# Topological Requirements of Figural Completion

Lance R. Williams
Department of Computer Science
University of Massachusetts at Amherst
Amherst, Massachusetts 01003

**Abstract**

Figural completion, in its most common form, is the preattentive inference of the shape and location of the occluded parts of visible surfaces. Yet it also underlies illusory contour phenomena, which are associated with completion of occluding, rather than occluded surfaces. Phenomenological accounts, such as Kanizsa's[5], though insightful, lack the precision necessary to formulate a computational theory. The mathematics relevant to understanding the problem is the topology of surfaces. Accordingly, the topology of surfaces is reviewed with the aim of more precisely characterizing the computational goal underlying figural completion. A contour labeling scheme incorporating a minimum set of necessary constraints on the appearance of boundaries in flat scenes is then defined. These constraints, when enforced at points of contour intersection, are subsequently shown to be both necessary and sufficient, in the sense that sets of contours satisfying the labeling scheme can always be interpolated to form topologically valid surfaces.

# Topological Requirements of Figural Completion

**Abstract**

Figural completion, in its most common form, is the preattentive inference of the shape and location of the occluded parts of visible surfaces. Yet it also underlies illusory contour phenomena, which are associated with completion of occluding, rather than occluded surfaces. Phenomenological accounts, such as Kanizsa's[5], though insightful, lack the precision necessary to formulate a computational theory. The mathematics relevant to understanding the problem is the topology of surfaces. Accordingly, the topology of surfaces is reviewed with the aim of more precisely characterizing the computational goal underlying figural completion. A contour labeling scheme incorporating a minimum set of necessary constraints on the appearance of boundaries in flat scenes is then defined. These constraints, when enforced at points of contour intersection, are subsequently shown to be both necessary and sufficient, in the sense that sets of contours satisfying the labeling scheme can always be interpolated to form topologically valid surfaces.

# 1 Introduction

The input to the human visual system is a set of simple pointwise measurements of the image brightness function. Among human vision theorists, one opinion holds that the huge difference in level of organization between the visual system's input, and the abstract inferences about the world we derive from it, necessitates the existence of more highly structured intermediate representations. The proponents of this theory, which is often associated with Marr[8], maintain that computing these intermediate representations is the goal of human early visual processing.

The cornerstone of Marr's theory is an intermediate representation called the $2\frac{1}{2}$-*D sketch.* The $2\frac{1}{2}$-D sketch is a representation of depth and surface orientation (and discontinuities in these quantities) at every image point. The word "sketch" denotes the fact that the representation is spatially indexed. In formulating the $2\frac{1}{2}$-D sketch, Marr was clearly influenced by Horn's pioneering work on shape-from-shading[4]. Horn demonstrated that under certain circumstances, surface orientation can be recovered directly from image brightnesses. Barrow and Tenenbaum[1] proposed that techniques similiar to Horn's could, in principle, recover each of the physical parameters underlying the image brightness function, including depth, surface orientation, reflectance and illumination at every image point. Unfortunately, shape-from-shading has proved to be a poor first example in this respect.

Shape-from-shading is relatively unique among "shape-from" methods since image brightness constrains surface orientation everywhere, simplifying the computation of a dense array of depth values. Yet partly because of this, and partly because Marr emphasized random dot stereograms and other densely textured smooth surfaces in his computational study of human stereopsis, the field has been saddled with a too simple view of the $2\frac{1}{2}$-D sketch, a view plainly in conflict with phenomena which suggest a representation with richer topological structure. Principal among these is *figural completion.*

Figural completion, in its most common form, is the preattentive inference of the shape and location of the occluded parts of visible surfaces. Yet it also underlies illusory contour phenomena, which are associated with completion of occluding, rather than occluded surfaces. The need to complete occluding surfaces derives from the fact that under certain conditions,

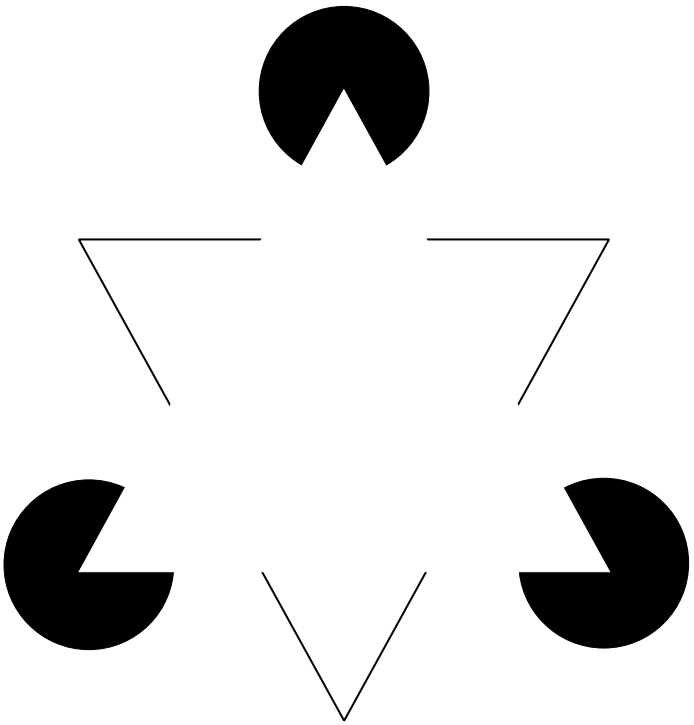

Figure 1: The Kanizsa triangle.

parts of surfaces which are in plain view are effectively invisible. Consider a planar surface which partially occludes a second planar surface of similiar orientation and reflectance. Under these circumstances, there will be little or no change in image brightness associated with the change in depth. Figure 1 is the well known Kanizsa triangle[5]. In Kanizsa's figure, an illusory triangle appears to partially occlude three black discs and a second triangle rendered in outline. Of course none of these elements is objectively present. All are the products of the figural completion process.

Confronted with evidence of figural completion like Kanizsa's triangle, let us assume that the computational goal of early visual processing is to build as complete a representation of visible surfaces as is possible, and not simply to build a representation of the visible portions of visible surfaces. What particular ability, or "competence," does the human visual system exhibit when it manufactures the illusory triangle in Kanizsa's figure? Does possession of this ability contradict the standard view of the $2\frac{1}{2}$-D sketch as simply a dense array of depth values? If the answer is yes, and I believe that this is self-evident, then what *are* the minimum topological requirements of a representation sufficient to explain this competence? To begin to answer these questions, it will be necessary to first consider more precisely what is meant by "surface."

# 2 Computational Goal

In mathematics, surfaces are classified as belonging to different families based upon a relatively small number of properties. Since these properties determine the ways in which a surface can be embedded in physical space, they ultimately limit its range of appearances in an image. Yet, despite this variety, a model of any surface can be constructed by "gluing" together some combination of the edges of one or more "paper" panels. Panels are the elementary building blocks from which more complex surfaces are constructed. For the purposes of this discussion, a panel can be envisioned as a surface cutout of a single piece of paper with a boundary consisting of a cycle of edges. A surface is completely determined by a set of panels, together with a set of *identifications*, which prescribe the gluing operations to be performed among the panels' edges. These elements together define a *paneling*. Although every surface can be constructed in this

fashion, not every paneling represents a surface.

The definition of identification given in the previous paragraph is incomplete. In particular, it does not take into account the fact that two edges of equal length may be glued, one to another, in two different ways. The ambiguity can be removed by first orienting each edge so that it matches the direction of a clockwise traversal of the panel boundary. The *sense* of the identification then depends on whether two edges are glued such that their orientations are the same or opposed.

For example, a surface can be constructed by identifying two opposite edges of a rectangularly shaped panel and leaving the other two non-identified (See Figure 2). If the edges are identified such that their orientations are opposed, then an *annulus* is created. However, if their orientations are the same, then the surface which results is a *Moebius strip*. These two surfaces are qualitatively very different: The annulus has two sides and a boundary consisting of two components while the Moebius strip has a single side and a boundary consisting of a single component. A surface with a single side is said to be *non-orientable*. Orientability affects the way in which a surface can be embedded in space. In particular, a non-orientable surface can not be flattened; any attempt to flatten a Moebius strip will result in the creation of a crease or twist. In the Kanizsa triangle figure, the subjective experience is of flat surfaces, overlapping one another in depth. Since all of these surfaces are two-sided, further consideration of non-orientable surfaces would seem to offer little insight into the problem at hand. Henceforward, we will assume that the orientations of all pairs of identified edges are opposed, and that the surfaces are, consequently, orientable.

Apart from orientability, surfaces can also be classified by the number of components in their boundary. A surface possesses a boundary if and only if at least one edge of some panel from which it is assembled remains non-identified. Within the paneling, non-identified edges form *boundary edges* and pairs of identified edges form *interior edges*. Each boundary component consists of a distinct cycle of boundary edges. For example, the panelings of the annulus and Moebius strip both contain a single interior edge (i.e. $a$) and two boundary edges (i.e. $b$ and $c$). But the boundary edges in the paneling of the annulus form two boundary components while

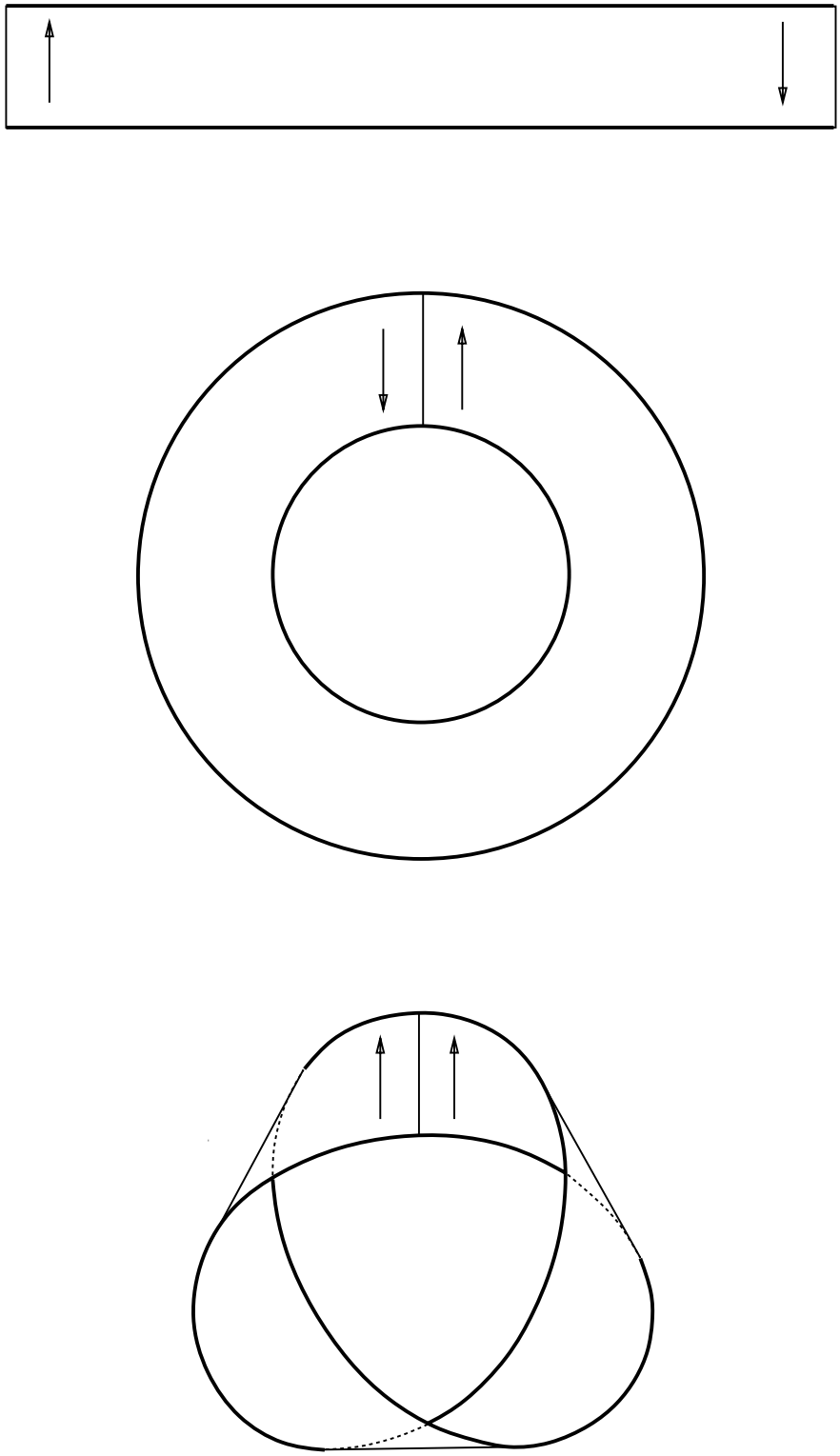

Figure 2: A surface can be constructed by identifying two opposite edges of a rectangularly shaped panel and leaving the other two non-identified. Depending on the sense of the identification, the surface which results is either an *annulus* or a *Moebius strip*.

