PERCEPTUAL COMPLETION OF OCCLUDED SURFACES

A Dissertation Presented

by

LANCE R. WILLIAMS

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 1994

Department of Computer Science

© Copyright by LANCE R. WILLIAMS 1994

All Rights Reserved

PERCEPTUAL COMPLETION OF OCCLUDED SURFACES

A Dissertation Presented

by

LANCE R. WILLIAMS

Approved as to style and content by:

Allen R. Hanson, Chair of Committee

Ivan Mirkovic, Member

Robin J. Popplestone, Member

Edward M. Riseman, Member

Steven W. Zucker, Member

W. Richards Adrion, Department Head
Department of Computer Science

ACKNOWLEDGMENTS

I consider it a great privilege to have been a member of the computer vision group at the University of Massachusetts these past eight years. My thesis advisor, Al Hanson, patiently gave me what I needed to succeed: freedom, opportunity, guidance and encouragement. For this, I am deeply grateful. With Al, Ed Riseman built a computer vision group and brought together a diverse group of talented people. This group thrived and prospered because of their vision and hard work. Among its current members and alumni are many good friends—some of whom may yet revolutionize the field. If I have a Ph.D. today, it is probably because Ed and Al hoped I would do something really useful. And I may yet.

I also wish to acknowledge the contributions of my other committee members, Ivan Mirkovic, Robin Popplestone and Steve Zucker. It was only with Ivan's help that the proof of the sufficiency of the labeling scheme was accomplished. Thank you, Ivan. Robin was a continual source of inspiration—he is the only full professor of computer science I know who has actually written a computer program in the last week. I hope I remain as willing to test theory with experiment as Robin has. I first met Steve in a seafood restaurant in Florida in 1988. Before I could finish my gumbo, he had explained the structure and function of the visual cortex. It was Steve who suggested the need for Chapter 2, and more recently, who encouraged me to better connect my work to the main problems of the field.

When I first came to U. Mass., I was greatly influenced by two senior graduate students. P. Anandan, with his outgoing personality, set the intellectual and social tone of the vision group. Michael Boldt shared his philosophy of perceptual organization and encouraged me and others to rebel against the "tyranny of the pixels."

I also wish to thank Harpreet Sawhney, who has a rare combination of passion for the subject and clarity of thought, for his contributions to this thesis. Teddy Kumar, R. Manmatha, and John Oliensis have been good friends and colleagues. I have benefited from dialogue and lively debate in seminars, hallways and in front of monitors with Ross Beveridge, Brian Burns, Bob Collins, Chris Connolly, John Dolan, Bruce Draper, Claude Fennema, Phil Kahn, Inigo Thomas, and Zhongfei Zhang. Thanks also to Laurie Waskiewicz and Janet Turnbull.

Jonathan Lim assisted in many ways, including generating the ray-traced images in Chapter 1, and finessing my Postscript output with various nawk hacks. Robert Heller, who answers more dumb questions in a day than anyone should have to in a lifetime, won't read this, but I thank him also.

I have had many excellent teachers over the years. In particular, I would like to thank Philip Buttermore and Charles Landis, who are truly credits to their profession, and who were especially kind to me. Thanks also to Bill Sakoda, who introduced me to vision, and who remains a good friend.

I wish to thank my new colleagues at NEC Research Institute, Ingemar Cox and David Jacobs, for comments which improved this dissertation, and for giving me the opportunity to investigate the unsolved questions which it raises.

On a different note, I would like to thank the past and present members of the Graduate Employee Organization/District 65 U.A.W. for fighting to make higher education in this country accessible to all.

Most of all, I would like to thank my mother and father, and my grandparents, who have consistently encouraged me in all my endeavours, and who have given me much love.

ABSTRACT

PERCEPTUAL COMPLETION OF OCCLUDED SURFACES

FEBRUARY 1994

LANCE R. WILLIAMS

B.S., PENNSYLVANIA STATE UNIVERSITY

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

PH.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Allen R. Hanson

Researchers in computer vision have primarily studied the problem of visual reconstruction of environmental structure that is plainly visible. In this thesis, 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.

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.

In this thesis, we show that the boundaries of the anterior surfaces can be represented in viewer-centered coordinates as a *labeled knot diagram*. The interior neighborhoods of the anterior surfaces are explicitly represented by a combinatorial model called a *paneling*, which is produced from a labeled knot diagram by means of a straightforward construction. Conventional "shape-from" methods formulated as variational problems and defined over image neighborhoods can be applied to the neighborhoods of the paneling equally well.

The labeling scheme and paneling construction provide a solid theoretical foundation for a working experimental system which computes surface representations from illusory contour displays, including well known figures from the visual psychology literature. 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.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

"Vision is the art of seeing things invisible." — Jonathan Swift

The subject of this thesis is perceptual completion of occluded surfaces. In human vision, perceptual completion processes "fill in the gaps" in the *visible world* and in doing so, construct the far richer experience we call the *visual world*. Because of perceptual completion, the visual world is populated by "objects" with boundaries which are independent of viewpoint. And while the visible world consists solely of those surfaces which reflect light onto the retina, the visual world contains objects with backsides and solid interiors.

The existence of a visual world, distinct from the "real world," but "going beyond" that which is plainly visible is familiar to philosophers of human perception. See Gibson[13] or Kanizsa[29] for especially lucid accounts. The field of computer vision, in contrast, has for all practical purposes ignored the problem of visual reconstruction of environmental structure that is not in plain sight. Because the problem of perceptual completion is largely unrecognized, its importance can not be established by citing the computer vision literature, but instead derives from the phenomenology of human vision. Stated differently, because perceptual completion plays a major role in what human beings call "seeing," it is worth studying. The fact is that almost all problems studied in computer vision today trace their ancestory to this same source.

This is because human vision still provides the ultimate measure of what is and is not possible to compute from images.

This chapter begins with a review of the salient features of the "standard model" of visual reconstruction. The standard model is the now familiar theory that the purpose of early human visual processing is to compute a viewer-centered representation of visible surfaces which is most often called the $2\frac{1}{2}$-*D sketch*[38]. However, the position taken in this thesis is that by considering only the visible surfaces, the standard model fails to address the most difficult part of the visual reconstruction problem, which is deducing the topology of three-dimensional scene structure. In the standard model, the topology of the reconstructed scene is trivially determined by the assumption that the imaging process maps visible surface neighborhoods[1] to image neighborhoods in one-to-one fashion.

In this thesis, 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. The result is that the topology of three-dimensional scene structure can no longer be taken for granted, but must instead be inferred from the fragmentary evidence provided by image contours. Solving this perceptual organization problem is the subject of this thesis.

[1] A subset of a topological space is a neighborhood of a point p if and only if it contains every other point of that space which lies within a sufficiently small ball centered on p.

1.1 Visual Reconstruction

The input to the human visual system is an array of brightnesses originating in light reflected from environmental surfaces. Using only this information, human beings exhibit a nearly limitless range of behavior requiring detailed knowledge of the physical structure of the environment. These range from navigating unfamiliar terrain to recognizing the potential utility of completely novel objects as tools. In short, vision underlies much of what we call intelligence. 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 environment we derive from it, necessitates the existence of intermediate representations of some sort. These intermediate representations presumably encode information about the environment useful in many different problem solving contexts. Adherents of this theory maintain that computing these general purpose intermediate representations is the purpose of human early visual processing.

Any meaningful description of environmental structure requires two components. The first is the topological component, which is a specification of a set of neighborhoods (i.e. a topology). The second component describes how those neighborhoods are embedded in three-space. Without a description of the neighborhoods, it is meaningless to talk about shape. For example, even computing something as simple as an orientation requires an appropriately defined neighborhood. This is why the phrase "structure from motion," as commonly used, is somewhat of an exaggeration. The "structure" consists of a set of points in three-space and is therefore trivial from a topological standpoint. Since the topological component of structure is the set of environmental neighborhoods, it is worth asking: What kind of neighborhoods are important?

