E_{\text{completions}}$ is a contour joining two elements of $V_{\text{endpoints}}$. Finally, a planar graph, G_{planar}, is created by splitting the edges of $G_{\text{nonplanar}}$ wherever two intersect and creating a vertex at that point called a *crossing*. If $V_{\text{crossings}}$ is the set of crossings, and $E'_{\text{fragments}} \cup E'_{\text{completions}}$ is the set of edges after the splitting operation, then $G_{\text{planar}} = (V_{\text{endpoints}} \cup V_{\text{crossings}}, E'_{\text{fragments}} \cup E'_{\text{completions}})$. Figure 8 illustrates what G_{planar} looks like in the case of the Kanizsa Triangle. Here the endpoints (i.e., $V_{\text{endpoints}}$) are drawn as filled circles and the crossings (i.e., $V_{\text{crossings}}$) as nonfilled circles. The boundary fragments (i.e., $E'_{\text{fragments}}$) are drawn as solid lines and the set of potential completions (i.e., $E'_{\text{completions}}$) as dotted lines.

The problem of maximizing (or minimizing) a linear objective function subject to linear inequality (or equality) contraints is termed a *linear program* (or *LP*). An *integer linear program* (or *ILP*) is a linear program where the solution is further constrained to have integer components. Integer linear programming is a standard and powerful formalism for describing combinatorial optimization problems of all kinds. By writing a fixed number of integer linear constraints for each vertex and edge in G_{planar}, an integer linear program equivalent to the graph labeling problem is generated. The optimal labeled subgraph of G_{planar} is a labeled knot-diagram and is termed G_{knot}. G_{knot} defines the *surface organization*.

4.2. *Topological Validity*

The first constraint enforced is that every edge (whether boundary fragment or completion) has one of two orientations (i.e., $\rightarrow$ and $\leftarrow$) and this orientation represents its *sign of occlusion*. As always, the convention is that the

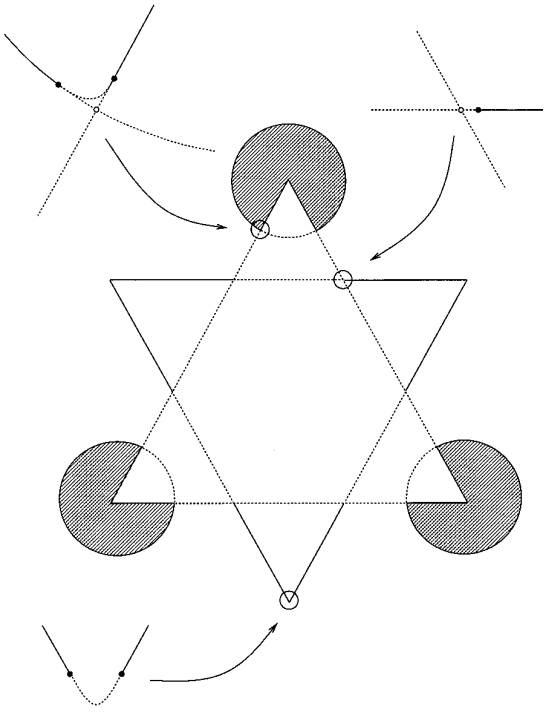

FIG. 8. What $G_{\text{planar}} = (V_{\text{endpoints}} \cup V_{\text{crossings}}, E'_{\text{fragments}} \cup E'_{\text{completions}})$ looks like for the Kanizsa Triangle. The endpoints (i.e., $V_{\text{endpoints}}$) are drawn in filled circles and the crossings, (i.e., $V_{\text{crossings}}$) as non-filled circles. The boundary fragments (i.e., $E'_{\text{fragments}}$) are drawn as solid lines and the set of potential completions (i.e., $E'_{\text{completions}}$) as dotted lines.

surface lies to the right as its boundary is traversed in the direction of its orientation.[4] The two possible orientations of a boundary fragment with respect to one of its endpoints, p, are represented as 0–1 valued integers x_p and x'_p. If p and q are opposite endpoints of a single boundary fragment, then the direction from endpoint q to endpoint p is represented by x_p (i.e., x'_q) and the direction from endpoint p to endpoint q is x'_p (i.e., x_q). Using this representation, the necessary constraint is the following integer linear inequality, enforced at every endpoint, $p \in V_{\text{endpoints}}$:

$$x_p + x'_p \leq 1. \tag{1}$$

Since the image projections of surface boundaries are closed plane-curves, all instantiated edges must form graph cycles in G_{knot}. It follows that there must be a unique completion through each endpoint. Furthermore, the sign of occlusion of the completion must be unique and compatible with the sign of occlusion of the sponsoring boundary fragments. Let *completions*(p) be the potential completions of the boundary fragment through endpoint p (Fig. 9). Two constraints per endpoint guarantee all of the above:

$$x_p = \sum_{j \in \textit{completions}(p)} x_j \tag{2}$$

$$x'_p = \sum_{j \in \textit{completions}(p)} x'_j . \tag{3}$$

Recall that as part of the process of constructing G_{planar}, at every point where one edge crosses another, the edges are split into four new edges and joined by a crossing vertex. Call the four edges u, d, l, and r and the crossing vertex c (Fig. 10). Associated with each of the four edges are 0–1 valued integers x and x' representing their signs of occlusion. Also associated with each edge is a positive integer variable n representing the boundary depth (i.e., the number of surfaces between the edge and the eye or camera). Certain constraints are immediately apparent. First of all, the signs of occlusion of edge u and edge d must be equal. Likewise for edge l and edge r. As simple equality constraints, they can be enforced by substitution and need not actually appear in the linear program: $x_u = x_d$, $x'_u = x'_d$, $x_l = x_r$, and $x'_l = x'_r$.

A second observation is that if u and l (and by implication d and r) are instantiated, then the surface which u bounds (call it $\mathcal{S}_u$) is either above or below the surface which l bounds (call it $\mathcal{S}_l$). This is independent of the specific signs of occlusion of u or l. When one considers that only the sign of occlusion of the uppermost surface has any effect on the relative depths of the four edges (i.e., n_u, n_d, n_l, and n_r), it becomes clear that crossing c can be in one of four principal states. The specific state is determined by which of u, d, l, or r is being occluded by the uppermost surface. When $\mathcal{S}_l$ is above $\mathcal{S}_u$, and the sign of occlusion of l is $r \rightarrow l$ (i.e., $x'_l = 1$), then the crossing is in the *up* state (denoted by $\top$). If edge l's sign of occlusion is $l \rightarrow r$ (i.e., $x_l = 1$) then the crossing is in the *down* state (denoted by $\perp$). When $\mathcal{S}_u$ is above $\mathcal{S}_l$ the crossing is either in the *left* ($\vdash$) state or the *right* ($\dashv$) state, depending on whether the sign of occlusion of u is $u \rightarrow d$ (i.e., $x'_u = 1$) or $d \rightarrow u$ (i.e., $x_u = 1$).

The four states are represented in the linear program with four 0–1 valued variables $x_c^{\top}, x_c^{\perp}, x_c^{\vdash}$, and $x_c^{\dashv}$. Crossing c is in the *up* state exactly when $x_c^{\top} = 1$ and $x_c^{\perp} = x_c^{\vdash} = x_c^{\dashv} = 0$. The other three states are represented similarly. Having established a representation, it is now possible to describe the first constraint enforced at every crossing. It ensures that one of the four crossing states will be true if both u and l are instantiated. For every $c \in V_{\text{crossing}}$ (with adjacent edges, u, d, l, and r) enforce the following:

$$x_u + x'_u + x_l + x'_l \leq x_c^{\top} + x_c^{\perp} + x_c^{\vdash} + x_c^{\dashv} + 1. \tag{4}$$

When u and l are instantiated, the left side of the inequal-

[4] Note that the direction of the sign of occlusion is distinct from the direction of the sign of contrast.