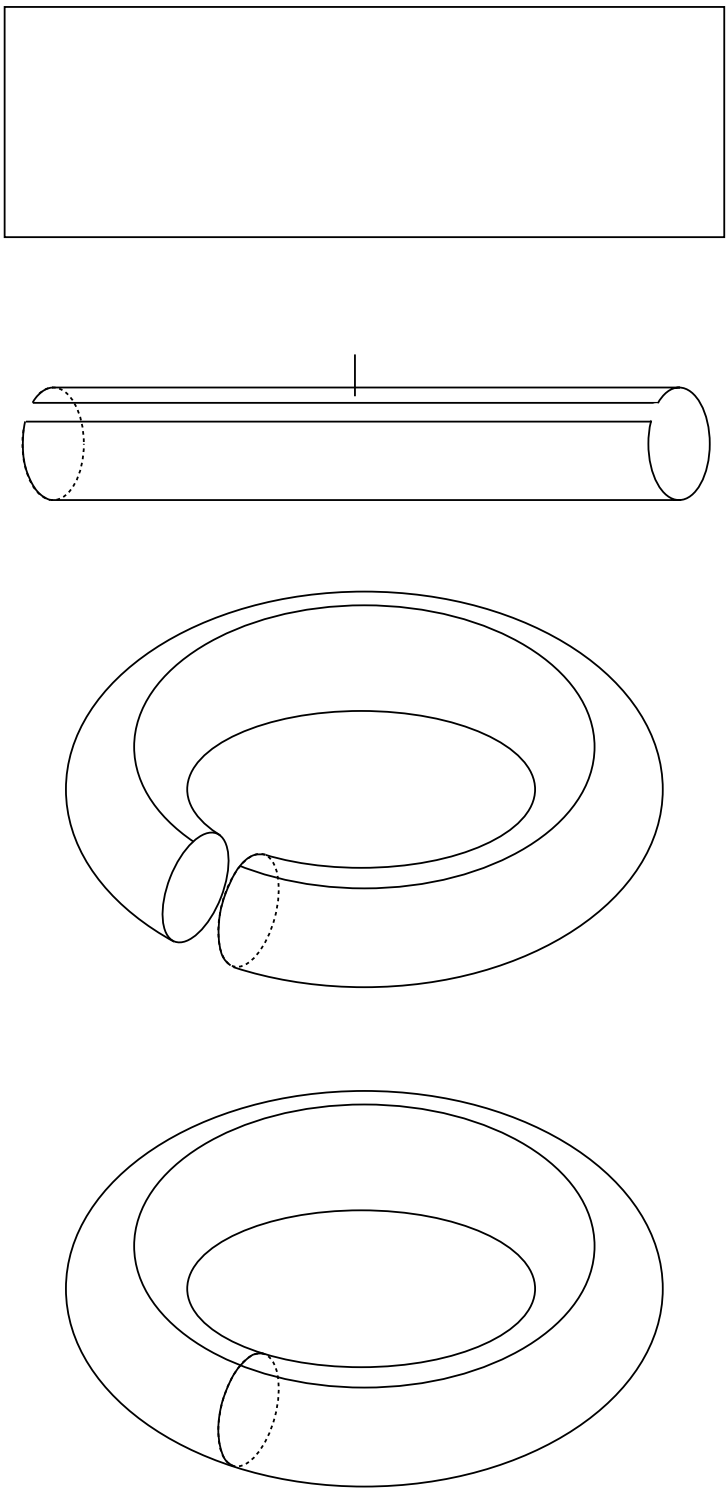

Figure 3: The *torus* is constructed by identifying the opposite edges of a rectangular panel as illustrated.

those of the Moebius strip form just one.

The *torus* is constructed by identifying the opposite edges of a rectangular panel as illustrated in Figure 3. Although the torus has no boundary, it still generates contours in the image. These are *occluding contours*, and are the projection onto the image plane of the set of points on the surface tangent to the lines of sight. Since its construction is effected without forming a Moebius strip, the torus, besides having no boundary, is also orientable. An orientable surface without boundary divides space into two disjoint sets. These sets consist, respectively, of interior points and exterior points. Orientable surfaces without boundary are important in geometric and solid modeling since the set of interior points can be interpreted as a *manifold solid*. Of course, Kanizsa's figure is experienced as a set of flat, overlapping surfaces, not as a set of smooth, occluding solids. The contours are interpreted as boundaries, not occluding contours. This is not to imply that a complete theory of human vision can ignore the representation and appearance of solids. Quite to the contrary. It is only saying that the perception of Kanizsa's triangle seems to specifically involve either the assumption or inference of flatness. It is therefore likely, that to understand human competence in this instance, surfaces without boundary, like non-orientable surfaces, can safely be ignored.

In the Kanizsa triangle figure, many different factors contribute to the appearance of flatness. Since the figure is printed on a flat page, stereo disparity suggests that all of the contours lie at constant depth. Monocular cues, such as the uniform brightness of the figure's regions, reinforce this percept. Also telling is the appearance of the completed surface boundaries, which are punctuated by sharp discontinuities in orientation but lack the cusps and terminations characteristic of occluding contours (see Koenderink[7]). Finally, it may be that flatness is inseparably connected with perception of the illusory surface itself, since it provides the necessary explanation for the triangle's lack of contrast against its uniformly bright background. But if flatness is part of the vocabulary of early vision (perhaps serving as a default assumption), then how exactly is it defined?

Planar embeddings of surfaces topologically equivalent to discs with zero or more holes can be represented in a straightforward fashion as the intersection of domains defined by sets of closed

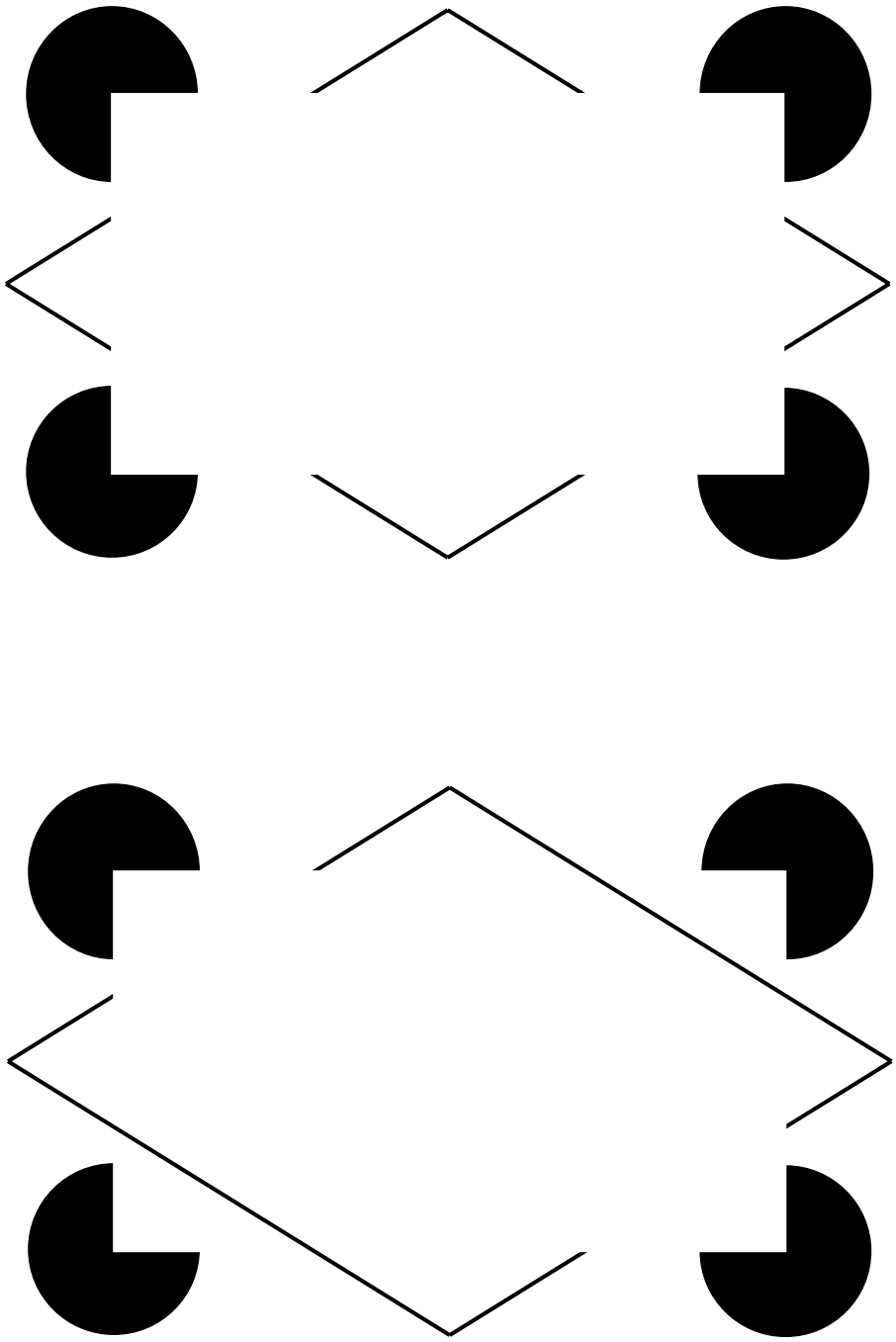

Figure 4: Compare the illusory surface in the top and bottom displays. Since there is no way to embed both the illusory surface and the diamond shaped outline in planes of constant depth, the human representation must be more general.

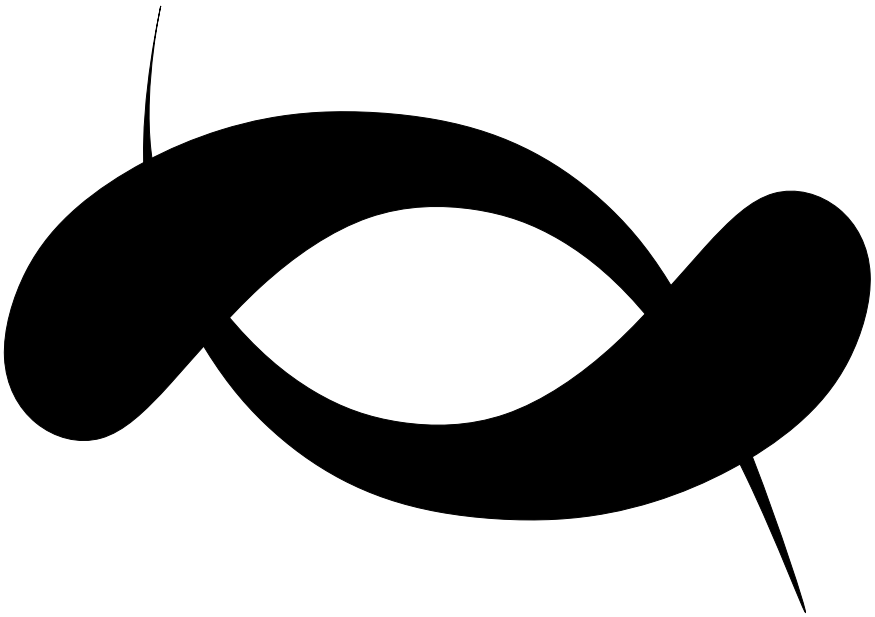

Figure 5: Although interpretations as surfaces embedded in planes at constant depth are plausible, they are not perceived. As Kanizsa points out, the slender "tail" of each surface seems to pass under the "head." The surfaces, nevertheless, remain flat.

non-intersecting plane curves, or *Jordan curves*. This observation leads naturally to the following question: Can a representation consisting of sets of Jordan curves at constant depth account for human competence in the Kanizsa triangle figure? This is the representation proposed by Nitzberg and Mumford[9]. Unfortunately, although sufficient for the Kanizsa triangle figure, the representation is not adequate for simple variations, such as Figure 4. In this display, a rectangularly shaped illusory surface seems to pass over and under a diamond shaped outline. Since there is no way to embed both the illusory surface and the diamond shaped outline in planes of constant depth, the human representation must be more general. Further evidence is provided by Figure 5, also designed by Kanizsa[5]. The striking thing about this figure is that even though an interpretation as two surfaces embedded in planes at constant depth is absolutely plausible, given the image evidence, it is not perceived. Nor is the interpretation as a single surface embedded in a plane at constant depth perceived, although this too, is entirely plausible. Instead, as Kanizsa points out, the slender "tail" of each surface seems to pass under the "head." The surfaces, nevertheless, remain flat.