A solid is a topological entity possessing a connected set of neighborhoods each of which can be thought of as a tiny ball. Probably because of the atomic theory of matter, computer vision researchers generally assume that the world contains only solid objects. In fact, a large fraction of the light falling on the retina is reflected from structure in the environment with neighborhoods more naturally thought of as non-solid. For example, the leaves of a tree, a pair of pants, a piece of paper and blades of grass are generally thought of as two dimensional, with neighborhoods that are like tiny discs, not balls. Similiarly, telephone wires, paperclips and pine needles can be thought of as one dimensional.

More complex neighborhoods are equally natural to think about. Consider the spine of a book, where all the book's pages meet in a neighborhood that is like neither a ball nor a disc—or the neighborhood formed by a pin stuck in a pin cushion. The point is that human beings are capable of imagining "compound" neighborhoods formed from various combinations of the "elementary" neighborhoods. This seems to hold for all manner of combinations up to dimension three.

The neighborhood structure is of course dependent on spatial scale. At one scale, the neighborhoods of a crumpled piece of newspaper are two dimensional, and disc-like. At a smaller scale, since the paper has non-zero thickness, the disc-like neighborhoods are revealed to be masses of fibers... At the largest scale, the crumpled piece of paper assumes the character of a solid object, with ball-like neighborhoods. This is how we think about it when we toss it in the trash after crumpling it.

The important thing is that we think about these things in all of these ways. English is filled with words which are only meaningful for objects possessing a particular topological character. Whether or not it makes sense to "crumple" something or "crush" it depends upon our judgement of whether it is more like a piece of newspaper or more like a boulder. Things with disc-like neighborhoods are crumpled, things

with ball-like neighborhoods are crushed. It is only logical that language develops in ways which are consistent with the human experience of the visual world.

By far the most difficult part of visual reconstruction is deducing the set of environmental neighborhoods. I maintain that this is the fundamental problem of perceptual organization. In this respect, this thesis goes quite a bit beyond what computer vision typically regards to be the purpose of perceptual organization: Grouping of image measurements of one kind or another into collections likely to belong to individual objects.

Ten years ago Witkin and Tenenbaum[60] wrote what many consider to be a kind of "P.O." manifesto. In "On the Role of Structure in Vision" they argued that the purpose of perceptual organization is to provide an explicit representation of "primitive image structure" embodying geometric coincidence which "demands explanation" from higher-level cognitive processes. When I say that the fundamental problem of perceptual organization is deducing the set of environmental neighborhoods, it is because I assume these neighborhoods have a manifestation in primitive structure. That is, they are its original source, and it forms their "precursors" in the reconstruction.

Of course, given such an ambitious goal, it is natural to ask what fraction of the environmental neighborhoods can realistically be reconstructed given only image information. At this point it is customary to mention the dimensionality reduction associated with projection of a three-dimensional world onto a two-dimensional retina, and how reconstruction is impossible without making additional assumptions about environmental structure. This is certainly true. But the point taken here is that the assumptions underlying the "standard model" of visual reconstruction are so strong that the problem of perceptual organization is completely trivialized.

The standard model of visual reconstruction can be viewed as an attempt to invert the *image irradiance equation*. The image irradiance equation relates the brightness of an image point to the depth, surface orientation, reflectance and illumination of

a visible surface point in the environment. In theory, a solution of the image irradiance equation would consist of a set of what Barrow and Tenenbaum[3] call *intrinsic images*, each of which represents one of the physical parameters underlying image brightness.

The intrinsic images representing depth and surface orientation together form what Marr[38] called the $2\frac{1}{2}$-*D sketch*. Specifically, for every image point x, y there is assumed to be a depth, $z = f(x, y)$ and a pair of values p and q where:

$$p = \frac{\partial f(x,y)}{\partial x}$$
$$q = \frac{\partial f(x,y)}{\partial y}$$

The values p and q represent the surface orientation of image point x, y as a location in *gradient space*. The unit surface normal of an image point is then given by the equation:

$$\hat{\mathbf{n}} = \frac{1}{\sqrt{1 + p^2 + q^2}}[p, q, -1]^T$$

The methods proposed for computing the $2\frac{1}{2}$-D sketch are generally referred to as "shape-from" methods. The inspiration for these methods is the influential early work of Horn[21] on *shape-from shading*. Surface orientation can be computed directly from image brightness when the direction of illumination is known and surface reflectance is predominantly diffuse. Where these assumptions do not hold, alternative shape-from methods are potentially applicable. For example, in *stereopsis* the binocular disparity between points in images from the left and right eyes is measured to determine depth

directly. These measurements can be interpolated to form a dense representation of depth and surface orientation. Alternatively, optical flow can be used to solve for the equations of motion of a moving sensor, providing estimates of depth equivalent to those provided by stereo. It is also possible to derive depth and surface orientation from texture and focus information.

Many shape-from methods are formulated as problems in the calculus of variations, which leads to relaxation algorithms defined in local image (i.e. surface) neighborhoods. Boundary conditions are either free or simple closed plane curves. Unfortunately, the constraints derived from the image irradiance equation or from stereo are seldom strong enough to determine a unique solution. Therefore, the functional to be minimimized usually consists of two terms, where the first term embodies the constraint derived from image brightness (in the case of shape-from shading) or points of known depth (in the case of stereo), and the second term is a *regularization* term, which introduces a bias for smooth solutions. For example, one of the simplest functionals of this sort is for shape-from shading[23]:

$$\int\int_I [(E(x,y) - R(p,q))^2 + (p_x^2 + p_y^2 + q_x^2 + q_y^2)]\, dx\, dy \tag{1.1}$$

where E is the image brightness function, R is the reflectance map in gradient space and p_x, p_y, q_x and q_y are the first partial derivatives of p and q with respect to x and y. Here $E(x,y) - R(p,q)$ represents the deviation from the ideal brightnesses predicted by the image irradiance equation and $(p_x^2 + p_y^2 + q_x^2 + q_y^2)$ represents the deviation from smoothness.

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 non-visibility of backward facing surface patches) and does not account for non-local occlusion (i.e. non-visibility 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 thesis, we distinguish between the *visible surfaces* and the *anterior surfaces*:

Defn. *visible surfaces - the locus of environmental surface points first incident along the lines of sight.*

Defn. *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 Figure 1.1. 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 1.2 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 Figures 1.3(a) and 1.3(b).

Although the standard model only addresses the problem of reconstructing the visible surfaces, 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. You would also be surprised if you walked around the pool table and discovered that the ball has a flat backside. 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? Specifically, the new information must contradict

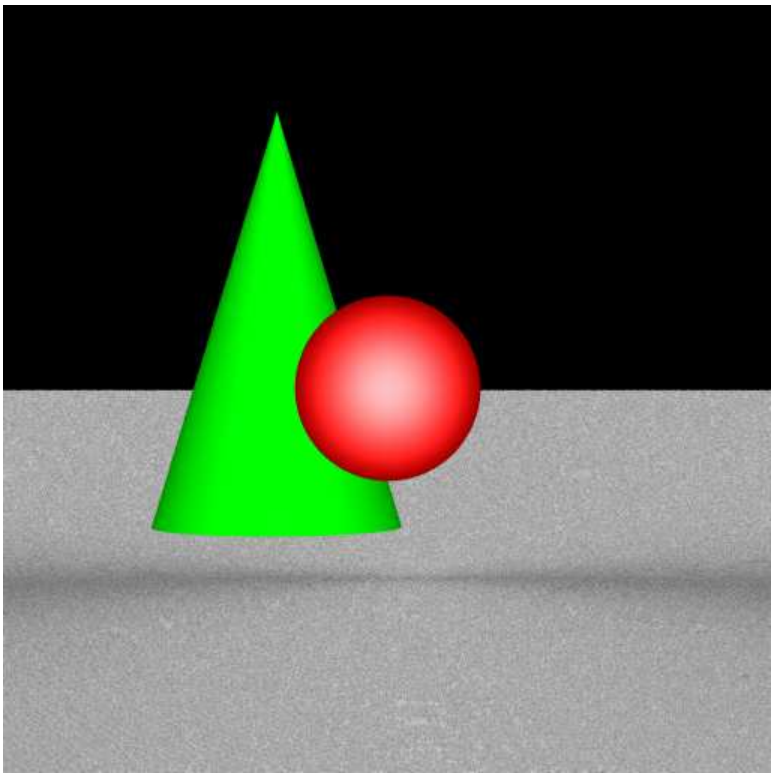

Figure 1.1: A simple ray-traced image of a sphere and a cone. This scene is illuminated by a point source coincident with the location of the viewer, so there are no visible shadows.