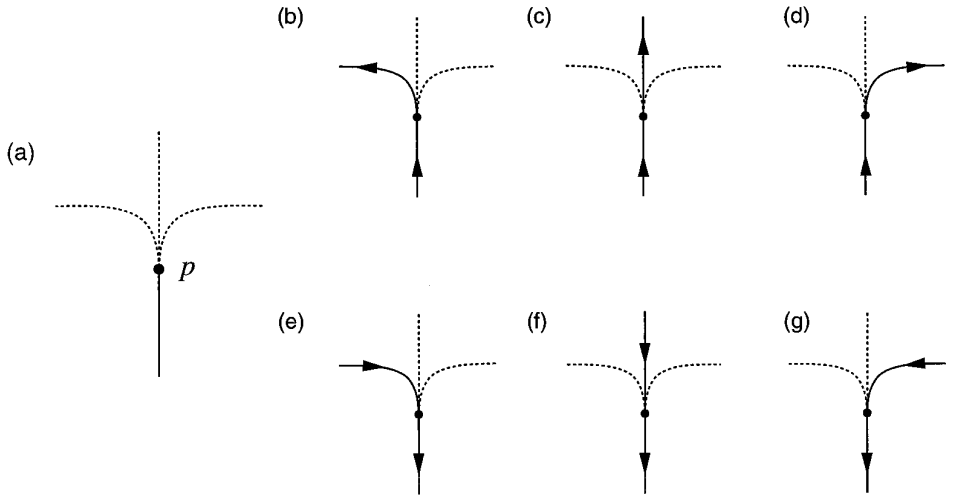

FIG. 9. (a) Potential completions at endpoint p of graph G_{planar}; (b, c, and d) unique continuation in the case where $x_p = 1$; (e, f, and g) unique continuation in the case where $x'_p = 1$.

ity equals two, so that at least one of $x_c^{\top}$, $x_c^{\perp}$, $x_c^{\vdash}$, and $x_c^{\dashv}$ must equal one. Another constraint makes the four states mutually exclusive:

$$x_c^{\top} + x_c^{\perp} + x_c^{\vdash} + x_c^{\dashv} \leq 1. \tag{5}$$

The specific signs of occlusion which are preconditions for each of the four states appear on the right sides of the inequalities which follow:

$$x_c^{\top} \leq x'_l \tag{6}$$

$$x_c^{\perp} \leq x_l \tag{7}$$

$$x_c^{\vdash} \leq x'_u \tag{8}$$

$$x_c^{\dashv} \leq x_u. \tag{9}$$

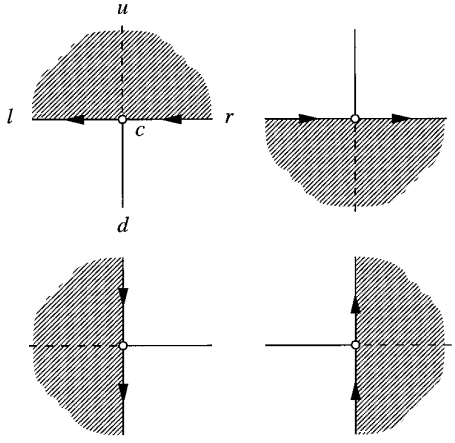

FIG. 10. Four principal crossing states.

For example, crossing c can only be in the *left* state (i.e., $x_c^{\vdash} = 1$) when edge u's sign of occlusion is $u \rightarrow d$ (i.e., $x'_u = 1$).

It is important to note that the four principal crossing states stand for specific differences in relative depth across the crossing vertex. The following constraints define the crossing states as relative depths:

$$n_u - n_d = x_c^{\top} - x_c^{\perp} \tag{10}$$

$$n_l - n_r = x_c^{\vdash} - x_c^{\dashv}. \tag{11}$$

4.3. *Stimulus Conformity*

Topological validity is a necessary but not a sufficient condition for feasiblity of G_{knot}. For a solution to be feasible, it must also conform to the image evidence. Most importantly, since input boundary fragments are visible in the image (i.e., they correspond to luminance boundaries), it is necessary that their depth indices in G_{knot} equal zero (i.e., for every visible boundary fragment $f \in E'_{\text{fragments}}$, we require that $n_f = 0$).

Also important, if a completion is instantiated, and its boundary depth equals zero (indicating that it should be visible) then the absence of a corresponding luminance boundary should be explainable. In Rock's [25] words, "the solution must not contradict the stimulus." Following Rock, we hypothesize that illusory contours occur only in those situations where the missing section of surface boundary presumably projects to the image plane with little or no brightness change.

A "pacman" from the Kanizsa triangle can be used to illustrate the visibility constraints which are written in the three possible cases (see Fig. 11). First, because contour i is a visible boundary fragment, its depth is constrained to

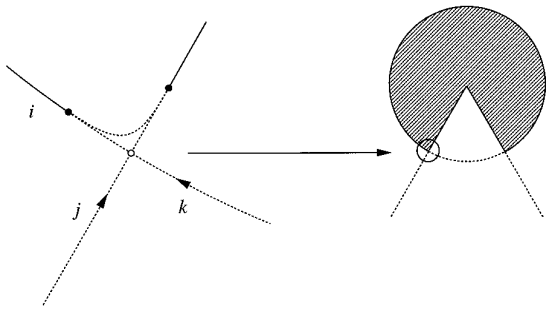

FIG. 11. A ''pacman'' from the Kanizsa triangle is used to illustrate the visibility constraints which are written in the three possible cases. First, because contour i is a visible boundary fragment, its depth is constrained to be zero (i.e., $n_i = 0$). Second, because contour j (with the sign of occlusion indicated) represents the completion of a white surface against a white background, no constraint is placed on its depth. Finally, contour k (again with the sign of occlusion indicated) represents the potential completion of the disc. Since the brightnesses are consistent with a dark surface against a light background, its depth is constrained to be greater than zero (i.e., $x_k \leq n_k$).

be zero (i.e., $n_i = 0$). Second, because contour j (with the sign of occlusion indicated) represents the completion of a white surface against a white background, no constraint is placed on its depth. Finally, contour k (again with the sign of occlusion indicated) represents the potential completion of the disc. Since the brightnesses are consistent with a dark surface against a light background, its depth is constrained to be greater than zero (i.e., $x_k \leq n_k$).

4.4. *Preference*

By committing to a set of potential completions of fixed shape, the surface organization problem was reduced to a graph labeling problem. However, the integer linear constraints defining the labeling problem only guarantee that the surface organization is topologically valid and conforms to the image evidence. Each of the integer points within the feasible region of the integer linear program can be viewed as a ''prediction'' about the actual state of the world. Usually, there will be more than one such integer point. If the goal of the computation is to choose the organization which is most likely to be correct (i.e., *veridicality*), then it is important that the predictions be equally specific. In our analysis, 0–1 valued expressions are interpreted as *events*. If two points in the feasible region are to be compared, then each must be interpreted as a prediction over the same set of events (i.e., 0–1 valued expressions). Following Lowe [19], we propose that the prediction consist of an assignment to each element of the set of potential completions, the label ''accidental'' or ''nonaccidental.'' Since the set of potential completions is of fixed size, these predictions are equally specific.

4.4.1. *Unit Preference.* In any given surface organization, some subset of completions is instantiated while a complementary set remains uninstantiated. One of the motivations for using curves of least energy to represent potential completions is a presumed relationship between the shape of the curve of least energy and the likelihood that two contour fragments have a common nonaccidental origin (i.e., the likelihood that they form consecutive segments of a single boundary). It is assumed that shape and completion likelihood are related through an unspecified probability density function.[5]

The instantiation of a completion is, in effect, an assertion that its origin is *nonaccidental*. Since a completion is instantiated if and only if $x_i + x_i' = 1$, the probability that completion i's origin is nonaccidental can be written as $\Pr(x_i + x_i' = 1)$. Although we will see very shortly that the converse does not hold, let us assume (for the moment) that the converse is also true. That is, failure to instantiate a completion is tantamount to an assertion that its origin is *accidental*. The probability that completion i's origin is accidental then becomes $1 - \Pr(x_i + x_i' = 1)$. Finally, let us also assume that whether a completion's origin is accidental or nonaccidental is independent of whether any other completion's origin is accidental or nonaccidental. That is, for any two completions, i and j:

$$\Pr(x_i + x_i' = 1 \mid x_j + x_j' = 1) = \Pr(x_i + x_i' = 1).$$

Because every completion either is or is not instantiated, the likelihood of a given surface organization (assuming that all completion's origins are independent) becomes

$$P_{\text{unit}} = \prod_{i \in E_{\text{completions}}} \Pr(x_i + x_i' = 1)^{(x_i + x_i')} (1 - \Pr(x_i + x_i' = 1))^{(1 - x_i - x_i')}.$$

This function is, in turn, a monotonically increasing function of the sum of the logarithms of the likelihoods, so that maximizing the sum of the logarithms is equivalent to maximizing the original function. The sum of the log-likelihoods has the advantage of being linear in the unknowns, which is essential if the objective function is to be part of a linear program. Since any completion, i, is instantiated if and only if $x_i + x_i' = 1$, the logarithm of the likelihood that a specific subset of completions will be instantiated is

[5] For the purpose of the experimental implementation, specific shape features and probability densities were chosen. These are described in the next section.