Strictly speaking, flatness requires that the vectors normal to a surface be perpendicular to the plane of projection. However, if the definition of flatness is relaxed, so that the surface normals are required to be simply close to perpendicular to the plane of projection, then a much broader class of scenes is possible. These include non-planar embeddings of surfaces which are still, for all intents and purposes, flat:

**Definition** *flat scene - A set of surfaces with boundary embedded in three space such that the surface normals everywhere are nearly perpendicular to the plane of projection.*

Orientability and number of boundary components are useful for classification of surfaces because they are invariants of *homeomorphism*. A homeomorphism is a continuous one-to-one mapping (with continuous inverse mapping) of one topological space to another. Just as euclidean geometry is the branch of mathematics concerned with properties of space which are invariant under rigid motion, topology is the branch of mathematics concerned with properties of space invariant under homeomorphism. Apart from orientability and number of boundary components, the only other topological invariant is *genus*. We will see that although the bound-

aries of surfaces in flat scenes form a more general class of curves than Jordan curves in the plane, they are still Jordan curves...Jordan curves embedded in orientable surfaces of arbitrary genus.

The orientable surface without boundary of lowest genus is the sphere (genus zero) while the next lowest is the torus (genus one). As part of an inductive definition of genus, it will be useful to introduce a technique for building surfaces of higher genus from surfaces of lower genus. To begin, a disc is subtracted from two surfaces, creating a hole in each. A "tube" is then used to connect the hole in the first surface to the hole in the second surface. The resulting surface is termed the *connected sum*, and its genus is the sum of those of the two original surfaces. If this operation is performed on a single surface, it is called adding a *handle*, and increases the genus by one. The genus of an orientable surface with boundary is the number of handles which must be added to a sphere to form a topologically equivalent surface. Since the genus of the torus is one, it is equivalent to a sphere with one handle.

Every orientable surface with boundary can be created by subtracting one or more discs from a sphere with zero or more handles. Since this operation partitions the sphere with handles into disjoint regions, the boundaries of the resulting surfaces are Jordan curves. Genus zero surfaces with boundary are a degenerate case. They are created by subtracting one or more discs from a sphere with zero handles. Members of this subset can be embedded in a plane, where their boundaries remain Jordan curves. However, all orientable surfaces with boundary, even those which can not be embedded in a plane, can still be "flattened." For example, an orientable surface with boundary of genus one is created when a disc is subtracted from a torus. This surface, known as a *punctured torus*, is constructed by gluing the edges of a paper panel as illustrated in Figure 6. Yet it can be readily verified that the flat surface depicted in Figure 7 is also a punctured torus. First, it should be clear that prior to gluing along edges $a$ and $b$, the surface consists of a single rather oddly shaped panel. Note that this panel is topologically equivalent to the rectangular panel from which the punctured torus was constructed in Figure 7; when the boundary of each panel is traversed in the same direction, the edges are encountered in the same order: $acbfaebd$. It is also clear that the identifications of edges $a$ and $b$ indicated in Figure 7 correspond exactly with the identifications of edges $a$ and $b$ in Figure 6. Furthermore,

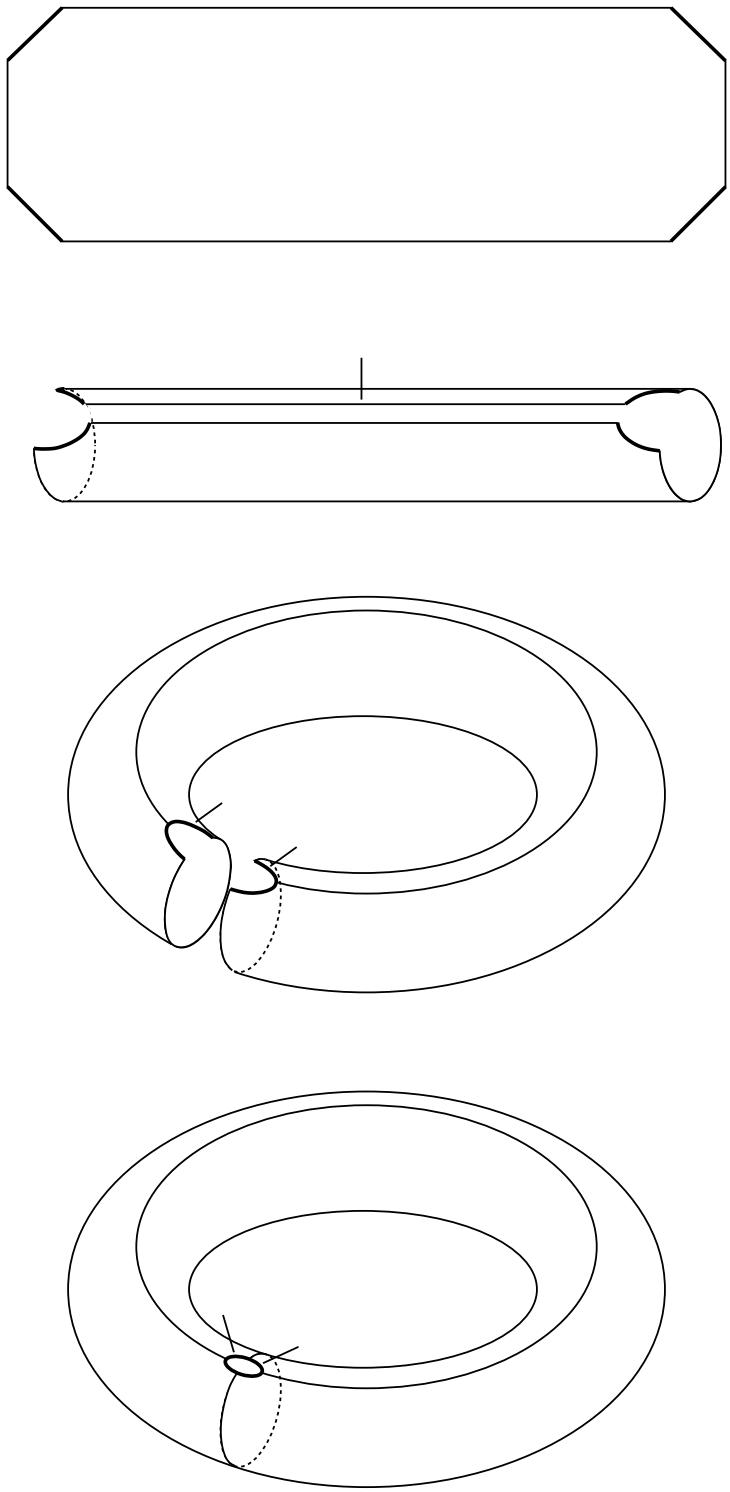

Figure 6: The *punctured torus*, an orientable surface with boundary, can be constructed from a single panel.

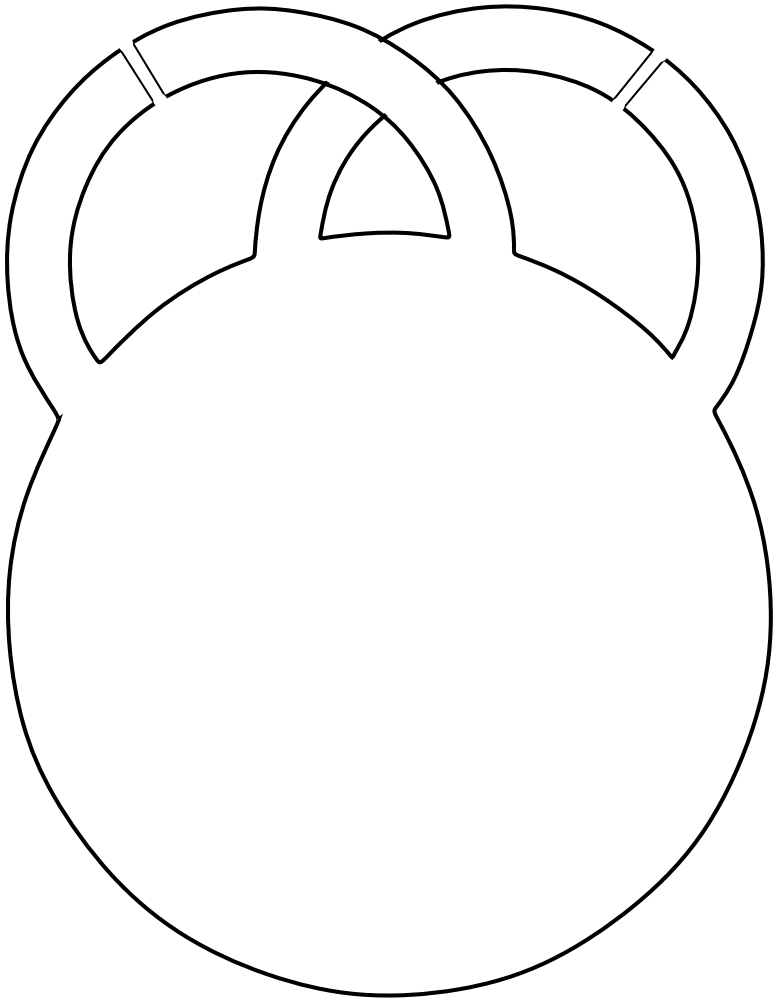

Figure 7: Although the punctured torus can not be embedded in a plane, it can still be flattened.

the cycle of free edges, $cfed$, forms the boundary of both surfaces. From a topological standpoint, the two surfaces are equivalent, even though the appearance of the boundary is very different in the respective figures.

# 3 Topological Requirements

Let us assume that the definition of flat scene is sufficiently general, and that the reconstruction of a flat scene adequately describes the competence exhibited in perception of the Kanizsa triangle figure. The associated completion problem can be divided into two interelated subproblems: 1) The problem of completing the set of surface boundary points; and 2) The problem of completing the set of surface interior points. To what extent does a solution to one of these problems determine the solution to the other? If only planar embeddings are allowed, the answer is very simple, since the shape of an embedded surface is uniquely determined by the shape of its boundaries. In this special case there is no "interpolation" problem, only the problem of identifying which boundary fragments match which. Of course this matching problem is strongly constrained by the requirement that boundaries must form sets of closed, oriented Jordan curves in planes at constant depth. Yet we have already seen that this requirement, while ensuring the topological validity of the completed surfaces, is somewhat artificial. It uneccessarily restricts the class of admissible scenes. Which leads to the following two questions: What topological requirements, at a minimum, are satisfied by the boundaries of a flat scene? How can a set of boundaries satisfying these less restrictive requirements be interpolated to build a representation of complete surfaces?

Because any surface's boundary is formed from cycles of non-identified edges in some paneling, each boundary component is topologically equivalent to a circle. An embedding of a topological circle in three space is called a *knot*. The abstract depiction of the projection of a knot onto a plane is called a *knot diagram*. A knot diagram is a closed plane curve which intersects itself at a finite number of points called *crossings*. Each of the closed plane curves which together comprise the projection of the boundary onto the image plane can be assigned a single orientation which everywhere indicates which side of the curve the image of the surface

lies. In this paper, we adopt the 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. This number will be referred as the boundary depth. If the view of the flat scene is generic, then crossings will be the only points of multiplicity two in the projection of the boundary onto the plane. Furthermore, boundary depth can change only at crossing points. In a knot diagram, crossings are drawn in a manner which explicitly indicates the relative depth of the two overlapping strands. 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 flat scenes. These constraints have been incorporated into the labeling scheme illustrated in Figure 8. The *writhe* of a crossing in a knot diagram is the sign of the cross product of the orientations of the upper and lower strands. Every crossing in a knot diagram is equivalent (after rotation) to a crossing with writhe equal to either $+1$ or $-1$. Crossings with opposite writhe are mirror images. For our purposes, the upper and lower strands represent the nearer and farther boundaries and the two different crossing labelings correspond to the two possible values for writhe. It can be easily verified that the depth labels of the different edges in the labeling scheme faithfully reproduce the effect of occlusion of the farther boundary as described in the previous paragraph. The labeling scheme can therefore be considered necessary in the sense that the image of the boundary of any flat scene satisfies the constraints. But does a set of closed contours satisfying the labeling scheme always represent a flat scene? Is the labeling scheme necessary *and* sufficient?