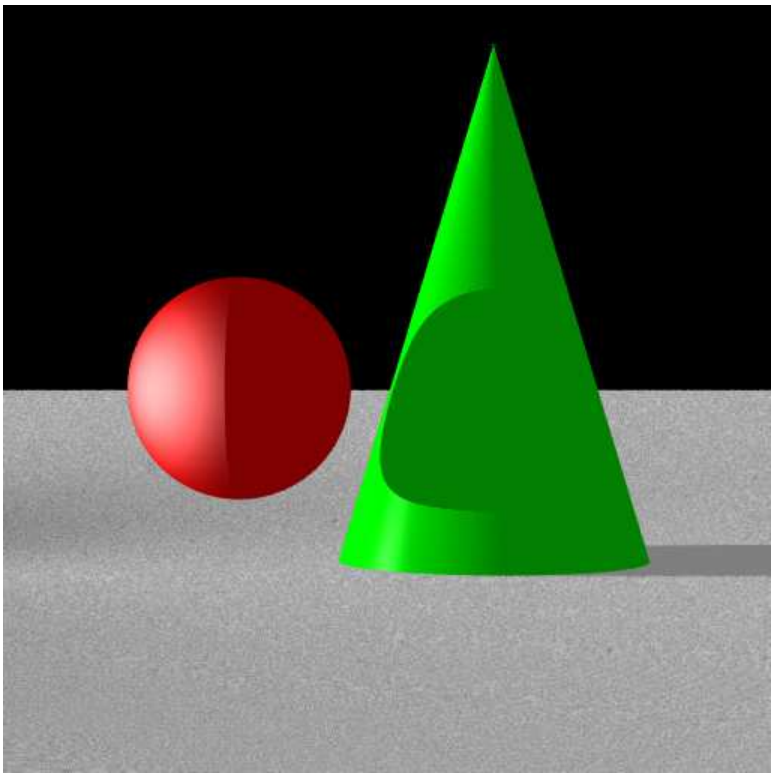

Figure 1.2: A sideview of the same scene with the illumination unchanged. Because of the location of the light source, surfaces not visible from the first viewpoint lie in shadow.

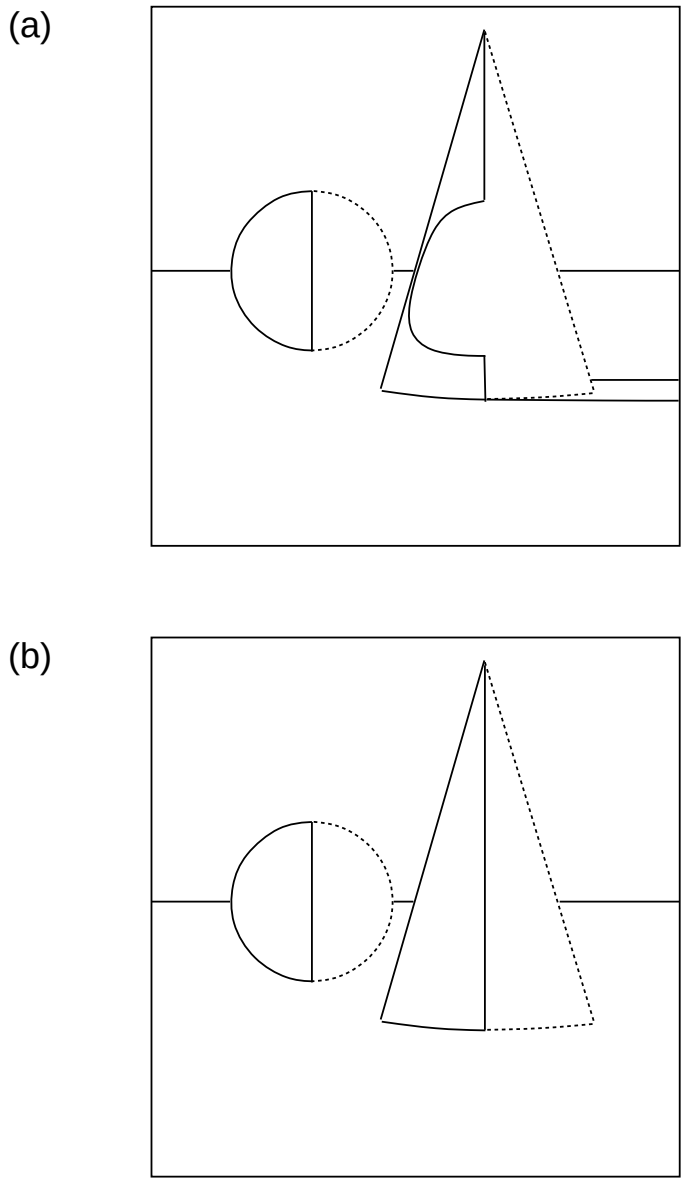

Figure 1.3: Distinguishing visible surfaces from anterior surfaces. The outlines of the visible and anterior surfaces (with respect to the first viewpoint) appear in sideview. (a) The *visible surfaces* are defined as the locus of surface points first incident along the lines of sight. (b) The *anterior surfaces* are 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. The standard model only addresses the problem of reconstructing the visible surfaces.

the assumption that the felt surface of the pool table continues underneath the ball, and that the ball's backside is hemispherical. These are two examples of *perceptual completion*, a collective term for those phenomena which "fill in" our experience of the visual world.

Examples of perceptual completion abound, and include the well known phenomena of *illusory contours*[2] which are epitomized by displays such as the Kanizsa triangle[29] (Figure 1.4). In Kanizsa's figure, a white triangle appears to partially occlude three black discs and a second triangle rendered in outline. Although the completion of the black discs is perceived, this perception is not manifest as a brightness change. In contrast, the completion of the white triangle is characterized by a significant difference in *apparent brightness* between the triangle and its background.[3] Of course neither the triangle nor the discs are objectively present. All are the products of a perceptual completion process.

Perhaps it is not surprising that occluded surfaces are ignored by the standard model, after all, surface patches which aren't visible don't contribute to image brightness. However, the most important reason is that the image irradiance equation 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 local iterative algorithms can operate on image

[2]Also called *subjective contours*.

[3]The terms *modal* and *amodal* are used by Kanizsa to distinguish between instances of perceptual completion which are accompanied by change in apparent brightness from instances which are not. Accordingly, Kanizsa describes the completion of the discs in the Kanizsa triangle figure as amodal, and the completion of the illusory triangle itself as modal.

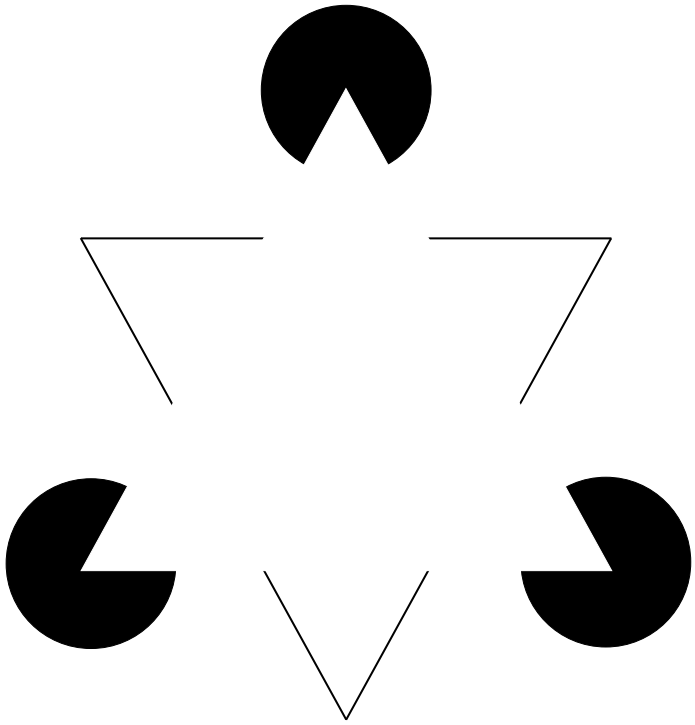

Figure 1.4: The Kanizsa triangle.

neighborhoods and reconstruct surface neighborhoods: They are assumed to be in one-to-one correspondence.[4]

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.