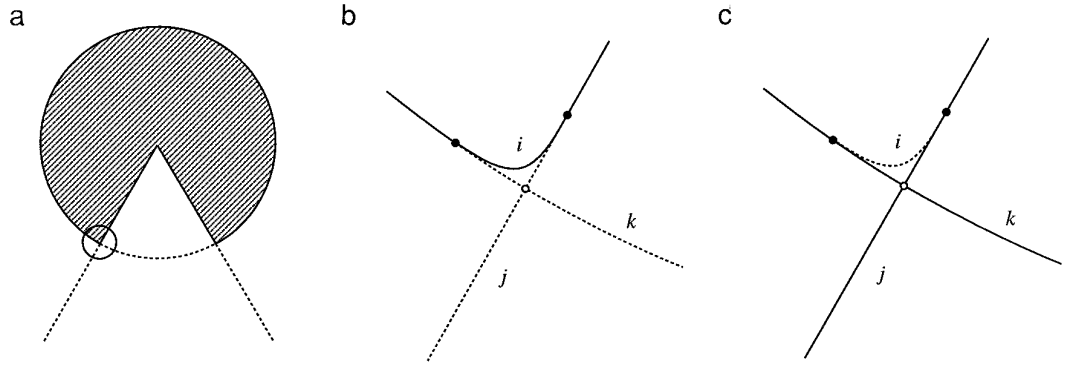

FIG. 12. (a) One corner of a "pacman." (b) Rock's "literal solution." The spatial coincidence represented by completions j and k is attributed to chance, so that the degree of surprise (i.e., S_{literal}) is $-\ln(1 - \Pr(x_j + x'_j = 1)) - \ln(1 - \Pr(x_k + x'_k = 1))$. (c) The "illusory contour solution" offers a single explanation for all of the completions, even though the corner is not instantiated. The degree of surprise (i.e., S_{illusory}) is therefore zero.

$$Q_{\text{unit}} = \sum_{i \in E_{\text{completions}}} \ln(\Pr(x_i + x'_i = 1))(x_i + x'_i) + \ln(1 - \Pr(x_i + x'_i = 1))(1 - x_i - x'_i).$$

Unfortunately, this analysis does not adequately account for the phenomenon of illusory contours. Let us again consider the example from Fig. 6, this time for the purpose of determining the different degrees of surprise implied by the two unit organizations. Figures 12b and 12c show magnified views of one neighborhood (i.e., the area within the circle in Fig. 12a) in the two unit organizations. In the "literal" solution (i.e., Fig. 12b) the only completion instantiated is the "corner," i. Since the other two completions (i.e., j and k) are not instantiated, the spatial coincidence they represent is attributed to chance, and the "degree of surprise," S_{literal} is equal to

$$S_{\text{literal}} = -\ln(1 - \Pr(x_j + x'_j = 1)) - \ln(1 - \Pr(x_k + x'_k = 1)).$$

Now compare this to the "illusory contour solution," where only completions j and k are instantiated (Fig. 12c). If we assume that because completion i is not instantiated that the spatial coincidence it represents is attributed to chance,[6] then the degree of surprise equals $-\ln(1 - \Pr(x_i + x'_i = 1))$. However, in the illusory contour solution, this assumption does not hold because all three completions originate in the same physical process: the occlusion of one surface by another. Since the corner is an artifact of occlusion, the illusory contour solution provides a single explanation for the existence of each of the potential completions, even though the corner is not instantiated. Since no spatial coincidence is attributed to chance, the degree of surprise, S_{illusory}, is actually zero.

Obviously, the first term of objective function Q_{unit} does not evaluate to S_{literal} and S_{illusory} as required. However, if all corners which are not orientation discontinuities are assumed to be artifacts of occlusion, then the desired effect can be achieved by assigning corners zero weight in the objective function:

$$Q'_{\text{unit}} = \sum_{i \in E_{\text{completions}} - E_{\text{corners}}} \ln(\Pr(x_i + x'_i = 1))(x_i + x'_i) + \ln(1 - \Pr(x_i + x'_i = 1))(1 - x_i - x'_i).$$

We hypothesize that where contrast stimuli are involved, there is an assumption that corners occur for two reasons, neither of which is accidental. The probability that corner i's origin is accidental is therefore zero, not $\Pr(1 - x_i - x'_i)$.

4.4.2. *Depth Preference.* Neither Q_{unit} nor Q'_{unit} incorporate all of the factors which contribute to preference in human vision. Both partition the feasible region into equivalent classes such that two integer points are in the same equivalence class if and only if they contain the same set of potential completions (i.e., they represent the same unit organization). As demonstrated in Fig. 7, a given unit organization will usually admit many different depth labelings, so that the majority of these equivalence classes will contain many distinct (although equally "likely") integer points. If the goal is to compute a unique organization, then additional preference criteria must be employed.

Additional preference criteria are described by Rock [25], who demonstrates the role they play with numerous examples. First among these is the correlation between the sign of occlusion and the sign of contrast. The effect of this preference in human vision is that against a white background, there is a strong tendency to perceive black as

[6] Recall that this was the first of two assumptions which led to the objective function. The second was the independent assumption.

figure. Similarly, the sign of occlusion is strongly correlated with the sign of curvature. Consequently, in human vision, figure tends to be perceived as convex and ground as concave. Finally, there is a strong tendency in human vision to perceive the space between closely spaced parallel lines as figure. This also holds for the space between pairs of contours exhibiting bilateral symmetry, although this tendency is weaker than the others.

At first glance, it appears to be a simple matter to incorporate the additional figure–ground preferences into a new objective function. Clearly, the variables required to form 0–1 valued expressions representing the competing depth labelings exist in the integer linear program (i.e., $x_i = 1$ and $x_i' = 1$). Unfortunately, matters are not so simple. The problem of determining the relative likelihood of a contour's two possible signs of occlusion as a function of contrast is presumably straightforward. However, it is difficult to design an objective function which differentially weights the two signs of occlusion without introducing a bias for (or against) organizations with larger numbers of completions. Solving this problem is a subject for future work.

5. EXPERIMENTAL SYSTEM

To simplify implementation of the experimental system, "off-the-shelf" components were used wherever possible. For example, the straight-line grouping algorithm developed by Boldt [3] was used to generate the input set of boundary fragments. Consequently, the experimental system was tested with straight-sided figures. This limitation is not as significant as it may seem, since it proved to be a simple matter to design straight-sided equivalents of some of the better known figures from the illusory contour literature. Four of these are shown in Fig. 13.

5.1. *Minimum Energy Cubic Bezier Splines*

For simplicity's sake, cubic Bezier splines of minimum energy were used to represent completion shape, not true curves of least energy. Consequently, only tangent continuity (not tangent and curvature continuity) was enforced at fragment endpoints. This is a reasonable simplification, since the total bending energy in the minimum energy cubic Bezier spline still provides strong evidence of a mutual nonaccidental origin.

Let p and p' be the ends of the two boundary fragments which are to be joined and $\hat{t}$ and $\hat{t}'$ be vectors tangent to the fragments at those points (see Fig. 14). A cubic spline which smoothly passes through both points with orientations matching those of the fragments is easily constructed in the Bezier form by specifying the locations of four control points. While the first and fourth control points are located at the ends of the fragments (i.e., $b_0 = p$ and $b_3 = p'$), the locations of b_1 and b_2 are underdetermined. In general, there is a two-parameter family of smooth cubic splines passing through two points with specified tangents. The free parameters are the distance between b_1 and b_0 in the direction of $\hat{t}$ (i.e., $d = |b_1 - b_0|$) and the distance between b_3 and b_2 in the direction of $\hat{t}'$ (i.e., $d' = |b_3 - b_2|$). The total bending energy, $E = \int k(s)^2\, ds$, is readily computed for a particular Bezier control polygon by Simpson's method, and the d and d' minimizing this quantity can be computed through a multivariate minimization technique, such as the downhill simplex method described in [24].