We now prove that a set of closed contours satisfying the labeling scheme illustrated in Figure 8 always defines a flat scene. First, constraints on the number of interior surface points which project to a single image point are identified. We then demonstrate that given an oriented knot diagram, values satisfying these constraints can always be found. This is the precondition of a procedure for constructing a paneling from a knot diagram satisfying the labeling scheme. Finally, we show that every paneling constructed with this procedure represents a flat scene

which projects generically as the labeled knot diagram.

**Theorem 1.** *Every knot diagram satisfying the labeling scheme illustrated below represents a generic view of a flat scene.*

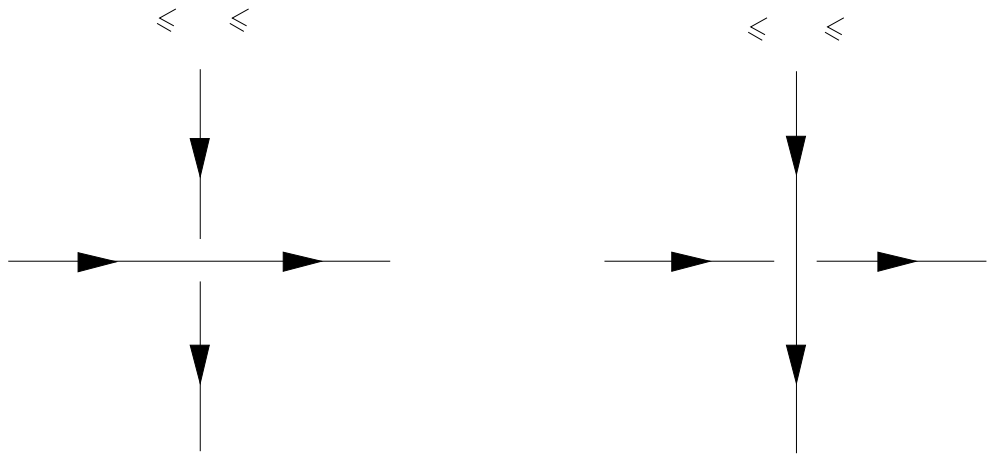

Figure 8: A labeling scheme incorporating a set of necessary constraints on the appearance of surface boundaries in flat scenes.

**Proof** Observe that a knot diagram partitions the plane into a set of disjoint planar regions. The boundary of each planar region is a cycle of oriented edges separated by crossing vertices. Every edge forms the side of exactly two planar regions, one lying to its right, the other to its left (where right and left are with respect to the edge's orientation). Note that if an edge is the projection of part of the boundary of a flat scene, then the multiplicity of the projection of interior surface points to image points is one greater on the right side of the edge than on the left. Furthermore, the multiplicity of the projection of interior surface points to image points will be constant within a planar region. Quite obviously, this is only true of flat embeddings; Folding a surface will result in changes in the multiplicity of the projection of interior surface points at image contours other than region boundaries.

Let $A$ and $B$ be neighboring regions in an oriented knot diagram and let $A$ lie to the right of $B$. If $\gamma_A$ and $\gamma_B$ are the multiplicities of the projection of interior surface points within regions $A$ and $B$, then $\gamma_A - \gamma_B = 1$. Observe that the set of difference constraints among all neighboring planar regions form the node-edge incidence matrix of a network. Let the nodes of the network corresponding to $A$ and $B$ be $v_A$ and $v_B$ respectively. We adopt the convention that the edge

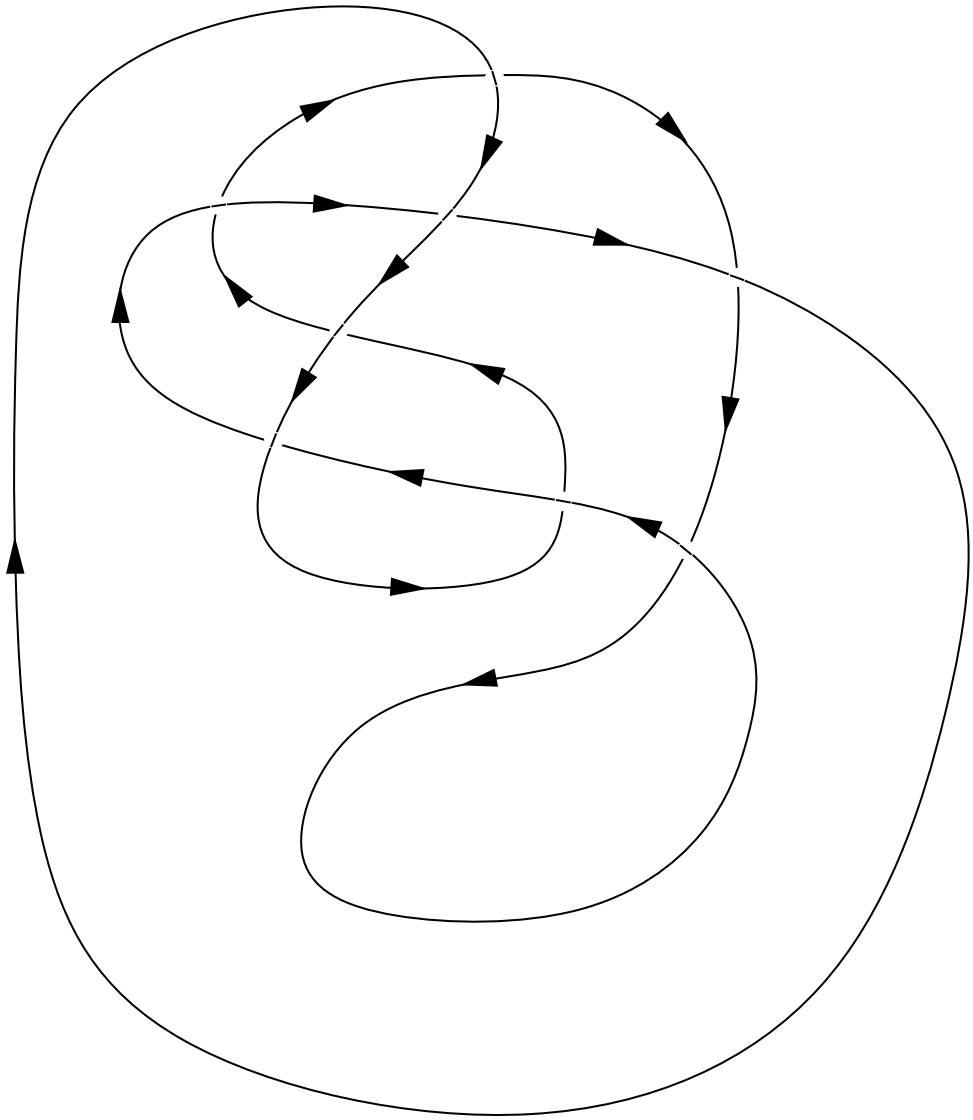

Figure 9: A labeled knot diagram partitions the plane into regions.

of the network joining $v_A$ with $v_B$ is directed from $v_A$ to $v_B$ when region $A$ lies to the right of region $B$ in the knot diagram so that the weight of an edge in the network is equal to one when traversed in the direction of its orientation.

**Example**

Figure 10 illustrates a network constructed in this fashion for the planar partition depicted in Figure 9. The linear system of difference equations represented by this network appear below:

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \gamma_A \\ \gamma_B \\ \gamma_C \\ \gamma_D \\ \gamma_E \\ \gamma_F \\ \gamma_G \\ \gamma_H \\ \gamma_I \\ \gamma_J \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Recall that a system of difference equations has a solution if and only if the sums of the weights of every cycle in its corresponding network equal zero (where the weight of an edge is interpreted as 1 or $-1$ depending on the direction of traversal). We demonstrate not only that a solution to this system of difference constraints always exists but also that a solution exists where the value of $\gamma$ for every planar region is greater than the largest depth label among all edges bordering that region in the knot diagram. Fortunately, this second condition is easy to satisfy, since it is always the case that if $\{x_1, x_2, ..., x_n\}$ is a solution to a system of difference equations, then $\{x_1 + c, x_2 + c, ..., x_n + c\}$ is also a solution for any constant $c$. Clearly, a sufficiently large $c$ can always be found, so it is sufficient to prove that the sums of the weights around every closed cycle in a network constructed as described equal zero.

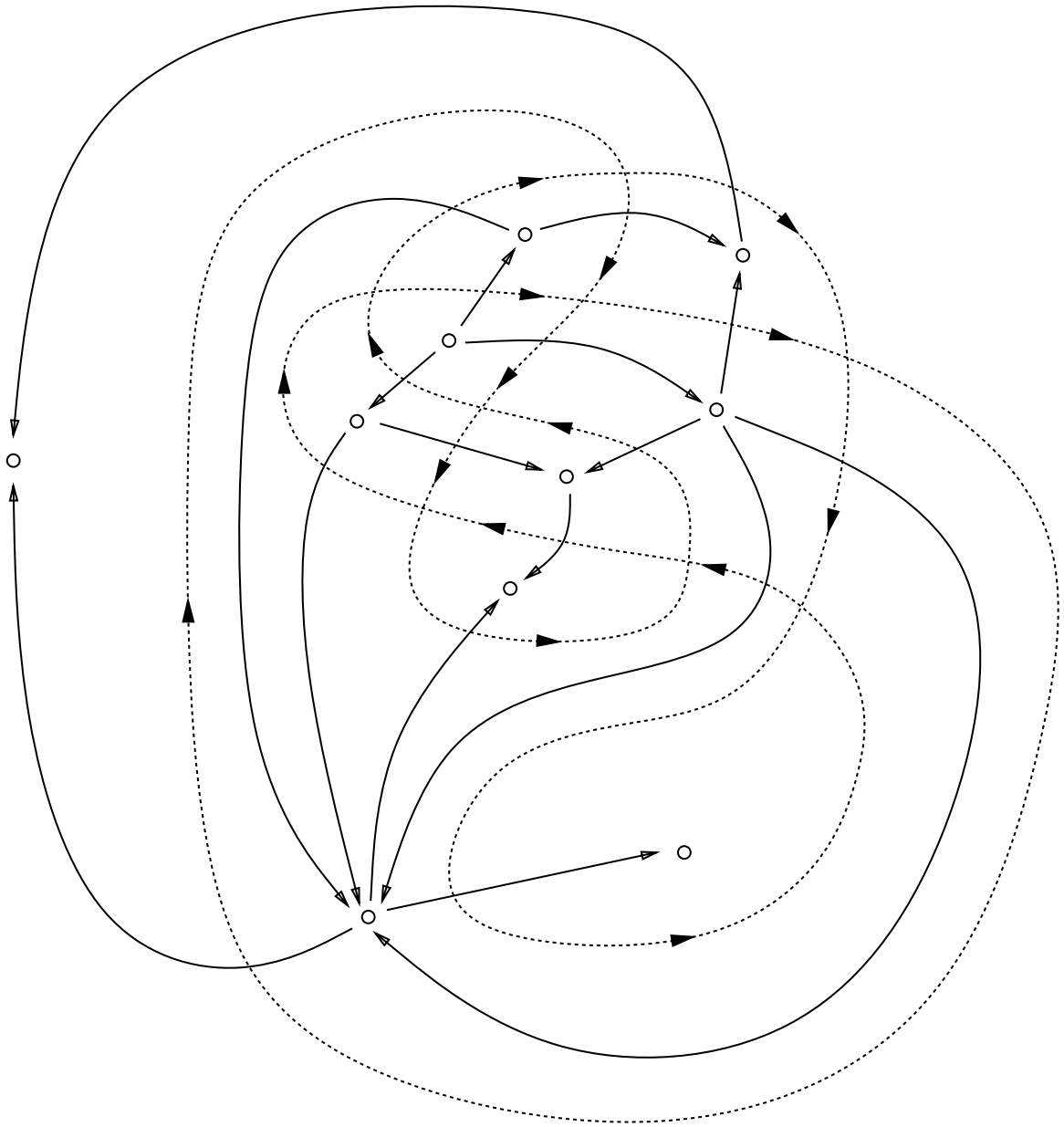

Figure 10: A network representing difference constraints on the multiplicity of the projection of interior surface points onto adjacent planar regions. The direction of the edges in the network is chosen so that the network weights are one when traversed in the direction of the arrows.