[4]It is important to distinguish between discontinuities in the visual mapping and discontinuities in surface orientation and depth. Marr[38] proposed that one-dimensional discontinuities in surface orientation and depth are explicitly represented in the $2\frac{1}{2}$-D sketch. Recently, Blake and Zisserman[5] have developed a method by which these discontinuities can be identified during reconstruction, so that smoothness assumptions have appropriate scope. But this addition does not change the fundamental assumption about the visual mapping (i.e. that it is a homeomorphism). One-dimensional discontinuities in surface orientation and depth are simply specialized neighborhoods. Identifying these neighborhoods in the course of reconstruction, while an important contribution, does not fundamentally alter the fact that image neighborhoods are assumed to be in one-to-one correspondence with visible surface neighborhoods.

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? To say that this problem is hard is an understatement. 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.[5]

Although the phenomenological evidence for perceptual completion in human vision is unassailable[29], one might argue that certain forms of interpolation go beyond what is reasonable to represent and compute in a viewer-centered representation.[6] Certainly, the phrase "perceptual completion" could be used to describe the process by which an object-centered, three-dimensional representation of surfaces is derived from the $2\frac{1}{2}$-D sketch. This argument has merit to the extent that reconstruction of anterior and posterior surfaces would result in a description of the environment that is independent of viewpoint. Therefore, the question is not whether perceptual completion is an important and valid phenomenon, rather it is which completion problems are naturally formulated in viewer-centered coordinates. Stated differently, what forms of completion are native to the $2\frac{1}{2}$-D sketch?

Marr[38] was curiously contradictory on this point. While suggesting that "continuation" (i.e. completion) processes operate within the $2\frac{1}{2}$-D sketch, and even sug-

[5] In a recent article, Barrow and Tenenbaum[4] confirm this. They state that "Our 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..."

[6] A representation using a two-dimensional coordinate system registered with the image.

gesting that "the viewer-centered representation of surfaces may be capable of representing more than one surface at once," he also states that "This formulation [the $2\frac{1}{2}$-D sketch] avoids all the difficulties associated with the terms *figure* and *ground*, *region* and *object* —the difficulties inherent in the image segmentation approach..." Very clearly, it is not possible to have it both ways. If one allows more than one surface at a point, one must accept the "can of worms" called "segmentation" as well.

In this thesis, 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 but a subset. In support of this claim, in Chapter 2, a method for representing the anterior surfaces of smooth manifold-solids in viewer-centered coordinates is introduced. The assumption of global homeomorphism between visible surface neighborhoods and image neighborhoods (i.e. embedding), is generalized to the assumption of local homeomorphism between anterior surface neighborhoods and image neighborhoods, or *immersion*. The difference in the two assumptions is shown diagrammatically in Figure 1.5. For scenes composed of smooth manifold-solids, immersion models the visual mapping of anterior surface neighborhoods to image neighborhoods almost exactly.[7] This allows a very good approximation of the complete anterior half of a scene composed of smooth manifold-solids to be represented in image coordinates.

Before the neighborhoods of the anterior surfaces can be explicitly represented, the complete image trace of their boundaries must be deduced. Where boundaries are not occluded and surface reflectance contrasts with that of the background, the boundaries will be marked by luminance change in the image. However, where bound-

[7]Local homeomorphism is violated only at points where the direction of the contour generator coincides with the viewing direction (i.e. at *cusps*[33]).

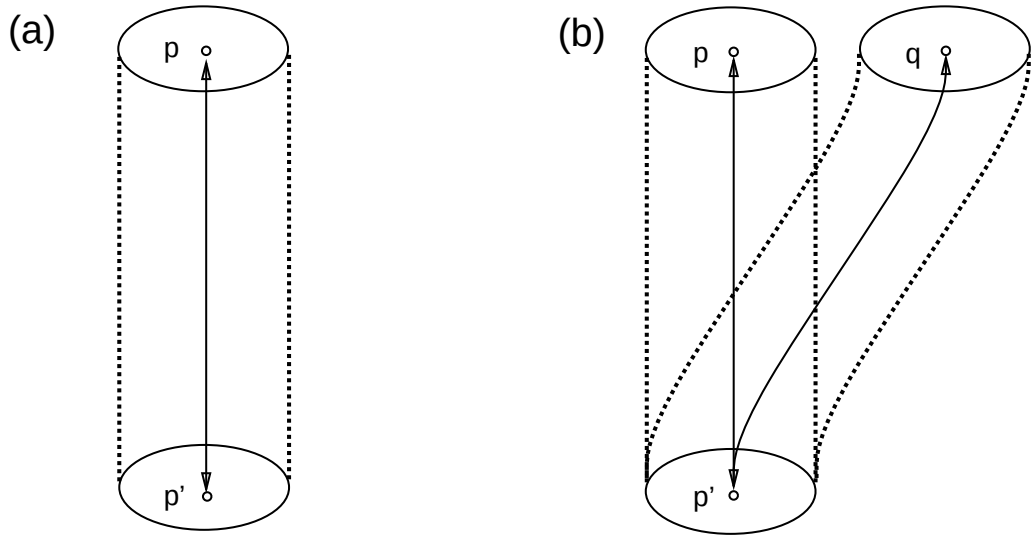

Figure 1.5: The difference between embedding and immersion. (a) An embedding is a global homeomorphism between visible surface neighborhoods and image neighborhoods first incident along the lines of sight. (b) An immersion is a local homeomorphism between anterior surface neighborhoods and image neighborhoods. This allows multiple anterior surface neighborhoods (i.e. of p and q) to project to the same image neighborhood (i.e. of p').

aries are occluded, or where surface reflectance matches background reflectance, there will be no detectable change in image brightness.

In this thesis, *figural completion* is defined as perceptual completion of the boundaries of the anterior surfaces. We show that the boundaries of the anterior surfaces can be represented in viewer-centered coordinates as a *labeled knot diagram.* A labeled knot diagram is a set of closed and oriented plane curves satisfying the labeling scheme depicted in Figure 1.6. 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. In Chapter 2, it is shown that labeled knot diagrams (as opposed to arbitrary closed plane curves) always define unique and topologically valid surfaces. The labeled knot diagram is therefore a boundary representation. Consequently, completion of the surface boundary is tantamount to completion of the surface interior. A labeled knot diagram representing the boundaries of the anterior surfaces of the ray-traced image of the sphere and cone is shown in Figure 1.7.

The neighborhoods of the interior are explicitly represented by a combinatorial model called a *paneling.*[8] The paneling is produced by applying a straightforward procedure called the *paneling construction* to a labeled knot diagram representing the boundaries of the anterior surfaces. We note that any shape-from method formulated as a variational problem and defined over image neighborhoods can be applied to the neighborhoods of the paneling equally well. In this scheme, both the constraint and smoothness terms of the shape-from functional would be applied to the visible

[8]The term "paneling" is used by Griffiths[14] in his informal but very accessible account of the topology of surfaces. Roughly speaking, a paneling is a model of a surface consisting of a set of paper *panels* taped together in prescribed ways. Panelings will be described in greater detail in the next chapter. More formal treatments of the subject (e.g. [18]) employ *triangulations*, which are panelings satisfying other criteria (e.g. all panels must be three-sided).