5.2. *Completion Features and Categories*

The probability density function defining the distribution of completion shape features is at least bimodal, since gaps originate in at least two distinct physical processes (i.e., gaps due to occlusion and gaps due to orientation discontinuity). If straight-sided surfaces occur with sufficient frequency, then it is likely that the probability density function describing the distribution of completion shape features is even more complex. Consider the joint conditional probability densities of completion shape features given gaps caused by:

- Orientation discontinuity
- Partial occlusion of a straight boundary
- Generic occlusion

Although the likelihood that a gap between two fragments is nonaccidental and originates in occlusion rapidly increases as distance between their near endpoints decreases, a gap having nearly zero length is much more likely to be caused by an orientation discontinuity. Similarly, although likelihood increases as energy in the interpolating curve decreases, the existence of a smooth interpolating curve with nearly zero total bending energy has special significance when straight-sided surfaces are allowed.

For these reasons,[7] it proved useful to divide the set of completions into three categories and employ a multicategory Bayes classifier [5] to select among them. The category minimizing squared Mahalanobis distance (i.e., covariance-weighted Euclidean distance) between observed and mean shape feature values (with zero mean) is used as a heuristic measure of likelihood,

$$\Pr(\textit{completion}|\mathbf{x}) \propto \exp(-\min(\mathbf{x}^t C_{\text{corner}}^{-1}\mathbf{x}, \mathbf{x}^t C_{\text{straight}}^{-1}\mathbf{x}, \mathbf{x}^t C_{\text{generic}}^{-1}\mathbf{x})), \quad (12)$$

where the components of $\mathbf{x}$ are

[7] Together with the need to distinguish corners from other completions, as required by objective function Q_{unit}'.

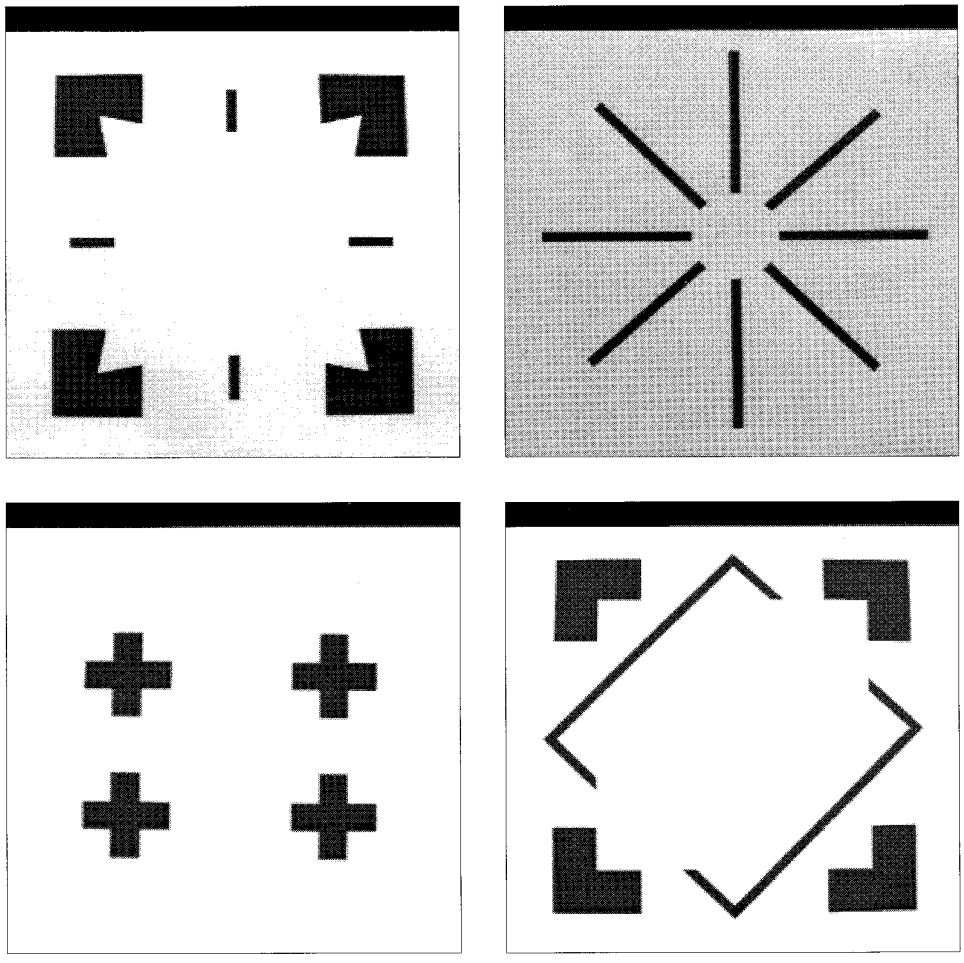

FIG. 13. Four test figures from the illusory contour literature. Upper left: *Warped Square.* Upper right: *Ehrenstein.* Lower left: *Kanizsa Plusses.* Lower right: *Woven Square.*

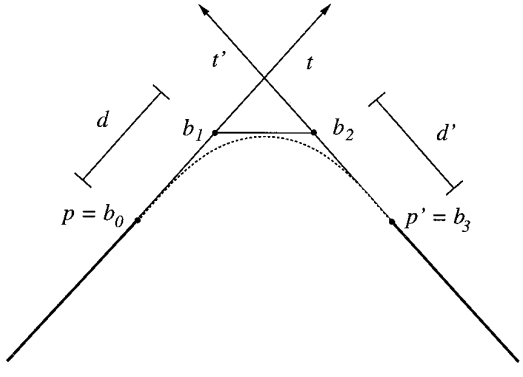

FIG. 14. A cubic Bezier spline. In general, there is a two-parameter family of smooth cubic splines passing through any two points p and p' with specified tangents. The free parameters are the distance between b_1 and b_0 in the direction of $\hat{t}$ (i.e., $d = |b_1 - b_0|$) and the distance between b_3 and b_2 in the direction of $\hat{t}'$ (i.e., $d' = |b_3 - b_2|$).

$$x_1 = |b_0 - b_3| \int \kappa(s)^2 \, ds \tag{13}$$

$$x_2 = \int ds \tag{14}$$

$$x_3 = \frac{l + \int ds + l'}{\int ds} \tag{15}$$

$$x_4 = |\theta - \theta'| \tag{16}$$

$$x_5 = \begin{cases} 0 & \text{if } \kappa(0)\kappa(\int ds) \geq 0 \\ 1 & \text{otherwise.} \end{cases} \tag{17}$$

Here x_1 is the total bending energy normalized by the distance between the fragment endpoints, x_2 is total arc length, x_3 is gap size relative to the size of the boundary fragments, x_4 is change in orientation, and x_5 is the presence of an inflection point. The matrices defining the feature probability densities for each category are

$$C_{\text{corner}} = \begin{bmatrix} 1.0 \times 10^9 & 0 & 0 & 0 & 0 \\ 0 & 5.0 & 0 & 0 & 0 \\ 0 & 0 & 0.05 & 0 & 0 \\ 0 & 0 & 0 & 1.0 \times 10^9 & 0 \\ 0 & 0 & 0 & 0 & 1.0 \times 10^{-9} \end{bmatrix} \quad (18)$$

$$C_{\text{straight}} = \begin{bmatrix} 0.001 & 0 & 0 & 0 & 0 \\ 0 & 2500.0 & 0 & 0 & 0 \\ 0 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0.001 & 0 \\ 0 & 0 & 0 & 0 & 1.0 \times 10^9 \end{bmatrix} \quad (19)$$

$$C_{\text{generic}} = \begin{bmatrix} 5.0 & 0 & 0 & 0 & 0 \\ 0 & 1000.0 & 0 & 0 & 0 \\ 0 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & \frac{\pi}{5} & 0 \\ 0 & 0 & 0 & 0 & 1.0 \times 10^{-9} \end{bmatrix}. \quad (20)$$

In the heuristic likelihood function, C_{corner}, C_{straight}, and C_{generic} act like covariance matrices since each is positive definite and is used to define a multivariate density resembling a Gaussian. Although these ''densities'' are not normalized, the heuristic likelihood function returns values between 0 and 1 and these values are used as probabilities.