We begin by proving the following lemma:

**Lemma 1.** *Let $J$ be an oriented Jordan curve in the plane and let $C$ be an arbitrary, oriented, closed plane curve. If $J$ intersects $C$ at $m$ points, and if $\vec{j}_i$ and $\vec{c}_i$ are the vectors tangent to $J$ and $C$ at these points, then $\sum_{i=0}^{m-1} sgn(\vec{j}_i \times \vec{c}_i) = 0$.*

**Proof** A Jordan curve divides the plane into two disjoint regions which we call the *inside* and the *outside*. We adopt the convention that the inside lies to the right as the Jordan curve is traversed in the direction of its orientation while the outside lies to the left. If in the course of traversing oriented plane curve $C$, an ant crosses Jordan curve $J$ at crossing $i$, then the ant is conveyed either from the inside to the outside or from the outside to the inside. In the first case, $sgn(\vec{j}_i \times \vec{c}_i) = 1$ while in the second case $sgn(\vec{j}_i \times \vec{c}_i) = -1$. Since successive crossings, $i$ and $i+1$, must occur in opposite directions:

$$sgn(\vec{j}_i \times \vec{c}_i) + sgn(\vec{j}_{i+1} \times \vec{c}_{i+1}) = 0$$

Since in the course of a complete circuit, $C$ must intersect $J$ an even number of times,

$$\sum_{i=0}^{m-1} sgn(\vec{j}_i \times \vec{c}_i) = \sum_{i=0}^{m/2-1} sgn(\vec{j}_{2i} \times \vec{c}_{2i}) + sgn(\vec{j}_{2i+1} \times \vec{c}_{2i+1}) = 0 \quad \square$$

We now proceed with the proof that the sums of the weights around every closed cycle in a network constructed as described equal zero. Assign locations in the plane to the vertices in the network, such that each vertex is located within its respective planar region. Since edges only connect vertices located in adjacent planar regions, the network clearly has a planar embedding. We further note that every edge in the network need only cross an edge in the knot diagram once: At the boundary between adjacent regions. Furthermore, at these crossing points, the signs of the cross products of vectors tangent to edges of the network and edges of the knot diagram are everywhere equal to 1, which is the appropriate interpretation of the network edge weights when traversed in the direction of their orientation. Conversely, if the edge is traversed in the opposite direction, then the sign of the cross product is $-1$, which again corresponds to the appropriate interpretation of the weight of the network edge. Now, since the network is

a planar graph, the traversal of every simple cycle (i.e. a cycle in which no vertex is visited twice) traces a Jordan curve in the plane. Complex cycles, in turn, are the sums of one or more simple cycles, each of which is a Jordan curve. By Lemma 1.0, the sum of the signs of the cross products of vectors tangent to oriented Jordan and closed plane curves at their points of intersection is zero. Clearly then, the sums of the weights around every cycle in the network also equals zero, so that the system of difference equations always has a solution.

Let us summarize the proof to this point. We began with the observation that a knot diagram partitions the plane into regions. We then described a system of difference equations which the multiplicities of the projection of interior surface points onto the different regions must satisfy if the knot diagram is an image of a surface boundary. It was subsequently shown that a solution to this system of difference equations can always be found.

The second part of the proof is a description of a procedure for constructing a paneling given a knot diagram labeled according to the scheme illustrated in Figure 8 and a solution to the system of difference equations. We then prove that the paneling actually does represent a surface with boundary, by demonstrating that the neighborhood of every point is homeomorphic to either a disc or half disc.

Since each region of the planar partition induced by the knot diagram is a topological disc, flat panels of the same shape and size can be cutout from a sheet of paper. For each region, $R$, create $\gamma_R$ copies of the paper panel, where $\gamma_R$ is the multiplicity of the projection of interior surface points onto region $R$. Let the copies of region $R$ be $R(1), R(2), ..., R(\gamma_R)$ and let them be arranged in a stack above region $R$ in the plane such that $R(1)$ is the uppermost region and $R(\gamma_R)$ is the lowermost region.

Let $A$ and $B$ be neighboring regions and let $n$ be the boundary depth of the edge in the knot diagram separating them. Note that if $A$ lies to the right of $B$ then $\gamma_A - \gamma_B = 1$. Unless $n$ equals zero, identify the side (bordering $B$) of each panel (above region $A$) numbered 1 through $n$ with the adjacent side of the corresponding copy of region $B$ (i.e. $A(1) \leftrightarrow B(1), ..., A(n) \leftrightarrow B(n)$). Let the side of $A(n+1)$ adjacent to $B$ remain non-identified. Now, unless $\gamma_A$ equals $n+1$, identify the side (bordering $B$) of each panel (above region $A$) numbered $n+2$ through $\gamma_A$ with

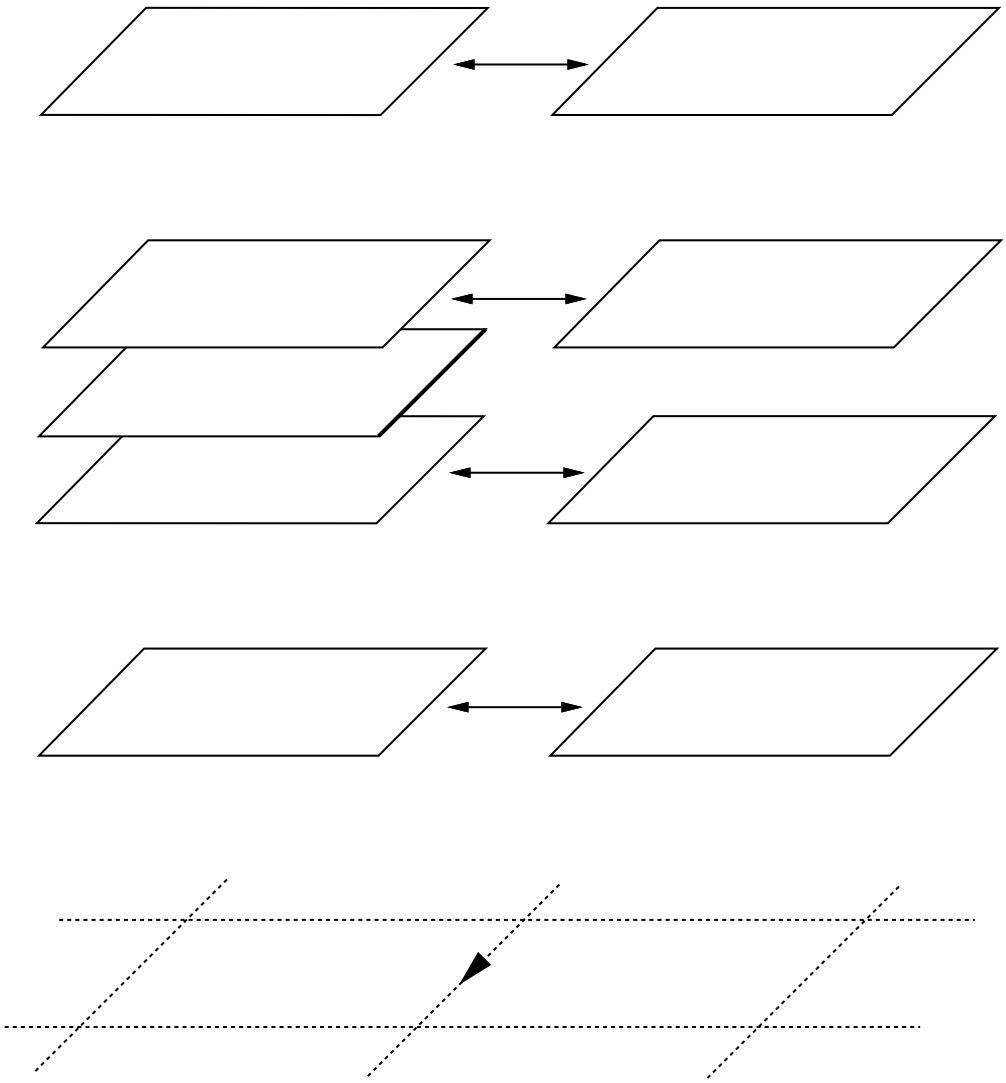

Figure 11: Paper panels stacked above regions $A$ and $B$ in the plane. Following the *identification scheme*, all copies of regions $A$ and $B$ but $A(n+1)$ are glued along their adjacent sides. The free side of $A(n+1)$ becomes part of the boundary of the surface.

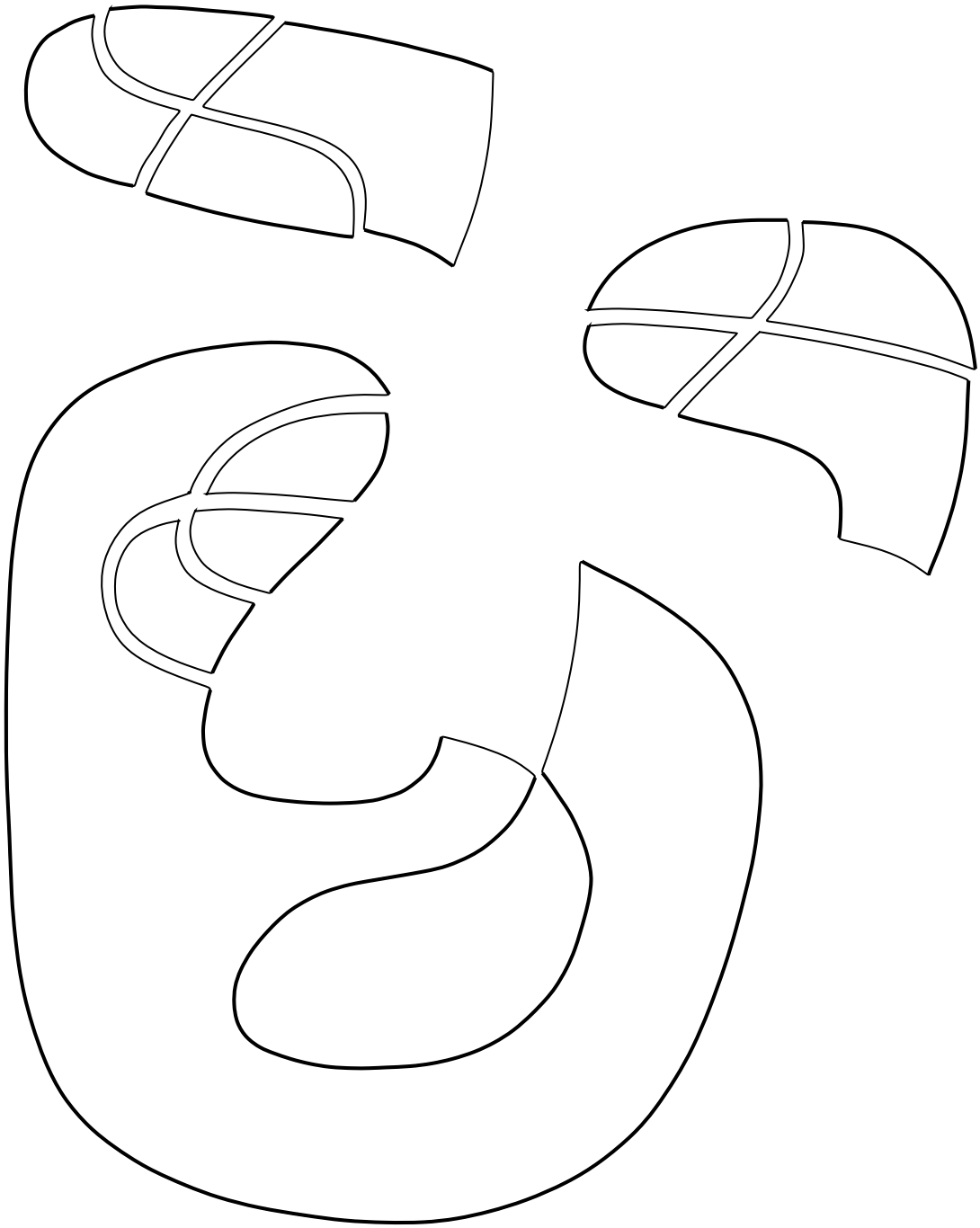

Figure 12: The paneling resulting from the identification scheme. Bold edges remain free, and form the boundary. Additional identifications are indicated by $x$ and $y$.

the adjacent side of the copy of region $B$ numbered $n+1$ through $\gamma_A - 1$ (i.e. $A(n+2) \leftrightarrow B(n+1), ..., A(\gamma_A) \leftrightarrow B(\gamma_A - 1)$). We refer to this implicitly defined set of edge identifications as the *identification scheme*. The effect of the identification scheme is to create $n$ interior edges above and $\gamma_A - n - 1$ interior edges beneath a single boundary edge in the paneling. The set of identifications can be divided into three subranges, the first and last of which are potentially empty:

(a) Unless $n = 0$, $A(1) \leftrightarrow B(1), ..., A(n) \leftrightarrow B(n)$.

(b) $A(n+1) \rightarrow \emptyset$.