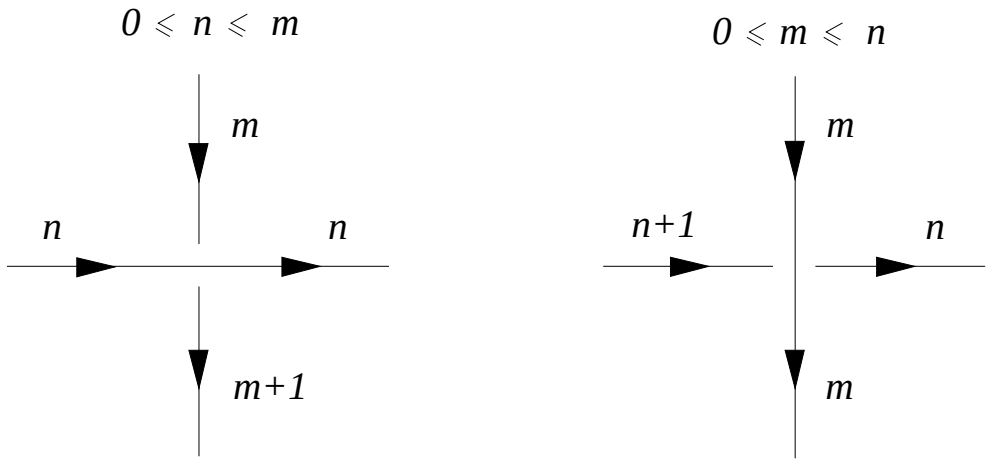

Figure 1.6: A boundary labeling scheme. The image of the surface lies to the right when the image of its boundary is traversed in the direction of its orientation. Each boundary point is also assigned an integer value equal to the number of surfaces lying between the boundary and its projected image (i.e. its *depth*). 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.

neighborhoods of the paneling, while the smoothness term alone would be applied to the occluded neighborhoods. A paneling representing the neighborhoods of the anterior surfaces of the image of the sphere and cone is depicted in Figure 1.8.

1.2 Visual Recognition

Although "visual recognition" in human beings is tremendously subtle and likely uses large amounts of world knowledge in ways which are not completely understood, computer vision has made some impressive initial progress in solving more limited recognition problems. Visual recognition (in this more limited sense) is often understood to mean correspondence between components of a stored geometric model and contours in an image, together with knowledge of the object's pose, or spatial orientation relative to the camera[47, 37]. Of course, the fundamental difficulty in establishing such a correspondence is the combinatorially prohibitive number of possible

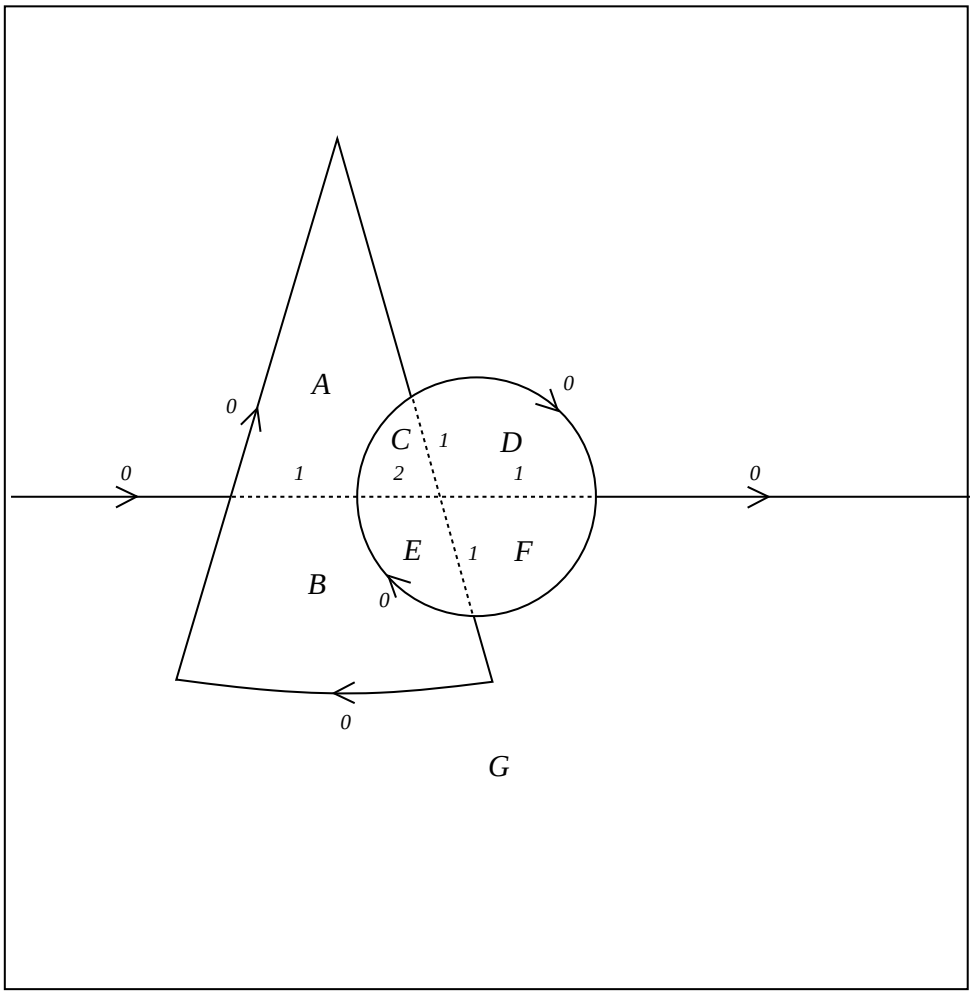

Figure 1.7: Labeled knot diagram representing boundaries of anterior surfaces. This figure depicts the output of the figural completion process for the ray-traced image of the sphere and cone.

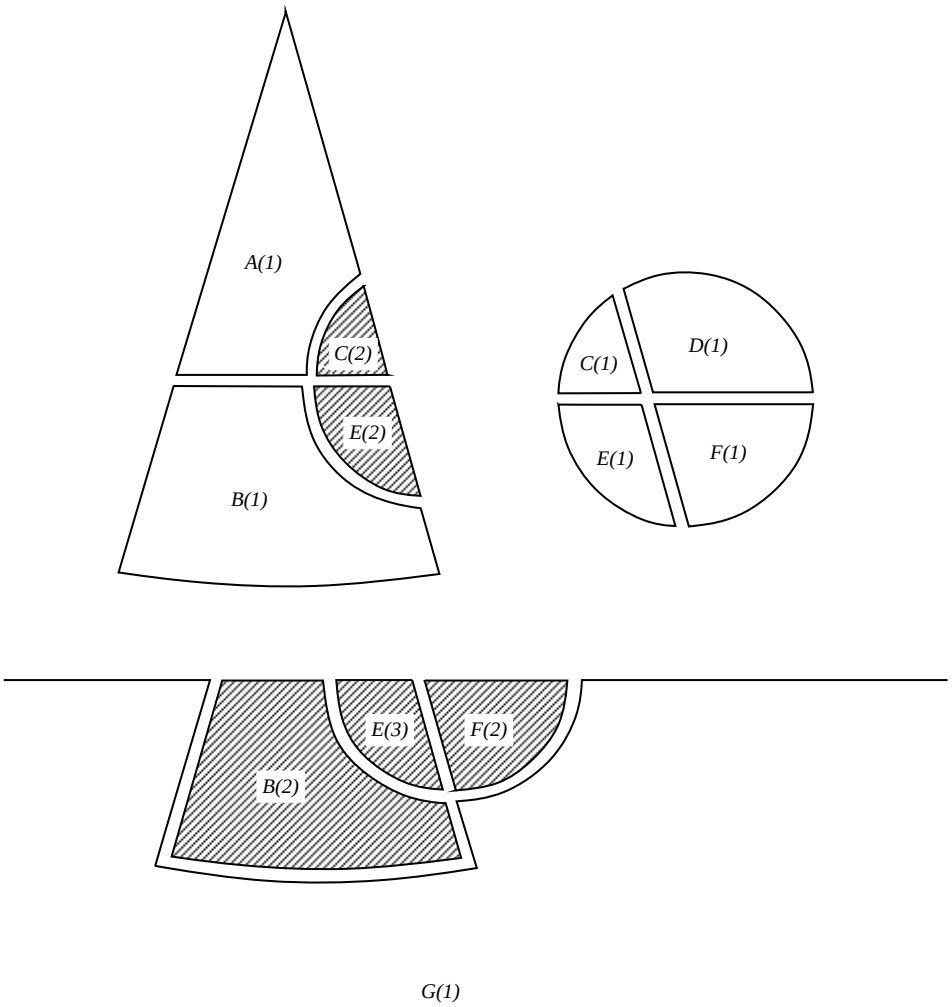

Figure 1.8: Paneling construction applied to the labeled knot diagram. The construction produces a paneling with neighborhoods that approximate those of the anterior surfaces. For clarity's sake, the paneling's three connected components are drawn separately. Any shape-from method formulated as a variational problem and defined over image neighborhoods can be applied to the neighborhoods of the paneling equally well. Both the constraint and smoothness terms of the shape-from functional would be applied to visible neighborhoods of the paneling, while the smoothness term alone would be applied to occluded neighborhoods (shown hatched).