The specific values which comprise C_{corner}, C_{straight}, and C_{generic} were chosen partly by design and partly by trial and error. For example, since gaps caused by orientation discontinuity are very short, length (i.e., x_2) must be nearly zero if a completion is to be classified as a ''corner.'' The values of energy and relative orientation (i.e., x_1 and x_4), on the other hand, are completely irrelevant. Accordingly, the diagonal entry for the length feature in C_{corner} was set to the value of 0.5 (pixels) and the entries for energy and relative orientation were set to 1.0×10^9. This illustrates that certain entries function more like ''switches'' than covariances. Because the diagonal entries for energy and relative orientation are so large, these values do not affect the ''likelihood'' of a completion classified as a corner.

Similarly, the two features which together characterize the ''straight'' category are energy and relative orientation (i.e., x_1 and x_4). A completion joining two collinear lines should have nearly zero energy and its orientation should remain constant. However, whether or not an essentially straight completion contains an inflection point (i.e., x_5) is immaterial. Accordingly, the diagonal entries for energy and relative orientation in C_{straight} were both set to 0.001 while the diagonal entry for inflection was set to 1.0×10^9 (i.e., ''don't care'').

Finally, the ''generic'' category functions as a ''catch all.'' The only shape feature which is of special importance is the presence or absence of an inflection point. To control the number of potential completions, the diagonal entry for the inflection feature in C_{generic} was set to 1.0×10^{-9}. The result is that inflection points are not allowed. The values of the other diagonal entries were determined by a process of trial and error.

5.3. *Building the Graphs*

The straight line segments which serve as input to the experimental system are produced by Boldt's zero crossing grouping algorithm [3]. In general, if there are n boundary fragments with $2n$ endpoints, there will be $O(n^2)$ potential completions and $O(n^4)$ crossing vertices in G_{planar}. Since there are a constant number of variables and constraints per crossing vertex, the constraint matrix for the integer linear program will be of size $O(n^8)$. Since the test figures typically contain about 50 boundary fragments, the space requirements alone of the ''naive'' formulation are unacceptable.

Fortunately, there are several good strategies for limiting the number of potential completions which are explicitly represented by edges in $G_{\text{nonplanar}}$. Since a factor of two reduction in the number of edges in $G_{\text{nonplanar}}$ can result in as much as a factor of four reduction in the number of crossings in G_{planar}, this is the obvious point to apply pruning strategies. These strategies are listed below in roughly their order of usefulness:

- Likelihood threshold.
- k-most likely at each endpoint.
- Contrast sign constraint.
- Overlap pruning.

The G_{planar} computed for the four test figures are shown in Fig. 15. An integer linear program (ILP) is generated by writing a fixed set of integer linear inequalities, as described in the last section, for each of the vertices and edges of G_{planar}. The ILP is solved by the method of *branch and bound* search [22, 26]. When an optimal feasible solution exists, it is interpreted as a labeling of G_{planar}, which defines G_{knot} (i.e., a labeled knot-diagram). The labeled knot-diagrams computed for the four test figures are shown in Fig. 16. These organizations maximize the following objective function:

$$Q'_{\text{surface}} = Q'_{\text{unit}} + Q_{\text{depth}}. \quad (21)$$

Although the Q'_{unit} term has already been defined, without the second term, the ILP is underconstrained. Accordingly, for the purposes of the experimental implementa-

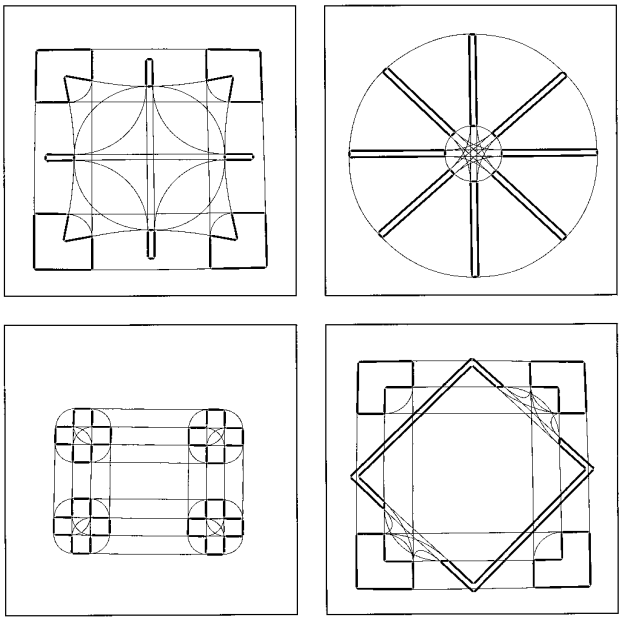

FIG. 15. G_{planar} for the four test figures. This graph represents the k-most-likely completions ($k = 3$) for every boundary fragment as cubic Bezier splines of least energy. Boundary fragments are drawn thick and potential completions are drawn thin. Graph vertices (endpoints and crossings) are omitted for clarity's sake.

tion, it was necessary to define Q_{depth} (even though the issue of the bias for larger organizations has not been satisfactorily resolved). If we adopt the convention that x_i represents the sign of occlusion with orientation matching the sign of contrast, then a preference for black figure and white ground can be implemented by assigning a positive unit weight to x_i for $i \in E_{\text{fragments}}$:

$$Q_{\text{depth}} = \sum_{i \in E_{\text{fragments}}} x_i. \tag{22}$$

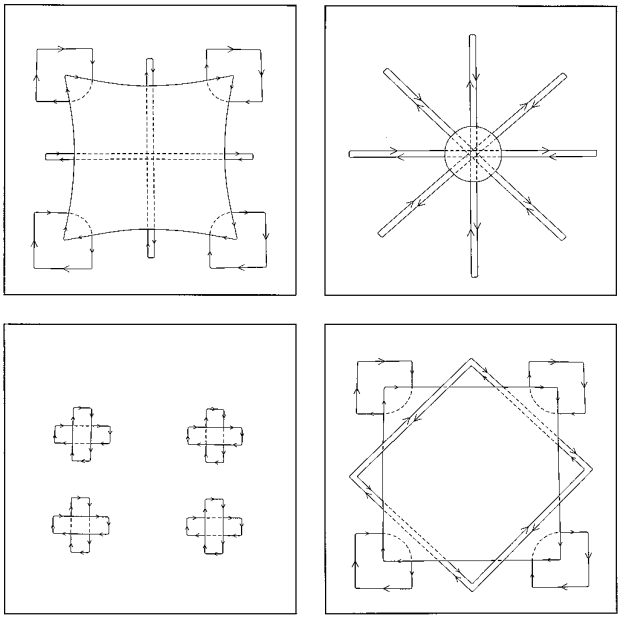

FIG. 16. G_{knot} for the four test figures. Note that the completed surfaces in the *Woven Square* test figure (lower right) can not be embedded in planes at constant depth.

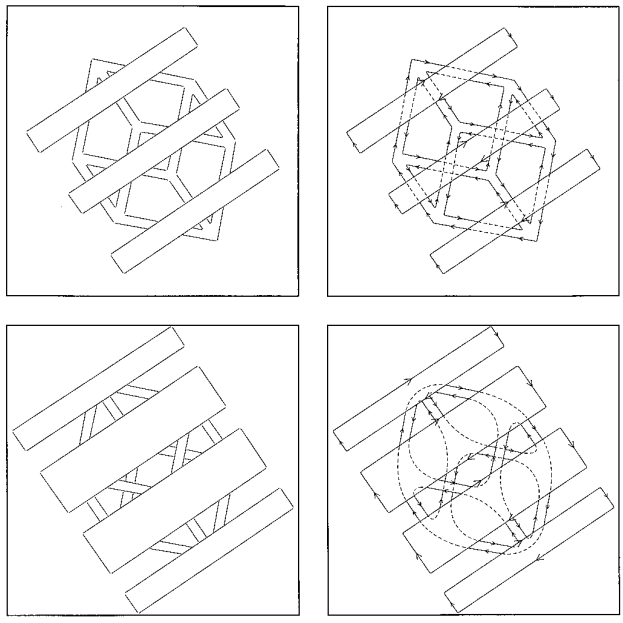

FIG. 17. A simple modification (described in [29]) allowed the experimental system to be demonstrated on Kanizsa's drawings of partially occluded cubes (note: in the original figures, the occluding bars are not closed). Upper left: The *Y's with Bars* test figure. Upper right: G_{knot} for the *Y's with Bars* test figure. The completed surface can not be embedded in planes at constant depth. Lower left: The *X's with Bars* test figure. Lower right: G_{knot} for the *X's with Bars* test figure. It is especially noteworthy that the experimental system has correctly organized the *X's with Bars* test figure because perceptual completion of this figure is challenging even for human vision. This result demonstrates the potential disambiguating power of topological constraints.