(c) Unless $\gamma_A < n+2$, $A(n+2) \leftrightarrow B(n+1), ..., A(\gamma_A) \leftrightarrow B(\gamma_A - 1)$.

As previously noted, by everywhere gluing along the edges specified by the identification scheme, a paneling is created. However, we still must show that this paneling represents a flat scene. This will be done by demonstrating that the neighborhood of every point of the paneling has structure characteristic of either an interior surface point or a boundary point. Towards this end, we observe that points of the paneling can be divided into the following categories: 1) Points interior to a panel; 2) Points lying on an identified edge; 3) Points lying on a non-identified edge; and 4) Vertex points. In each of these cases, we demonstrate that the neighborhood of the point has structure characteristic of an interior surface point or a boundary point.

The first three cases are trivial. First, it is clear that a point interior to a panel forms an interior point of the surface. Second, the nature of the identification scheme insures that every panel edge is identified with at most one other. Pairs of identified panel edges therefore form interior edges of the paneling. Third, it is also clear that non-identified panel edges form boundary edges of the paneling. This leaves only the fourth case.

We therefore consider the neighborhood structure of vertex points. These are points of the paneling where the corners of two or more panels are incident. They should not be confused with crossings in the knot diagram. Specifically, it is important to understand that two or more vertices in the paneling are created when the construction is applied to the edges incident at a crossing in the knot diagram. We note that the result need only be demonstrated for crossings

with writhe equal to $+1$ since the case of crossings with writhe equal to $-1$ follows from mirror symmetry.

To better appreciate the need for an explicit proof of the proposition that neighborhoods of vertex points produced by the construction are homeomorphic to discs or half discs, it will be useful to study a negative example. The knot diagram shown in Figure 13 satisfies all components of the labeling scheme but the boundary depth order requirement. More specifically, although boundary depths are positive and depth change is consistent with the orientation of the occluding strand, the boundary depth of the occluding strand is *greater* than the depth of the occluded strand at the four crossings bordering region $I$. Nevertheless, the construction still can be applied to this knot diagram. This results in the paneling shown in Figure 14. The structure of the neighborhoods of each of the four crossings violating the depth order requirement is fairly complex, and is best appreciated through a paper model, which is readily constructed with scissors and tape. But even without building a model, one consequence of the unusual neighborhood structure can be readily appreciated: Although the boundary set is connected, the set of interior points is not! Since none of panel $I(2)$'s four sides is identified, each forms part of the boundary. Consequently, there exists no unbroken path, wholly interior to the surface, connecting an interior point of $I(2)$ with an interior point of any other panel. It is precisely this type of pathology that we wish to demonstrate is impossible in a knot diagram satisfying the labeling scheme in all respects. Toward that end, we continue the proof.

Let the four regions incident at a crossing with writhe equal to $+1$ be $A$, $B$, $C$ and $D$ as illustrated in Figure 15. Note that the depth of the edges dividing regions $A$ and $B$ is $m$, regions $C$ and $D$ is $m$, regions $A$ and $C$ is $n+1$, and regions $B$ and $D$ is $n$, with $0 \leq m \leq n \leq \gamma_C$ as guaranteed by the labeling scheme. Since region $C$ lies to the right of both of the strands formed by these edges, the multiplicity of region $C$ is one greater than the multiplicity of regions $A$ and $D$ (i.e. $\gamma_C = \gamma_D + 1 = \gamma_A + 1$) and two greater than the multiplicity of region $B$ (i.e. $\gamma_C = \gamma_B + 2$). We will show that, after gluing, exactly two of the $\gamma_C$ copies of region $C$ will form boundary vertices (with neighborhoods homeomorphic to half discs) while the remainder will form interior vertices (with neighborhoods homeomorphic to discs). In the process, all copies of the other three regions will be accounted for.

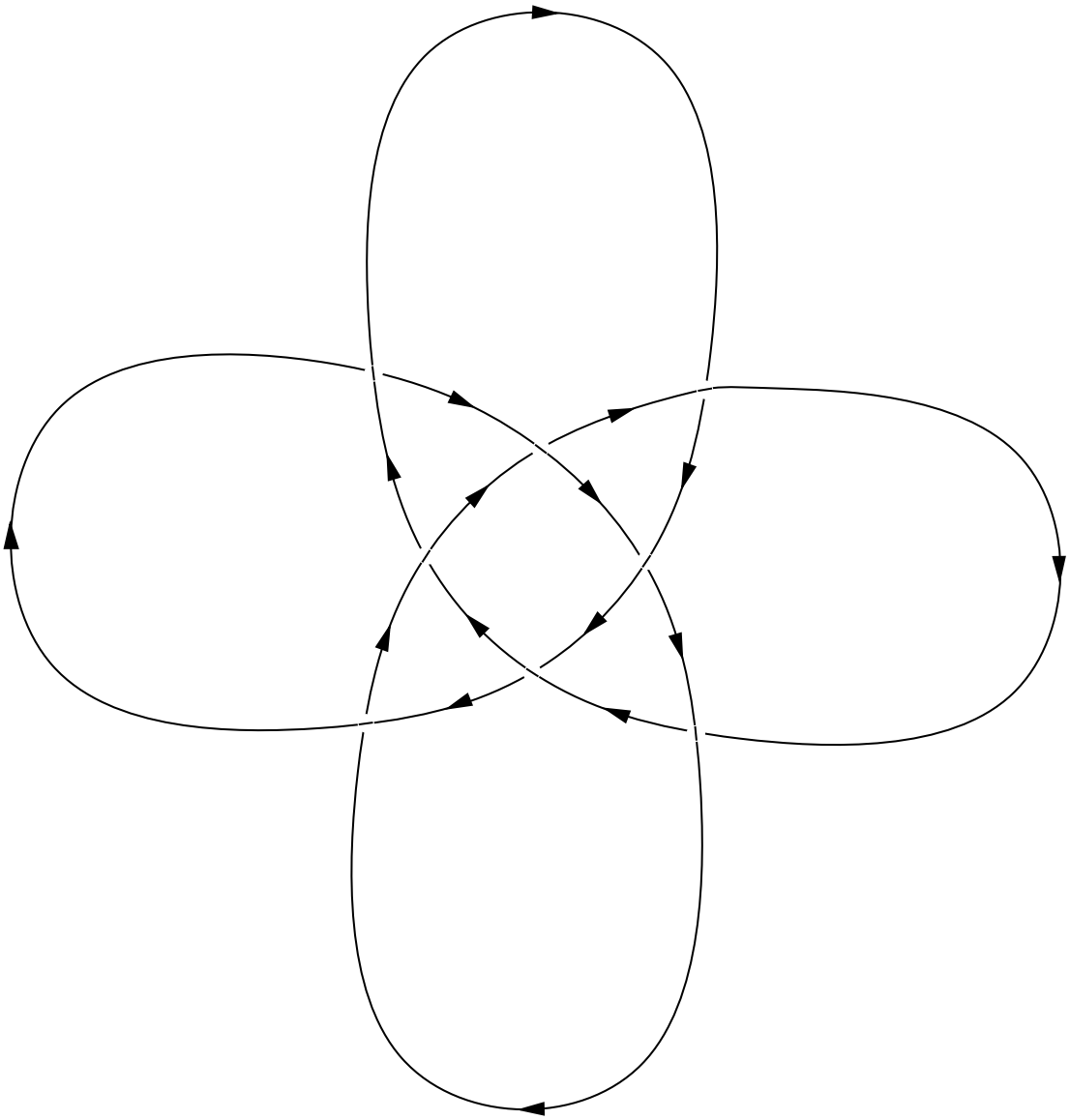

Figure 13: The labeling of the four crossings on the boundary of region $I$ is in violation of the depth order requirement.

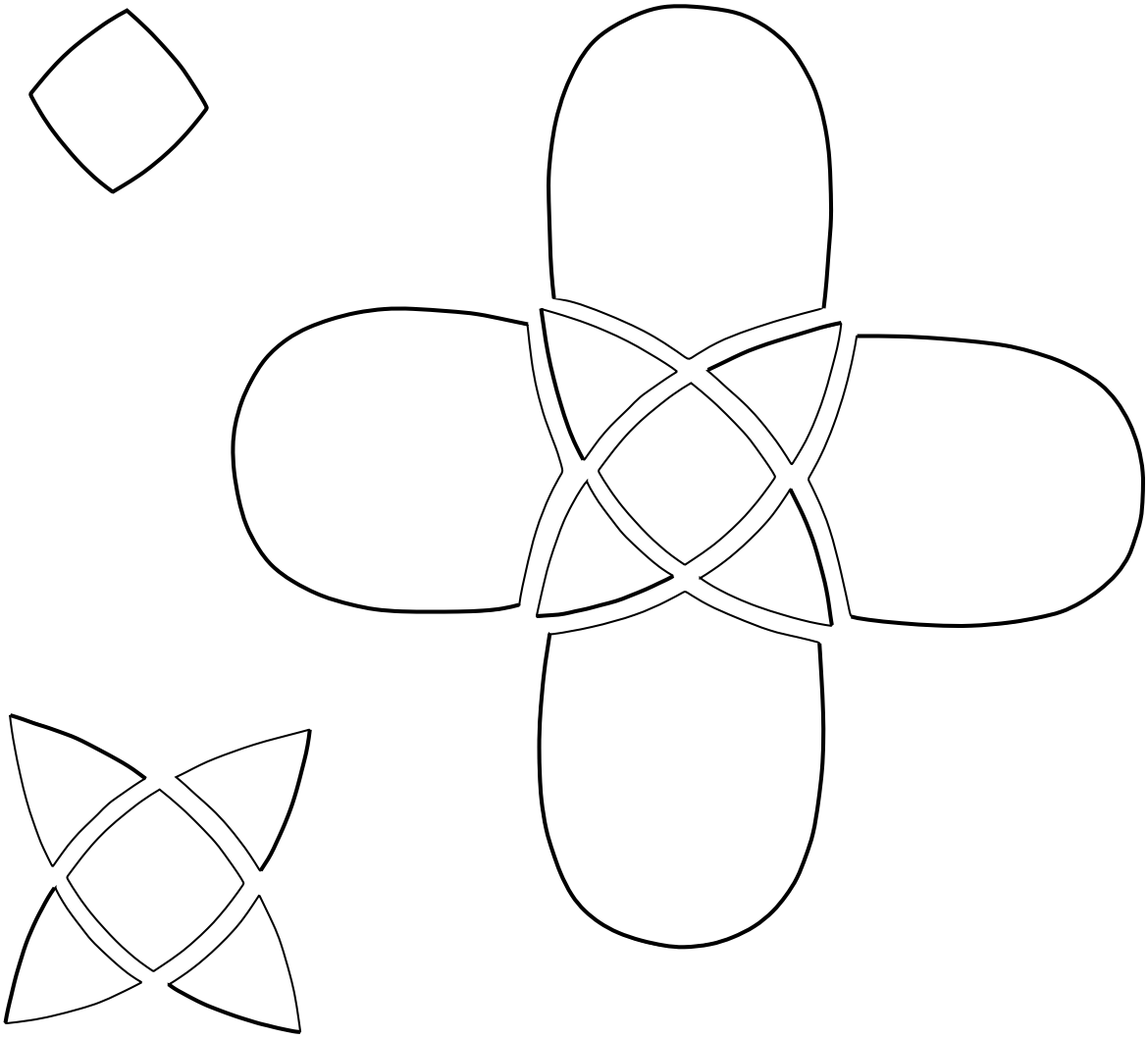

Figure 14: The paneling which results when the construction is applied to the knot diagram in the previous figure. Bold edges remain free, and form the boundary. Additional identifications are required along edges labeled $w, x, y$ and $z$. Although the set of boundary points form a single connected component, the set of interior points does not.

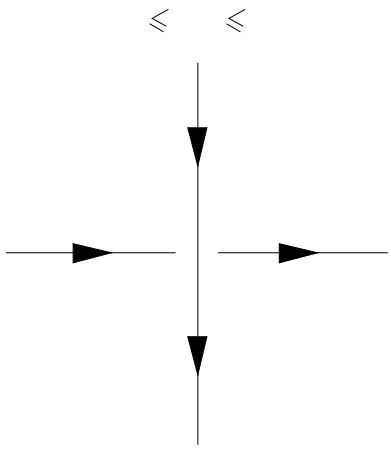

Figure 15: Four regions incident at a crossing with writhe equal to $+1$.