mappings between the components of a stored geometric model and the contours in an image. To make matters worse, the combinatorial problem is often aggravated by extraneous contours originating in background clutter and contours which are absent because of occlusion. It is the opinion of Lowe[37] and Jacobs[28] that the necessary computational savings can only be achieved if grouping processes are used to collect together subsets of contours likely to belong to the same object prior to matching. Kanizsa's cube demonstration is very telling in this respect. The same percentage of the cube is visible in Figure 1.9 and Figure 1.10. Yet the cube is easily recognized in Figure 1.10 and is only recognized with great difficulty in Figure 1.9. The obvious explanation is that the effect of perceptual organization is different in the two figures, and one organization facilitates recognition and the other hinders it. Furthermore, it suggests that perceptual organization processes employ generic knowledge of surfaces to function effectively in spite of background clutter and occlusion. It is likely that a similiar strategy will be useful in computer vision also. Clutter is not peculiar to natural environments; it also occurs in storage bins and among parts lying on conveyer belts. Common practices in digital image processing, such as thresholding edge operator output and the use of a limited number of grey levels, further aggravate the effect of clutter by producing exactly the conditions required for illusory contours.

1.3 Thesis Overview

This thesis approximately follows the "natural computation" methodology advocated by Richards[45] (see Table 1.1). This methodology is based upon Marr's[38] belief that complex information processing systems can only be understood by analyzing them at the three levels of 1) computational theory; 2) algorithm and representation; and 3) implementation. Since the first step of the natural computation methodology requires identifying the "computational goal," Chapter 2 is devoted to

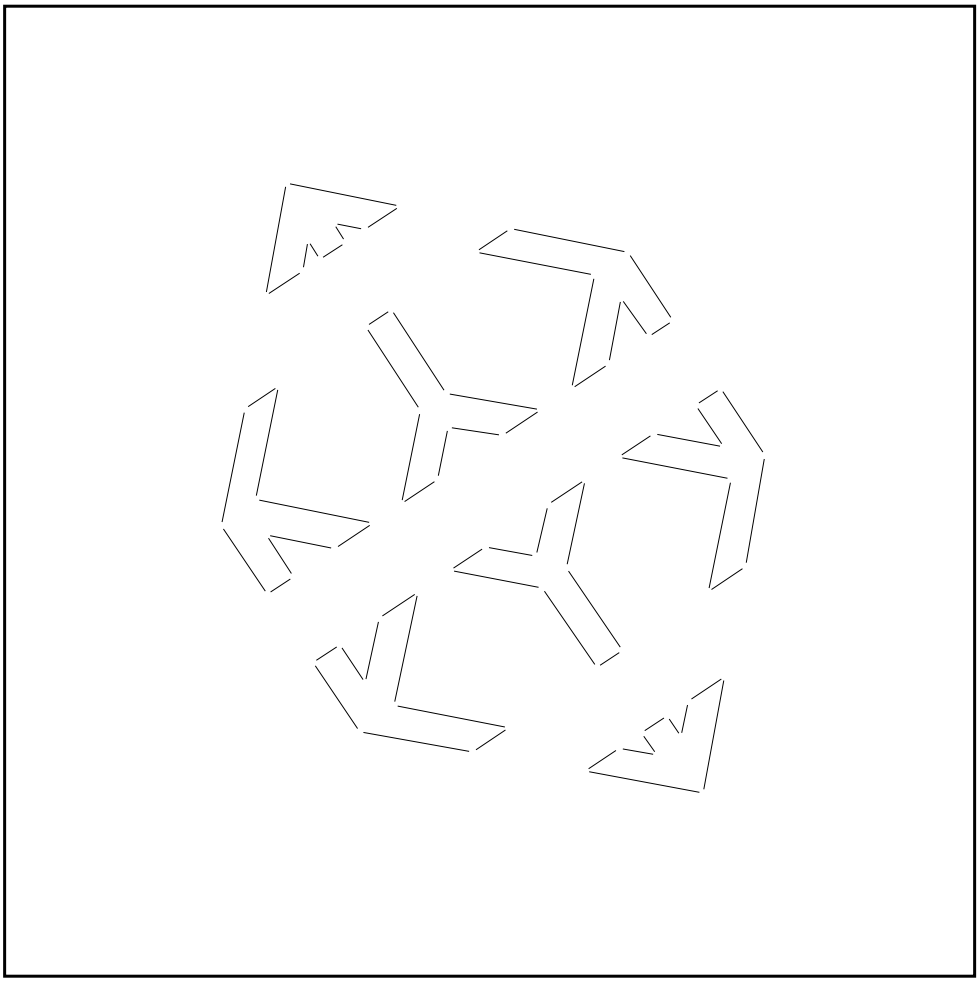

Figure 1.9: An array of odd shapes (from Kanizsa[29]).

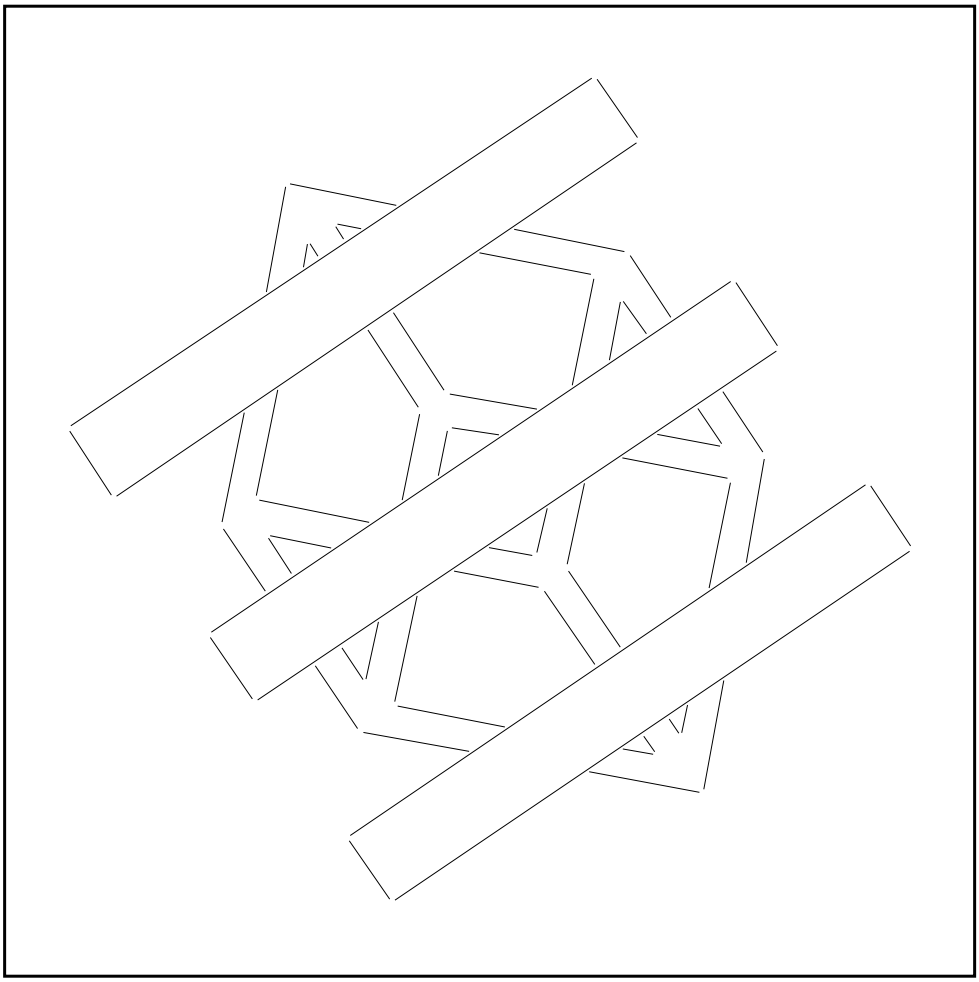

Figure 1.10: A partially visible cube (from Kanizsa[29]).

the introduction of the labeled knot diagram and the paneling construction. Together, these constitute an explicit viewer-centered representation of the neighborhoods of the anterior environmental surfaces.