At least superficially, the organizations produced by the experimental system resemble what humans perceive. The appropriate illusory surfaces are constructed for the *Warped Square, Ehrenstein,* and *Woven Square* test figures. In the case of the *Woven Square,* the computed illusory surface passes over and under the diamond-shaped frame, so as to conform with the "proximal stimulus." This is in agreement with the human percept. Finally, like the human visual system, the experimental system does not construct an illusory rectangle in the case of the *Kanizsa Plusses* figure. Additional experimental results are shown in Fig. 17.

6. A REVISED PROBLEM-LEVEL FORMULATION

Figural completion was portrayed earlier as the problem of computing a labeled knot-diagram representing an ante-

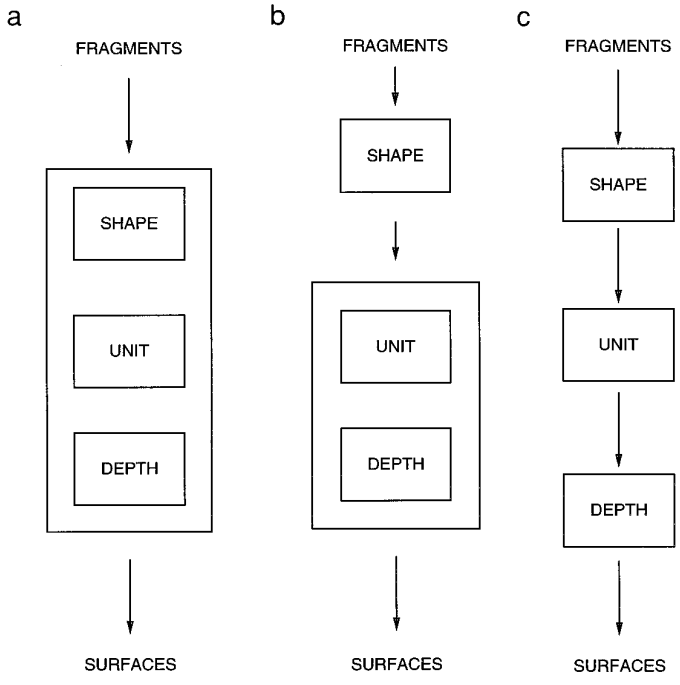

FIG. 18. Alternate problem decompositions. Figural completion requires three different sources of ambiguity to be overcome (i.e., shape, unit, and depth). In theory, all three can be resolved concurrently, which would result in the problem decomposition shown in (a). However, it was proposed earlier that the shape of a perceptual completion is independent of the role it plays in the organization. This permitted the decomposition shown in (b), which we refer to as the *surface organization model*. In this section, a further decomposition is proposed. The *unit/depth organization model,* shown in (c), requires the additional assumption that unit ambiguity can be resolved in advance and independently of depth ambiguity.

rior scene from a set of contour fragments representing image luminance boundaries. Given this computational goal and this input, three distinct sources of ambiguity were identified. These were termed *shape, unit,* and *depth*. Since the applicable physical constraints were insufficient to overcome these ambiguities, an additional assumption was introduced. Broadly speaking, this assumption was: *the shape of a perceptual completion is independent of the role it plays in the organization*. This assumption allowed the figural completion problem to be decomposed into two independent subproblems, the first devoted to shape, the second devoted jointly to unit and depth. Decomposed in this way, it proved possible to formulate the second subproblem as an integer linear program, which led to a unique solution.

In this section, a further decomposition is proposed. Again, this decomposition occurs along the lines of the inherent ambiguities (see Fig. 18). Specifically, it is proposed that *unit ambiguity is resolved in advance and independently of depth ambiguity*. The integer linear program for the *surface organization model* is replaced by two simpler integer linear programs in the *unit/depth organization model*. The results of a revised experimental implementation are compared to those of the original.

The motivation for this new decomposition is twofold. First, evidence from human vision supports the hypothesis that unit organization occurs in advance and independently of depth organizations (see Fig. 19). Second, the integer linear program for the unit organization subproblem can be shown to possess a property called *total unimodularity,* which allows it to be solved by numerical relaxation in a locally connected network (see [29]).

6.1. *Evidence from Human Vision*

Earlier, the figural completion problem was reduced to a graph labeling problem and posed as an integer linear program. This was made possible by assuming that completion shape is determined solely by the tangents and curvatures of the terminal ends of the occluded boundaries and not by a completion's role in any eventual surface organization. A still more radical notion is proposed by Kellman and Loukides [17], who argue that unit organization is accomplished in advance and independently of depth organization. More specifically, they propose that visible con-

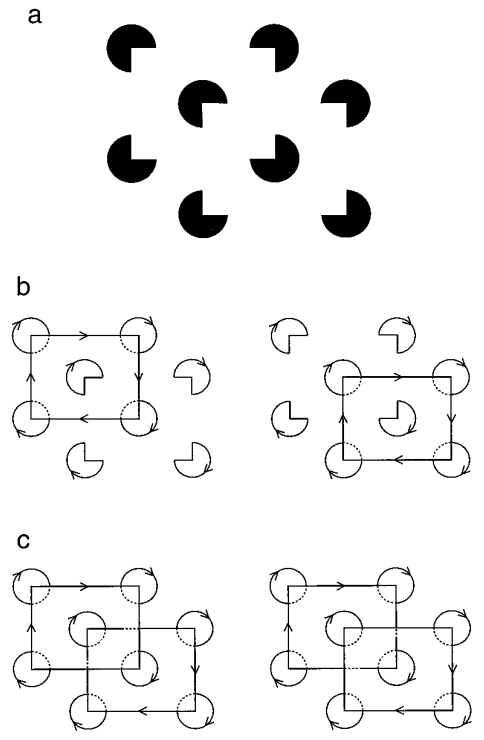

FIG. 19. Predictions of different models. (a) A variation of a display by Fahle and Palm [6]. Although only one illusory rectangle is physically possible, the arrangement of the inducing elements and the overall symmetry of the figure seems to suggest that two are present. (b) Bistable interpretation predicted by surface organization model. (c) Bistable interpretation predicted by unit/depth organization model.

tour fragments are organized into closed plane curves before their relative depths at points of intersection are determined.

If the goal of figural completion is to compute a topologically valid surface organization, then there is a fundamental theoretical problem with Kellman and Loukides' proposal. Unless unit and depth organization are accomplished together, there is no guarantee that the set of closed plane curves produced by the unit organization process will have a consistent depth labeling. Yet, recent evidence seems to suggest that the human visual system has not heard of this theory.

Consider the stimulus depicted in Fig. 19a, which is a variation of a figure designed by Fahle and Palm [6]. Although only one illusory rectangle is physically possible, the arrangement of the inducing elements and the overall symmetry of the figure seems to suggest that two are present. Since only one illusory rectangle is possible, the surface organization model predicts that the two organizations which appear in Fig. 19b will be perceived with equal likelihood. Each of these is a distinct unit organization, containing a different set of boundary components. Both are topologically valid and consistent with the image evidence.

However, an informal study of the visual systems of the author's colleagues suggests that nothing like Fig. 19b is perceived. In fact, most subjects experience something more closely resembling one of the two organizations depicted in Fig. 19c. Here two illusory rectangles exist in some degree of perceptual tension. Sometimes one is on top, sometimes the other. Some observers describe the illusory rectangles as intersecting one another. There are two significant conclusions which can be drawn about this. First, irrespective of which illusory rectangle is on top, the unit organization consists of the same set of boundary components. Second, neither organization is topologically valid. It is as if the visual system commits to a unit organization which subsequently can not be consistently labeled.