We begin by enumerating the set of edge identifications prescribed by the identification scheme for copies of regions $A$, $B$, $C$ and $D$. These identifications are understood to apply to the adjacent edges of the specified copies:

1. Identifications between copies of $A$ and $B$.
   (a) Unless $m = 0$, $A(1) \leftrightarrow B(1), ..., A(m) \leftrightarrow B(m)$.
   (b) $A(m+1) \rightarrow \emptyset$.
   (c) Unless $\gamma_A < m+2$, $A(m+2) \leftrightarrow B(m+1), ..., A(\gamma_A) \leftrightarrow B(\gamma_A - 1)$.
2. Identifications between copies of $C$ and $D$.
   (a) Unless $m = 0$, $C(1) \leftrightarrow D(1), ..., C(m) \leftrightarrow D(m)$.
   (b) $C(m+1) \rightarrow \emptyset$.
   (c) Unless $\gamma_C < m+2$, $C(m+2) \leftrightarrow D(m+1), ..., C(\gamma_C) \leftrightarrow D(\gamma_C - 1)$.
3. Identifications between copies of $A$ and $C$.
   (a) $C(1) \leftrightarrow A(1), ..., C(n+1) \leftrightarrow A(n+1)$.
   (b) $C(n+2) \rightarrow \emptyset$.
   (c) Unless $\gamma_C < n+3$, $C(n+3) \leftrightarrow A(n+2), ..., C(\gamma_C) \leftrightarrow A(\gamma_C - 1)$.

4. Identifications between copies of $B$ and $D$.

   (a) Unless $n = 0$, $D(1) \leftrightarrow B(1), ..., D(n) \leftrightarrow B(n)$.

   (b) $D(n+1) \rightarrow \emptyset$.

   (c) Unless $\gamma_D < n+2$, $D(n+2) \leftrightarrow B(n+1), ..., D(\gamma_D) \leftrightarrow B(\gamma_D - 1)$.

The identifications can be grouped into five consecutive subranges instead of three by exploiting the fact that the labeling scheme guarantees that $0 \leq m \leq n \leq \gamma_C$:

1. Identifications between copies of $A$ and $B$.

   (a) Unless $m = 0$, $A(1) \leftrightarrow B(1), ..., A(m) \leftrightarrow B(m)$.

   (b) $A(m+1) \rightarrow \emptyset$.

   (c) Unless $m+2 < n+1$, $A(m+2) \leftrightarrow B(m+1), ..., A(n+1) \leftrightarrow B(n)$.

   (d) None.

   (e) Unless $n+2 < \gamma_C - 1$, $A(n+2) \leftrightarrow B(n+1), ..., A(\gamma_C - 1) \leftrightarrow B(\gamma_C - 2)$.

2. Identifications between copies of $C$ and $D$.

   (a) Unless $m = 0$, $C(1) \leftrightarrow D(1), ..., C(m) \leftrightarrow D(m)$.

   (b) $C(m+1) \rightarrow \emptyset$.

   (c) Unless $m+2 < n+1$, $C(m+2) \leftrightarrow D(m+1), ..., C(n+1) \leftrightarrow D(n)$.

   (d) $C(n+2) \leftrightarrow D(n+1)$.

   (e) Unless $n+3 < \gamma_C$, $C(n+3) \leftrightarrow D(n+1), ..., C(\gamma_C) \leftrightarrow D(\gamma_C - 1)$.

3. Identifications between copies of $A$ and $C$.

   (a) Unless $m = 0$, $C(1) \leftrightarrow A(1), ..., C(m) \leftrightarrow A(m)$.

   (b) $C(m+1) \leftrightarrow A(m+1)$.

   (c) Unless $m+2 < n+1$, $C(m+2) \leftrightarrow A(m+2), ..., C(n+1) \leftrightarrow A(n+1)$.

(d) $C(n+2) \to \emptyset$.

(e) Unless $n+3 < \gamma_C$, $C(n+3) \leftrightarrow A(n+2), ..., C(\gamma_C) \leftrightarrow A(\gamma_C - 1)$.

4. Identifications between copies of $B$ and $D$.

(a) Unless $m = 0$, $D(1) \leftrightarrow B(1), ..., D(m) \leftrightarrow B(m)$.

(b) None.

(c) Unless $m+1 < n$, $D(m+1) \leftrightarrow B(m+1), ..., D(n) \leftrightarrow B(n)$.

(d) $D(n+1) \to \emptyset$.

(e) Unless $n+2 < \gamma_C - 1$, $D(n+2) \leftrightarrow B(n+1), ..., D(\gamma_C - 1) \leftrightarrow B(\gamma_C - 2)$.

The effect of gluing the panels according to the prescribed identifications is best illustrated by means of a diagram such as Figure 16. Pairs of identified edges are adjacent in the diagram. This diagram illustrates, in the most general case, the vertices of the paneling which are produced by the construction when applied at a single crossing. The fact that these and only these vertices are created can be verified by noting that: 1) Every identification prescribed by the identification scheme appears in the diagram; and 2) Every identification appearing in the diagram is prescribed by the identification scheme.

Whether or not the conditional identifications appear depends on the depths of the two strands overlapping at the crossing and the multiplicity of region $C$. The effect of the unconditional identifications (i.e. 1-4 (b) and (d)) is to create two vertices with neighborhoods homeomorphic to half discs in the paneling. These are boundary vertices. The effect of the conditional identifications (i.e. 1-4 (a),(c) and (e)) is: 1) To create $m$ interior vertices above the uppermost boundary vertex; 2) To create $n-m$ interior vertices between the uppermost and lowermost boundary vertices; and 3) To create $\gamma_C - n - 2$ interior vertices beneath the lowermost boundary vertex. Inspection of the diagram confirms that exactly four panels are incident at each interior vertex, and that the neighborhood structure of each interior vertex resembles a disc.

We now show that the image of the boundary of the flat scene produced by the construction corresponds to the knot diagram in every respect. First, the definition of the construction

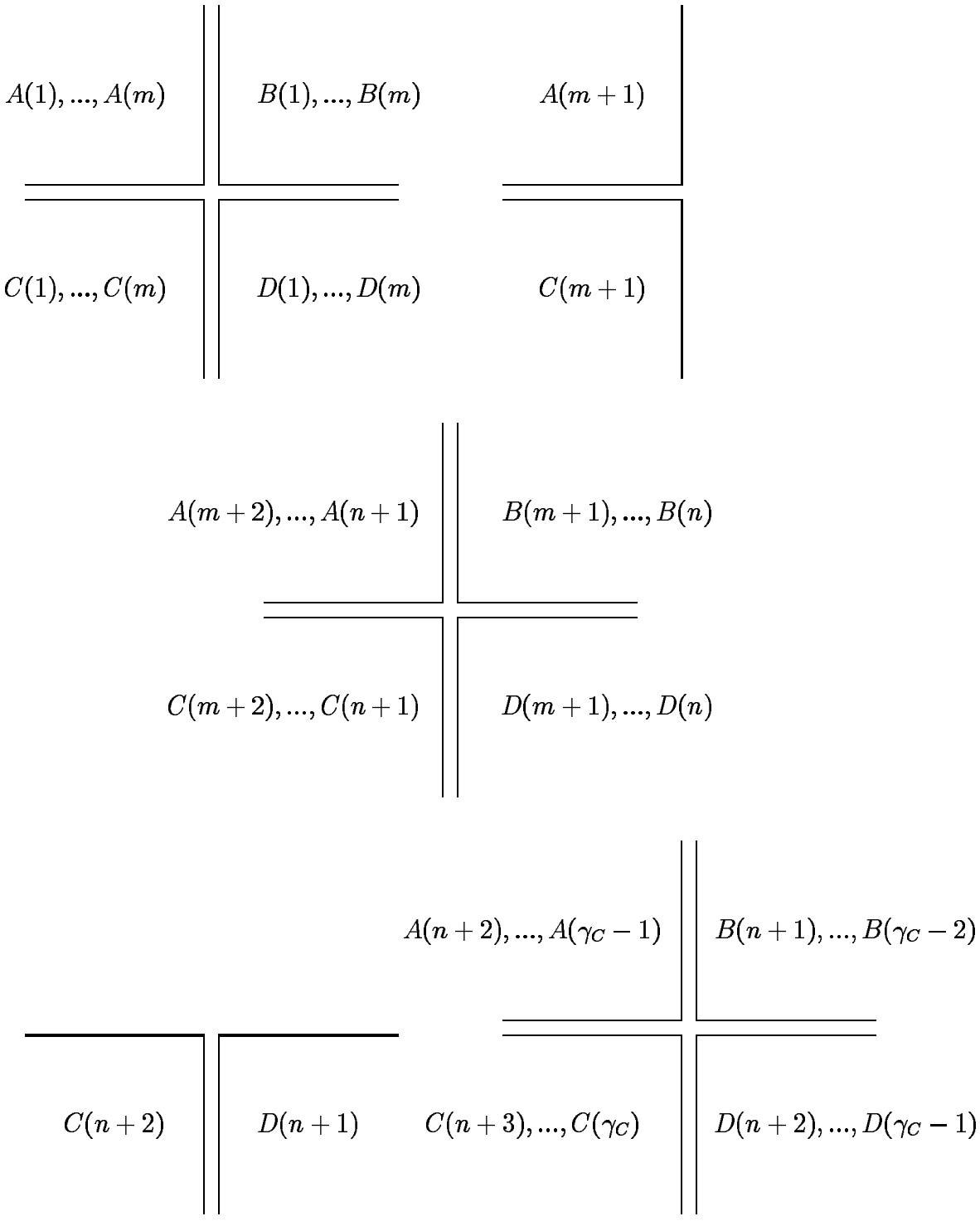


Figure 16: Paneling vertices produced by the construction when applied to edges incident at a crossing. Thick lines are boundary edges.

guarantees that each edge in the knot diagram produces exactly one non-identified edge in the paneling. The multiplicity of the projection of boundary points is therefore equal to one everywhere except at crossings. Furthermore, at crossings the multiplicity of the projection of boundary points is two, since exactly two boundary vertices are produced in the paneling when the construction is applied to the edges incident at a crossing. It follows that the image of the boundary is generic. Second, the definition of the construction guarantees that the image of the surface everywhere lies to the right of the image of its boundary, so that contour orientation is respected. Finally, the definition of the construction also guarantees that the boundary depth everywhere matches the depth attributes of the labeled knot diagram, since exactly $n$ interior panel edges are assembled above each boundary edge.□

# 4 Conclusion

This paper describes the theoretical basis for the working experimental system described in[10]. The experimental system employs a two stage process of completion hypothesis and combinatorial optimization. The labeling scheme, incorporating necessary and sufficient constraints for flat scenes, is enforced by a system of integer linear inequalities. The final organization is the optimal feasible solution of an integer linear program. This paper extends that work in the following ways:

1. The topological requirements implicit in the system of integer linear inequalities are explicitly identified and incorporated in a contour labeling scheme which is more easily understood.

2. The labeling scheme is shown to be necessary and sufficient, in the sense that it implicitly defines a domain of flat scenes more general than sets of surfaces embedded in planes at constant depth.

3. In the course of the sufficiency proof, a procedure for interpolating a set of boundaries satisfying the labeling scheme is described.