Chapter 3 begins with a discussion of the natural constraints which define the set of acceptable solutions (i.e. *topological validity* and *stimulus conformity*) and the ambiguities which remain after the constraints are exhausted (i.e. *shape*, *unit* and *depth*). Because the problem is underconstrained, it is proposed that the human visual system resolves the ambiguity in computing a unique completion shape in a way that affords inferential leverage[60] in resolving the other two ambiguities. Under this assumption, the problem of computing a labeled knot diagram is shown to be a graph labeling problem, which is posed as an integer linear program.

Chapter 4 describes an experimental implementation of the computational theory outlined in Chapter 2 and 3. Our intention was not to realistically model human vision at the "level of algorithm and representation[38]." 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 is demonstrated on a number of illusory contour figures from the visual psychology literature.

In Chapter 5, the problem formulation is revised in light of recent evidence from human vision. The revised model is 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.

Chapter 5 continues by reviewing the observation (of the Gestalt psychologists) that grouping phenomena are sensitive to non-local changes in image context. We argue that numerical relaxation in a locally connected network is consistent with this sensitivity and that this form of computation underlies figural completion phenomena in human vision. This leads to a discussion of *total unimodularity*, a property which

Table 1.1: Natural Computation Approach.

	Natural Computation Approach	**Mathematical Formalism**
Step 1	*Identify the Goals and Givens.* What is the desired representation? What data are available?	Define the domain and range of the function.
Step 2	*Show theoretically how a reliable representation can be computed.* What are the natural constraints which will force the unique interpretation?	Find a unique mapping (function).
Step 3	*Design a particular algorithm that correctly interprets the available input information.*	Describe a procedure for computing the function.
Step 4	*Test whether the primate visual system uses the particular algorithm.* (If not, identify which step above is incompatible and reiterate.)	Identify relation between the elements of the procedure and their biological (machine) implementation.

allows a certain class of integer linear programs to be solved in this fashion. We demonstrate that the first of the integer linear programs in the revised model is totally unimodular and argue that computational complexity considerations motivate this problem decomposition in human vision.

Chapter 6 summarizes the contribution of this thesis and presents a set of predictions of the computational theory. The thesis concludes with a discussion of directions for future research.

CHAPTER 2

A BOUNDARY REPRESENTATION FOR ANTERIOR SURFACES

In the previous chapter, two strong assumptions of the standard model of visual reconstruction were identified. The first concerned the goal of visual reconstruction, which the standard model assumes is a viewer-centered representation of the visible surfaces. In this thesis, this goal is generalized to include the fraction of the environmental surfaces which are potentially visible (because they are forward facing) but are possibly occluded by intervening surfaces. To distinguish these from the visible surfaces, these were termed the anterior surfaces.

The second assumption of the standard model is the nature of the visual mapping, which is conventionally modeled as an *embedding* (i.e. global homeomorphism) of the *visible surfaces* in the image plane. In this thesis, this assumption is generalized so that the visual mapping is instead modeled as an *immersion* (i.e. local homeomorphism) of the *anterior surfaces* in the image plane.

In this chapter, a new viewer-centered surface representation called a *labeled knot diagram* is introduced. Development of this idea is facilitated by a short review of elementary combinatorial topology, which begins the chapter. In the course of this review, the topological properties shared by surfaces which can be embedded in three-space so that parallel projection can be modeled as an immersion are identified. The set of scenes for which this projection model suffices are termed *anterior scenes*. Next a labeling scheme incorporating a set of necessary constraints on the appearance of

boundaries in anterior scenes is defined. The labeled knot diagram is simply a set of closed contours satisfying this labeling scheme.

A major part of the chapter is devoted to a proof that labeled knot diagrams can serve as boundary representations. Stated differently, we show that every set of closed plane curves satisfying the labeling scheme represents a topologically valid anterior scene. The core of the proof consists of a description of a procedure by which a "paper model" (i.e. a *paneling*) of an anterior scene can be constructed from a labeled knot diagram. The remainder of the proof involves showing that this procedure can always be applied, and that the paneling it produces satisfies the definition of a surface.

2.1 Topology of Surfaces

In topology, surfaces are divided into different families based upon a small number of properties invariant under continuous deformations. To facilitate arguments which appear later in this chapter, it will be useful to identify the topological properties of surfaces which can be embedded in three-space so that parallel projection induces a local homemorphism between surface neighborhoods and image neighborhoods. A mapping of this kind is called an *immersion*. Image neighborhoods where parallel projection cannot be modeled by local homemorphism are said to be *singular*. For arbitrary embeddings of smooth surfaces, the visual mapping induced by parallel projection will be singular only at the *occluding contours*, which are the image of points where a surface is tangent to the viewing direction.

Since singularities occur only where a surface is tangent to the viewing direction, they will never appear in images of surfaces embedded in three-space so that surfaces are nowhere tangent. This leads to the following definition:

Defn. *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 is an immersion.

Combinatorial topology begins with the idea that a paper model of any surface can be constructed by "gluing" together some combination of the edges of a set of *panels*. For the purposes of this discussion, a panel can be envisioned as a surface cut out 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 and a set of *identifications*, which are explicit indications that two edges of equal length are to be glued together (and in which way). These elements together define a *paneling*. Although a model of every surface can be constructed in this fashion, not every paneling represents a surface.

2.1.1 Orientability

Two edges of equal length may be glued, one to another, in two different ways. The ambiguity can be removed by assigning an orientation to the panel's 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 2.1). 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*. We observe that there is no image of a

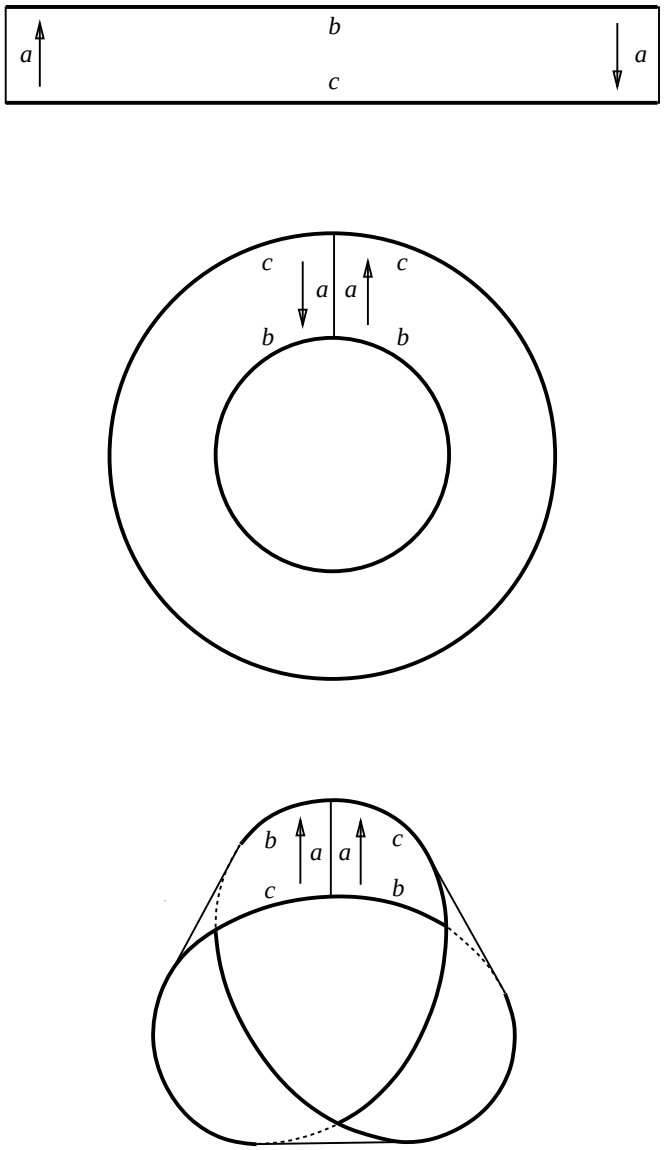

Figure 2.1: *Annulus* vs. *Moebius strip*. A surface can be constructed by identifying two opposite edges of a rectangularly shaped panel and leaving the other two non-identified. Depending on the sense of the identification, the surface which results is either an annulus or a Moebius strip.