6.2. *The Unit/Depth Organization Model*

The unit/depth organization model is defined by two independent and individually simpler graph labeling problems. Like the ILP for the surface organization model (i.e., ILP_{surface}), these are also formulated as integer linear programs, which we will refer to as ILP_{unit} and ILP_{depth}. While ILP_{surface} is a labeling problem on G_{planar}, the unit organization problem in isolation, ILP_{unit}, is a labeling problem on its precursor, $G_{\text{nonplanar}}$. The overall structure of the unit/depth organization model is compared to the surface organization model in Fig. 20. If x_j equals one when completion $j \in completions(p)$ is instantiated, and x_j equals zero when completion j is not instantiated, then ILP_{unit} is formed by generating one constraint per endpoint p of the form

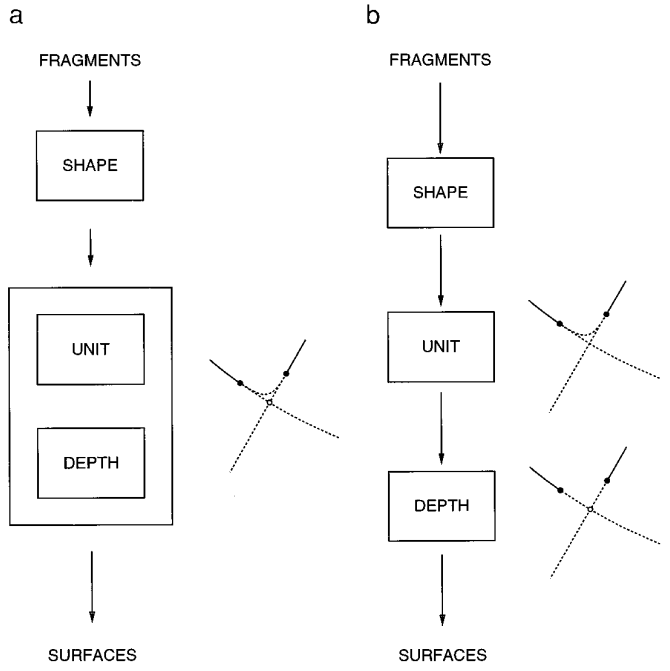

FIG. 20. A comparison of the labeling problems. (a) The integer linear program for the surface organization model (i.e., ILP_{surface}) is a labeling problem on G_{planar}. Unit and depth organization constrain one another, resulting in true surface organization. (b) In the unit/depth organization model, unit organization is accomplished in advance and independently of depth organization. ILP_{unit} is not a labeling problem on G_{planar} but on its precursor, $G_{\text{nonplanar}}$. Completions not instantiated in the unit organization are deleted. A much simpler G_{planar} is then labeled by ILP_{depth}.

$$x_p = \sum_{j \in completions(p)} x_j .$$

The objective function for ILP_{unit} and the optimal feasible solution defines the unit organization. Any completion not instantiated in the unit organization (i.e., any completion j, such that $x_j = 0$) is deleted from $G_{\text{nonplanar}}$. The structure of $G_{\text{nonplanar}}$ after pruning is very simple. Since the in-degree and out-degree of every endpoint vertex is equal to one, the connected components of $G_{\text{nonplanar}}$ are all simple graph cycles. If the unit organization has a consistent depth labeling (which is not guaranteed) then these cycles represent the boundary components of an anterior scene.

The unit organizations computed for the four contrast test figures are shown in Fig. 21. These are the optimal feasible solutions of ILP_{unit}. In three of the four cases, the unit organization can be consistently labeled; these act as the precursors of topologically valid anterior scenes. However, in the case of the *Kanizsa Plusses* test figure no consistent labeling is possible. This suggests that either the unit/depth organization model is incorrect or that the

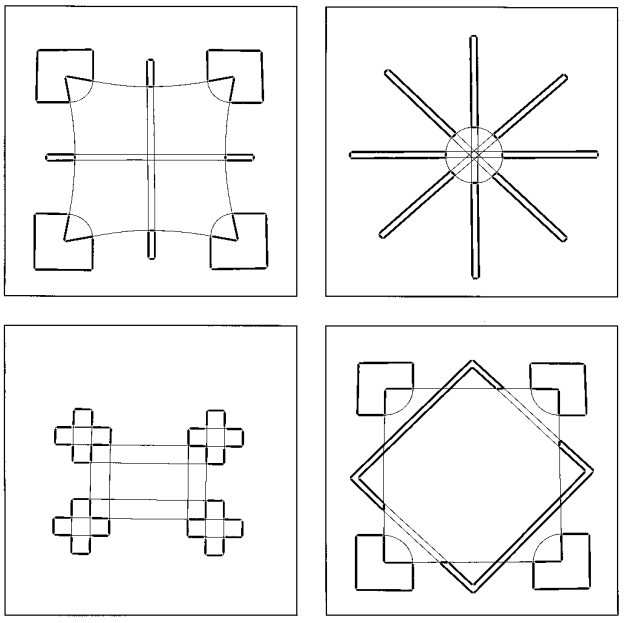

FIG. 21. Unit organizations for the contrast test figures. In three of the four cases, the unit organization can be consistently labeled; these act as the precursors of topologically valid anterior scenes. However, in the case of the *Kanizsa Plusses* test figure (lower left), no consistent labeling is possible.

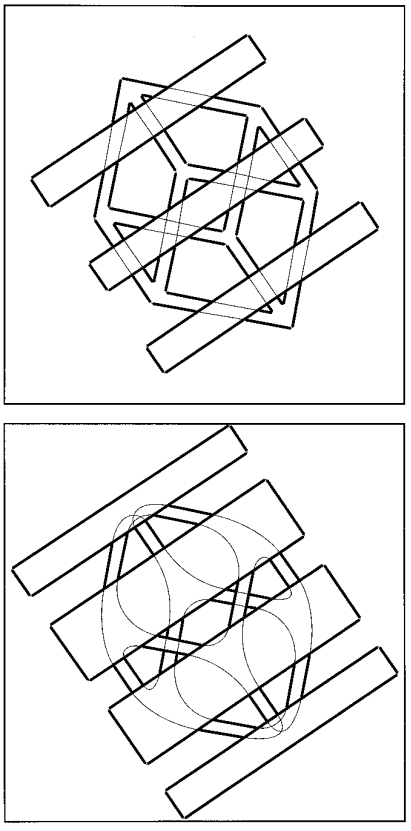

FIG. 22. Unit organizations for the outline test figures. In the case of the *Y's with Bars* test figure (top), the unit organization is the same as for the surface organization model. However, in the case of the *X's with Bars* test figure (bottom), the unit organization admits no consistent depth labeling.

objective function, as defined, does not incorporate all of the preference factors which operate in human vision.

Figure 22 shows the unit organizations computed for the two outline test figures. In the case of the *Y's with Bars* test figure, the unit organization is the same as for the surface organization model. However, in the case of the *X's with Bars* test figure, the unit organization admits no consistent depth labeling. This shows that in the experimental implementation of the surface organization model, the topological constraints can play a role in unit formation.

Like the ILP for the surface organization model, ILP_{depth} is a labeling problem on G_{planar}. As before, G_{planar} is created by splitting the edges of $G_{\text{nonplanar}}$ wherever two intersect and creating a crossing at that point. But ILP_{depth} is considerably simpler than ILP_{surface}, since it exploits the fact that every edge in G_{planar} must appear in G_{knot}.

Recall that a major problem with the surface organization model was the theoretical difficulty in combining unit and depth preference factors in a single objective function. Ideally, the objective function should interpret points within the feasible region as equally specific predictions about the actual state of the world. Unfortunately, as discussed earlier, the relative likelihoods of different figure-ground assignments can not be compared when the unit organizations are of different sizes. One advantage of dividing ILP_{surface} into ILP_{unit} and ILP_{depth} is that this theoretical objection disappears. The objective function for ILP_{depth} needs only to differentiate among alternative figure-ground assignments for a unit organization of fixed size.