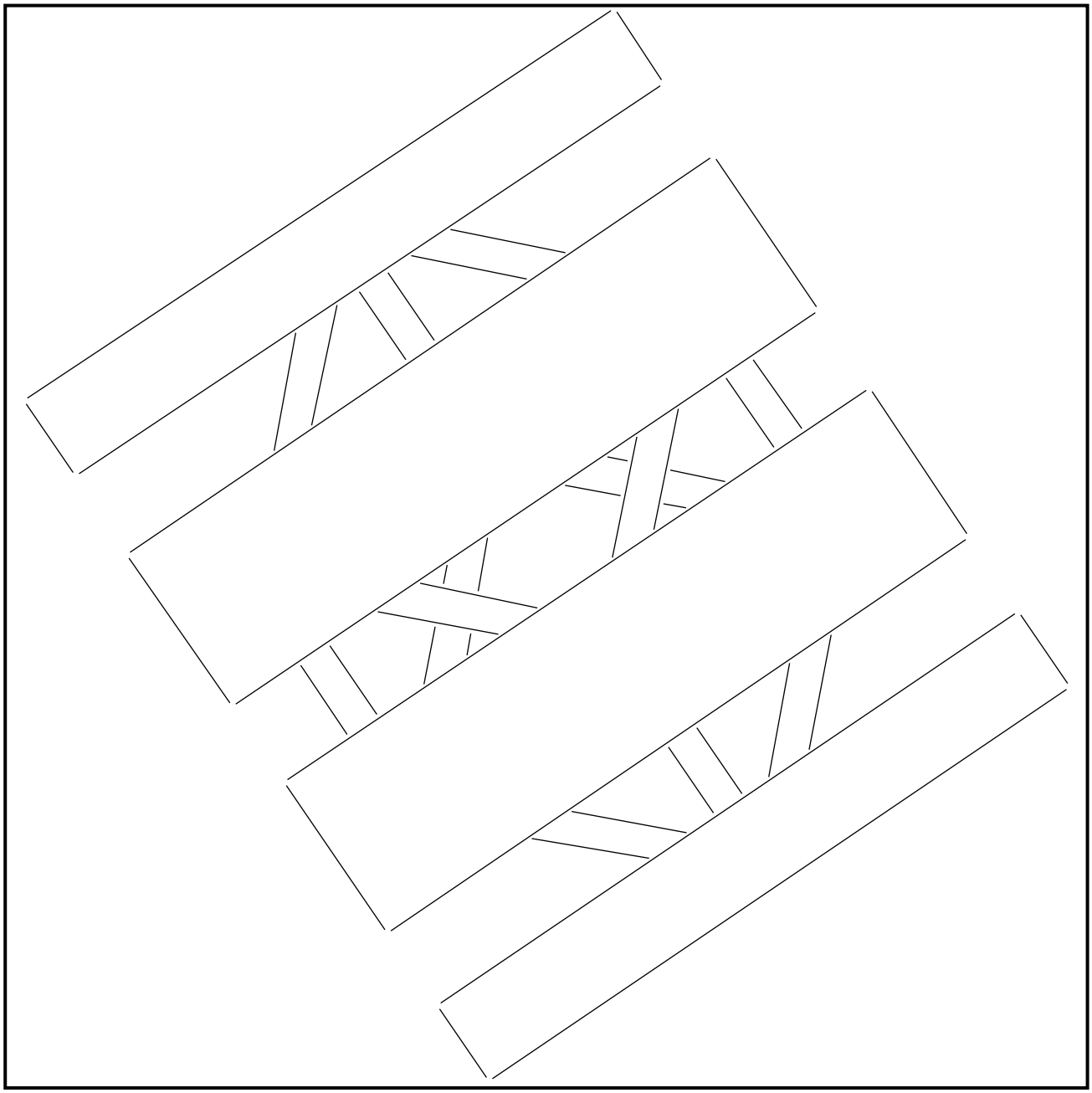

Figure 17: Kanizsa's partially occluded cube, a difficult completion problem.

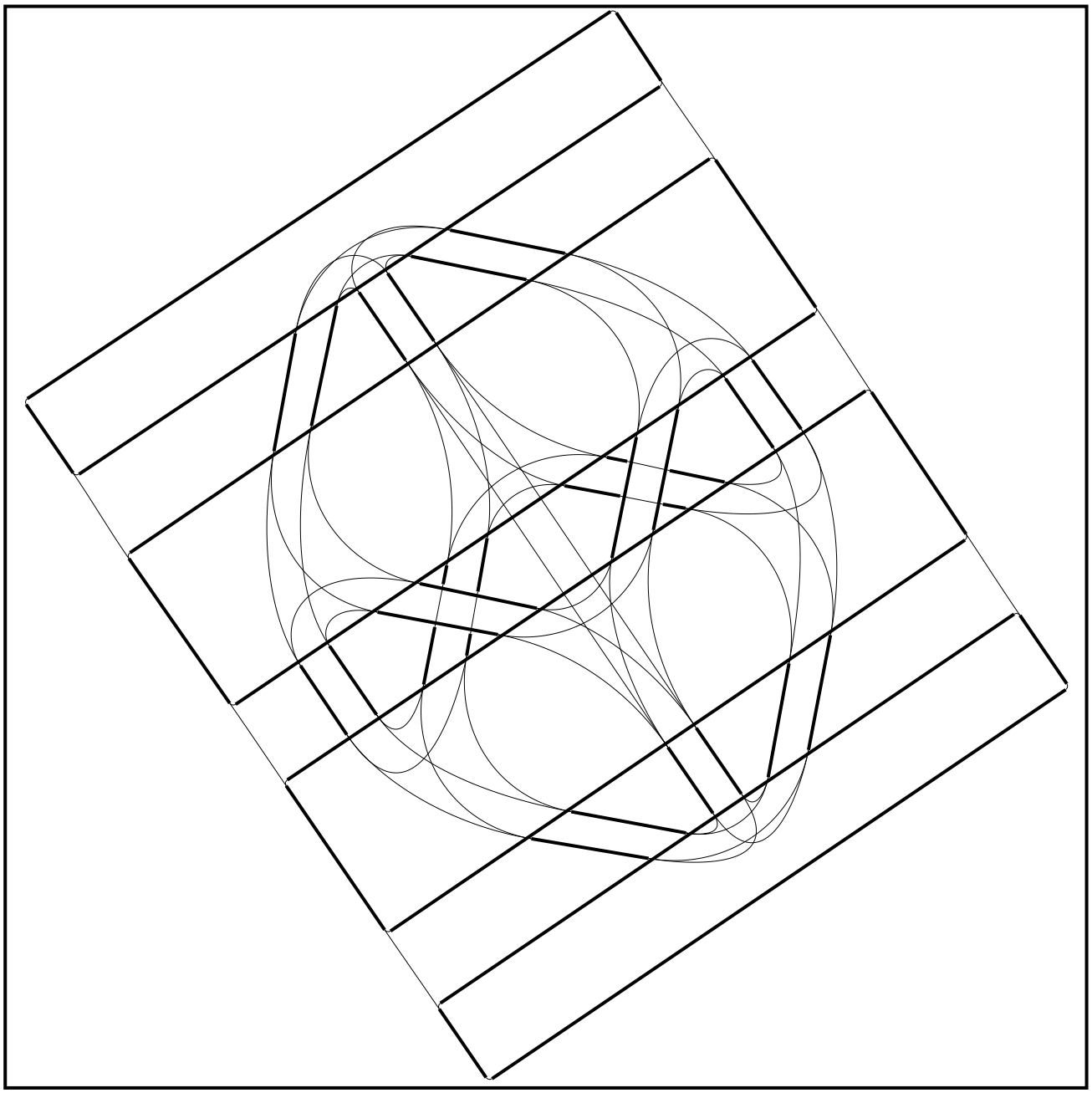

Figure 18: A subset of potential completions, selected solely on geometric criteria.

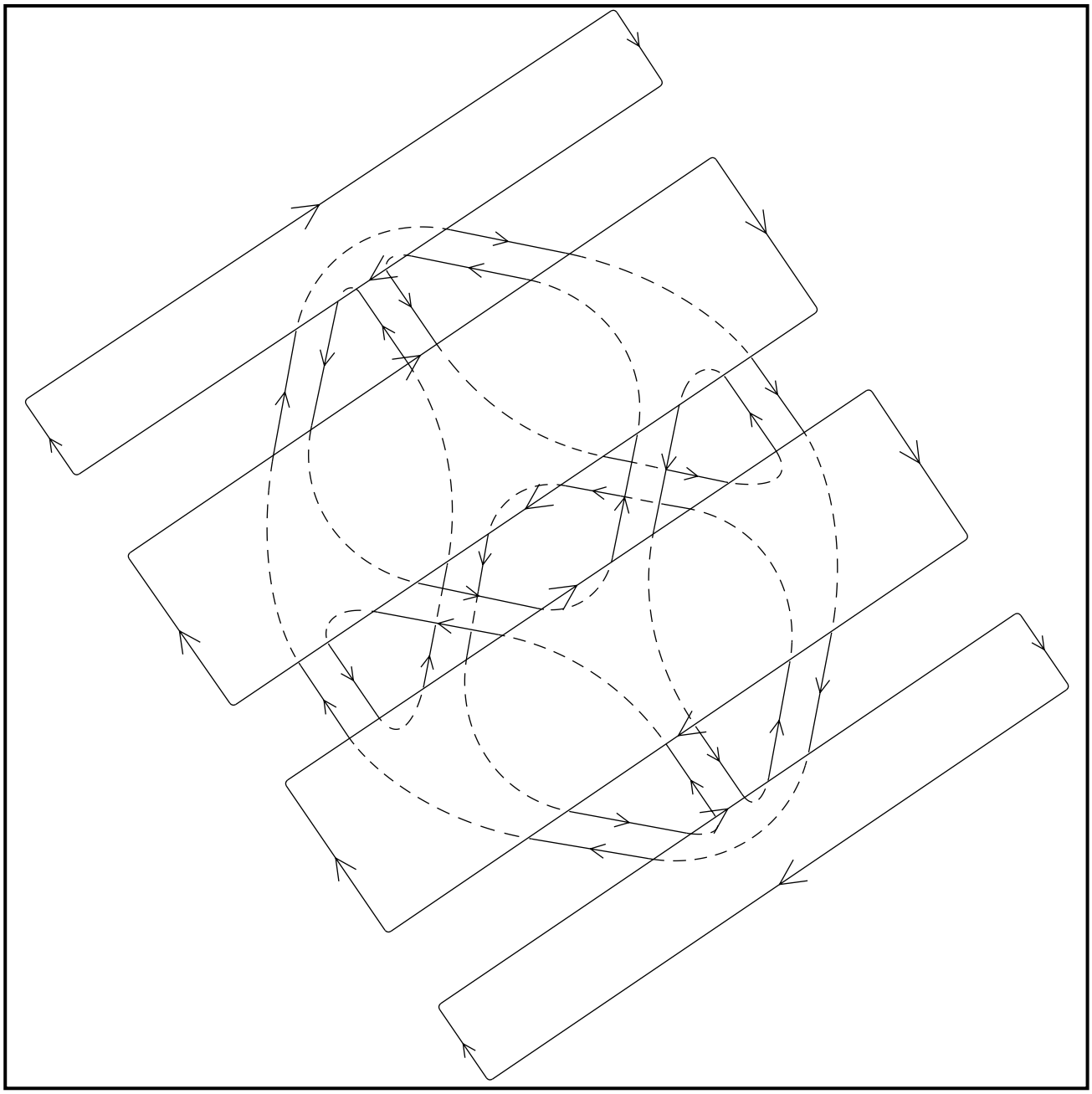

Figure 19: The completed cube, which in this case, is uniquely determined by topological requirements.

4. As a consequence of interpolation, the subsets of boundary components which together define individual surfaces are identified. This is a grouping problem in its own right.

Figure 17 depicts one of Kanizsa'a partially occluded cubes, a difficult completion problem which the experimental system is capable of solving. This example, though not flat, was selected because it illustrates the disambiguating power afforded by topological requirements alone. Potential boundary completions, represented by cubic bezier splines of least energy, are shown in Figure 18. This subset of completions was selected solely on geometric criteria. In this instance, the final organization (shown in Figure 19) is uniquely determined by the requirement that the completed boundaries define a topologically valid flat scene. A detailed description of an improved version of the experimental system is forthcoming.

## Acknowledgements

This work would not have been possible without the help of Ivan Mirkovic, of the Dept. of Mathematics, who proposed the construction used in the proof, and who patiently reviewed the entire argument. Thanks are also due to Steve Zucker, for originally encouraging me to investigate the problem at this level, and to Harpreet Sawhney and Allen Hanson for many valuable discussions.

## References

[1] Barrow, H.G. and J.M. Tenenbaum, Recovering Instrinsic Scene Characteristics from Images, *Computer Vision Systems*, A.R. Hanson and E.M. Riseman (eds.), Academic Press, New York, 1978.

[2] Griffiths, H.B., *Surfaces*, 2nd Ed., Cambridge University Press, London/New York, 1981.

[3] Henle, M., *A Combinatorial Introduction to Topology*, W.H. Freeman and Co., San Francisco, Cal., 1979.

[4] Horn, B.K.P., Obtaining Shape from Shading Information, *The Psychology of Computer Vision*, P.H. Winston (ed.), McGraw-Hill, New York, 1975.

[5] Kanizsa, G., *Organization in Vision*, Praeger, New York, 1979.

[6] Kauffman, L., *On Knots*, Princeton University Press, Princeton, New Jersey, 1987.

[7] Koenderink, J.J., The Shape of Smooth Objects and the Way Contours End, *Natural Computation*, W. Richards (ed.), MIT Press, Cambridge, Mass., 1988.

[8] Marr, D., *Vision*, W.H. Freeman and Co., San Francisco, Cal., 1982.

[9] Nitzberg, M. and D. Mumford, The 2.1-D Sketch, *Proc. of the 3rd Intl. Conf. on Computer Vision*, Osaka, Japan, 1990.

[10] Williams, L.R., Perceptual Organization of Occluding Contours, *Proc. of the 3rd Intl. Conf. on Computer Vision*, Osaka, Japan, 1990.