The computational advantage of decomposing the figural completion problem according to the unit/depth organization model are convincingly demonstrated in Tables 1 and 2. Table 1 compares the total number of simplex pivot steps for ILP_{surface} to the sum of the totals for ILP_{unit} and

TABLE 1
Surface vs Unit/Depth Organization Models (Total Pivot Steps)

Figure	Surface	Unit/depth	Percent cost	Percent savings
Warped Square	1124	162	14.4	85.6
Ehrenstein	1209	505	41.8	58.2
Kanizsa Plusses	1411	178	12.6	87.4
Woven Square	1263	170	13.5	86.5
Y's with Bars	13377	504	3.8	96.2
X's with Bars	5339	188	3.5	96.5

TABLE 2
Surface vs Unit/Depth Organization Models (Total Multiplies)

Figure	Surface	Unit/depth	Percent cost	Percent savings
Warped Square	7.34×10^8	1.35×10^7	1.8	98.2
Ehrenstein	1.03×10^9	8.47×10^7	8.3	91.7
Kanizsa Plusses	1.49×10^9	2.55×10^7	1.7	98.3
Woven Square	1.21×10^9	1.63×10^7	1.4	98.6
Y's with Bars	1.40×10^{11}	1.20×10^8	0.1	99.9
X's with Bars	6.35×10^9	2.94×10^7	0.5	99.5

ILP_{depth}. Table 2 similarly compares the total number of floating point multiplies. The average reduction in the total number of simplex pivot steps for the five (out of a total of seven) instances where the results of two models could be compared is 83%. The average reduction in the total number of floating point multiplies for the same five instances is 97%.

7. CONCLUSION

This paper has presented a computational theory of figural completion. Figural completion was defined as the problem of computing a *labeled knot-diagram* representing an *anterior scene* from a set of contour fragments representing image luminance boundaries. Given this computational goal and this input, three distinct sources of ambiguity were identified. These were termed *shape, unit,* and *depth*. Since the applicable physical constraints were insufficient to overcome these ambiguities, an additional assumption was introduced. Broadly speaking, this assumption was: *the shape of a perceptual completion is independent of the role it plays in the organization*. This assumption allowed the figural completion problem to be decomposed into two independent subproblems, the first devoted to shape, the second devoted jointly to unit and depth. Decomposed in this way, it proved possible to formulate the second subproblem as an integer linear program, which led to a unique solution.

Next we described an experimental implementation. Our intention was not to realistically model human vision at the ''level of algorithm and representation'' [20]. Instead, our intention was to validate the computational theory by demonstrating that a well-defined procedure for computing the mapping between input and output does in fact exist. The experimental system was demonstrated on a number of illusory contour figures from the visual psychology literature.

We also considered the hypothesis of Kellman and Loukides [17] that *unit ambiguity is resolved independently and in advance of depth ambiguity*. The revised model was implemented in an experimental system. In the majority of cases, the implementation of the revised model gives the same solution and at significant computational savings.

In conclusion, we note that there are many places where the scope of the existing theory is simply too narrow and is insufficient to account for human competence. For example, the human ability to understand line-drawings of smooth surfaces embedded in ways which violate the definition of anterior scene suggests that the representation computed by the figure completion process is more general than (and probably subsumes) the labeled knot-diagrams developed here.

Toward this end, it is worth reexamining Huffman's [13] influential paper ''Impossible Objects as Nonsense Sentences.'' While this paper is widely cited as one source of the Huffman–Clowes junction catalog for trihedral scenes, the last few pages are actually devoted to a labeling scheme for line-drawings of smooth surfaces. We optimistically predict that the methods described here can be generalized to the full Huffman labeling scheme, leading to a computational theory sufficient to explain the full range of perceptual completion phenomena at work in human vision.

REFERENCES

1. H. G. Barrow and J. M. Tenenbaum, Recovering intrinsic scene characteristics from images,'' in *Computer Vision Systems* (A. R. Hanson and E. M. Riseman, Eds.), Academic Press, New York, 1978.
2. H. G. Barrow and J. M. Tenenbaum, Retrospective on interpreting line drawings as three dimensional surfaces. *Artificial Intelligence* **59,** 1993, 71–80.
3. M. Boldt, R. Weiss, and E. M. Riseman, Token based extraction of straight lines, *IEEE Trans. Syst. Man Cybern.* **19**(6), 1989, 1581–1594.
4. J. Burnett and U. Buy, Solving integer programming systems using the IMINOS prototype, Technical report, Department of Computer Science, University of Massachusetts, in preparation.
5. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis,* Wiley, New York, 1973.
6. M. Fahle and G. Palm, Perceptual rivalry between illusory and real contours, *Biol. Cybern.* **66**(1), 1991, 1–8.
7. L. H. Finkel and P. Sajda, Object discrimination based on depth-from-occlusion, *Neural Computation* **4,** 1992, 901–921.
8. H. B. Griffiths, *Surfaces,* 2nd ed., Cambridge Univ. Press, London/New York, 1981.
9. S. Grossberg and E. Mingolla, Neural dynamics of surface perception: Boundary webs, illuminants, and shape from shading, *Comput. Vision Graphics Image Process.* **37,** 1987, 116–165.
10. G. Guy and G. Medioni, Inferring global perceptual contours from local features, *Proceedings, DARPA Image Understanding Workshop, Washington, DC, 1993,* pp. 881–892.
11. R. Heitger and R. von der Heydt, A computational model of neural contour processing, figure-ground and illusory contours, *Proceedings, 4th International Conference on Computer Vision, Berlin, Germany, 1993.*
12. B. K. P. Horn, The curve of least energy, MIT AI Lab Memo No. 612, Artificial Intelligence Laboratory, MIT, Cambridge, MA, 1981.

13. D. A. Huffman, Impossible objects as nonsense sentences, *Machine Intell.* **6,** 1971, 295–323.
14. G. Kanizsa, *Organization in Vision,* Praeger, New York, 1979.
15. M. Kass, A. Witkin, and D. Terzopolous, Snakes: Active minimum energy seeking contours, *Proceedings, First International Conference on Computer Vision, London, England, 1987.*
16. L. Kauffman, Planar surface immersions, *Illinois J. Mathematics* **23,** 1979, 648–665.
17. P. J. Kellman and M. G. Loukides, An object perception approach to static and kinetic subjective contours, in *The Perception of Illusory Contours,* (S. Petry and G. Meyer, Eds.), pp. 151–164, Springer-Verlag, New York, 1987.
18. J. J. Koenderink, The shape of smooth objects and the way contours end, in *Natural Computation,* (W. Richards, Ed.), MIT Press, Cambridge, MA, 1988.
19. D. Lowe, *Perceptual Organization and Visual Recognition,* Kluwer Academic, Dordrecht, 1985.
20. D. Marr, *Vision,* Freeman Press, San Francisco, 1982.
21. M. Nitzberg and D. Mumford, The 2.1-D sketch, *Proceedings, 3rd International Conference on Computer Vision, Osaka, Japan, 1990,* pp. 138–144.
22. C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity,* Prentice-Hall, Englewood Cliffs, NJ, 1982.
23. S. Petry and G. E. Meyer, Eds., *The Perception of Illusory Contours,* Springer-Verlag, New York, 1987.
24. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C,* Cambridge University Press, Cambridge, UK, 1988.
25. I. Rock, *The Logic of Perception,* MIT Press, Cambridge, MA, 1983.
26. H. M. Salkin and K. Mathur, *Foundations of Integer Programming,* North-Holland, New York, 1989.
27. A. Shashua and S. Ullman, Structural saliency: The detection of globally salient structures using a locally connected network, *2nd International Conference on Computer Vision, Clearwater, FL, 1988,* pp. 321–327.
28. S. Ullman, Filling-in the gaps: The shape of subjective contours and a model for their generation, *Biol. Cybern.* **21,** 1976, 1–6.
29. L. R. Williams, Perceptual Completion of Occluded Surfaces, Ph.D. Dissertation, Dept. of Computer Science, Univ. of Massachusetts at Amherst, Amherst, MA, 1994.
30. L. R. Williams and D. W. Jacobs, Stochastic completion fields: A neural model of illusory contour shape and salience, *Proceedings, 5th International Conference on Computer Vision,* Cambridge, MA, 1995.
31. A. P. Witkin and J. M. Tenenbaum, On the role of structure in vision, *Human and Machine Vision,* (Beck, Hope and Rosenfeld, Eds.), pp. 481–543, Academic Press, San Diego, 1983.