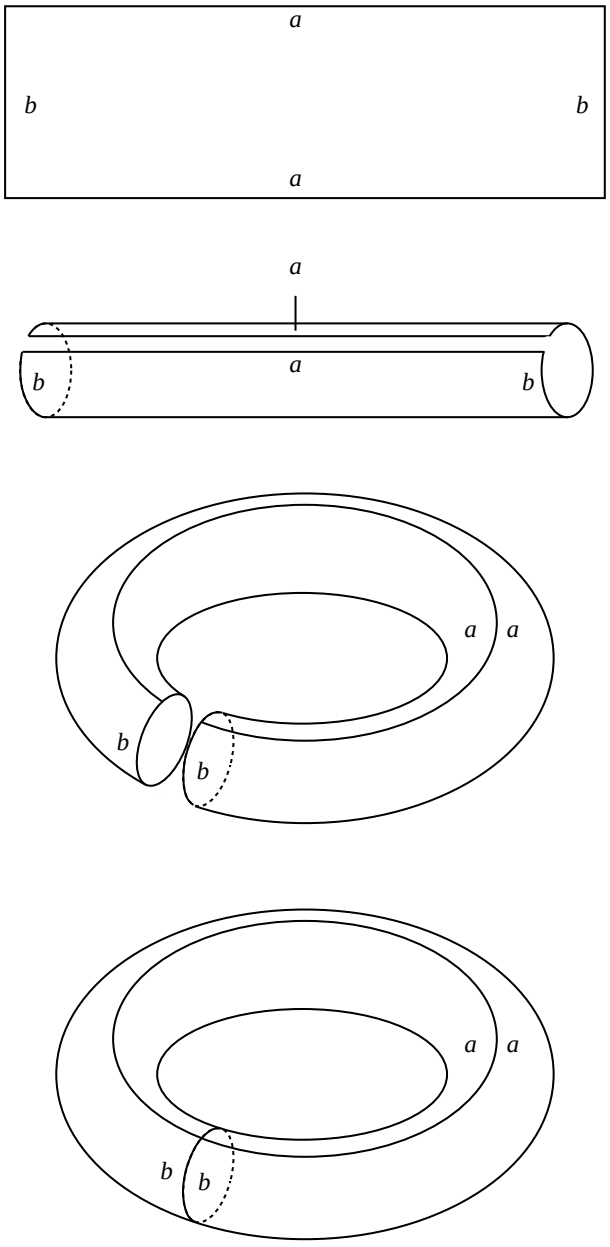

Figure 2.2: The *torus*. The torus is constructed by identifying the opposite edges of a rectangular panel as illustrated (after Hilbert and Cohn-Vossen[19]).

non-orientable surface which does not contain an occluding contour.[1] Non-orientable surfaces cannot be immersed in the plane. For this reason, we adopt the convention (simplifying later arguments) that the orientations of all pairs of identified edges are opposed, and that all surfaces are, consequently, orientable. [2]

2.1.2 Boundaries

Apart from orientability, a surface can also be classified by the number of components in its 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 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 those of the Moebius strip form just one.

A surface without boundary has no non-identified edges in its paneling. For example, a *torus* is constructed by identifying the opposite edges of a rectangular panel as illustrated in Figure 2.2. Surfaces without boundary which are also orientable divide space into two disjoint sets which can be interpreted as the interior and the exterior of a *manifold-solid*. It should be clear that no surface without boundary can be immersed in the plane.[3]

---

[1] This can be appreciated by considering the Moebius strip's Gaussian image. On the Gaussian sphere, there is a great circle of points representing surface orientations tangent to the viewing direction. Since there is a connected path on the Moebius strip joining every point with a point of opposite orientation, its Gaussian image will always cross this great circle.

[2] Equivalently, we could use panels that are painted black on one side and white on the other and adopt the convention that no two panel edges are glued in such a way that black and white meet.

[3] Since the Gaussian image of a surface without boundary occupies the entire Gaussian sphere, parallel projection must create an occluding contour.

2.1.3 Genus

After orientability and number of boundary components, the only remaining topological invariant of surfaces is *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 without 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.

An orientable surface with boundary is created by subtracting one or more discs from a sphere with zero or more handles. The only surfaces which can be embedded in the plane are genus zero surfaces with boundary, which are created by subtracting one or more discs from a sphere. When a genus zero surface is embedded in the plane, its boundary forms a set of closed non-self-intersecting plane curves, or *Jordan curves*. However, all orientable surfaces with boundary, even those which can not be embedded in the plane, have planar immersions. Unfortunately, the boundaries of surfaces with genus greater than zero cannot be represented by Jordan curves, since they always self-intersect in the image plane.

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 2.3. Although

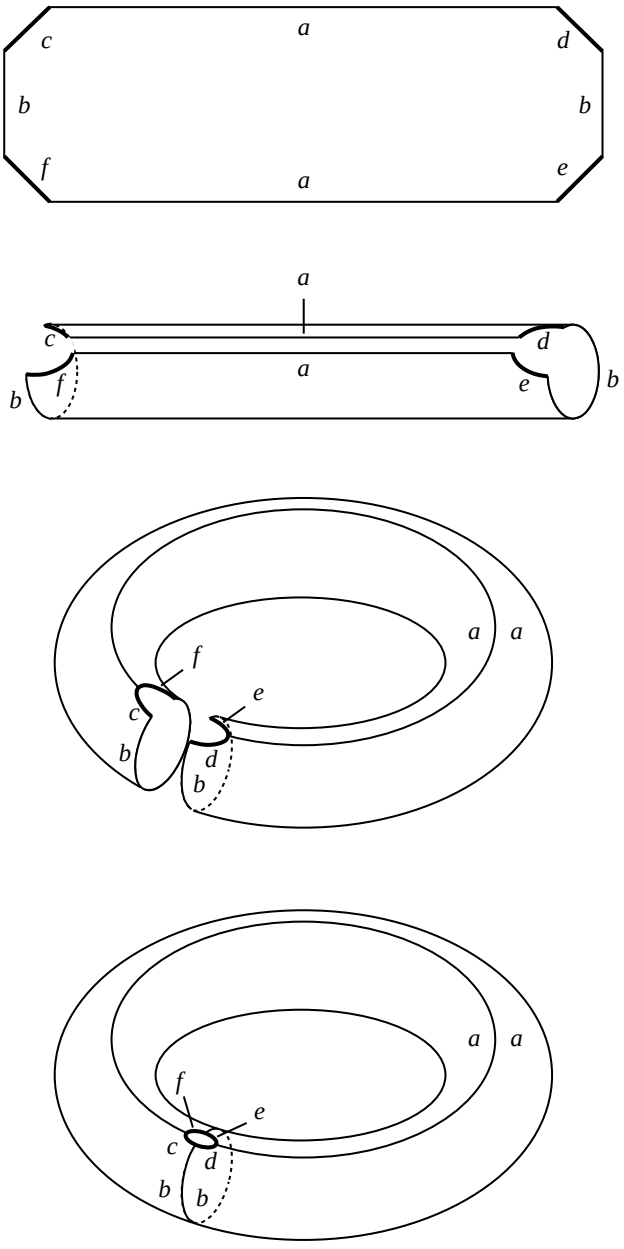

Figure 2.3: The *punctured torus*. This orientable surface with boundary can be constructed from a single panel.

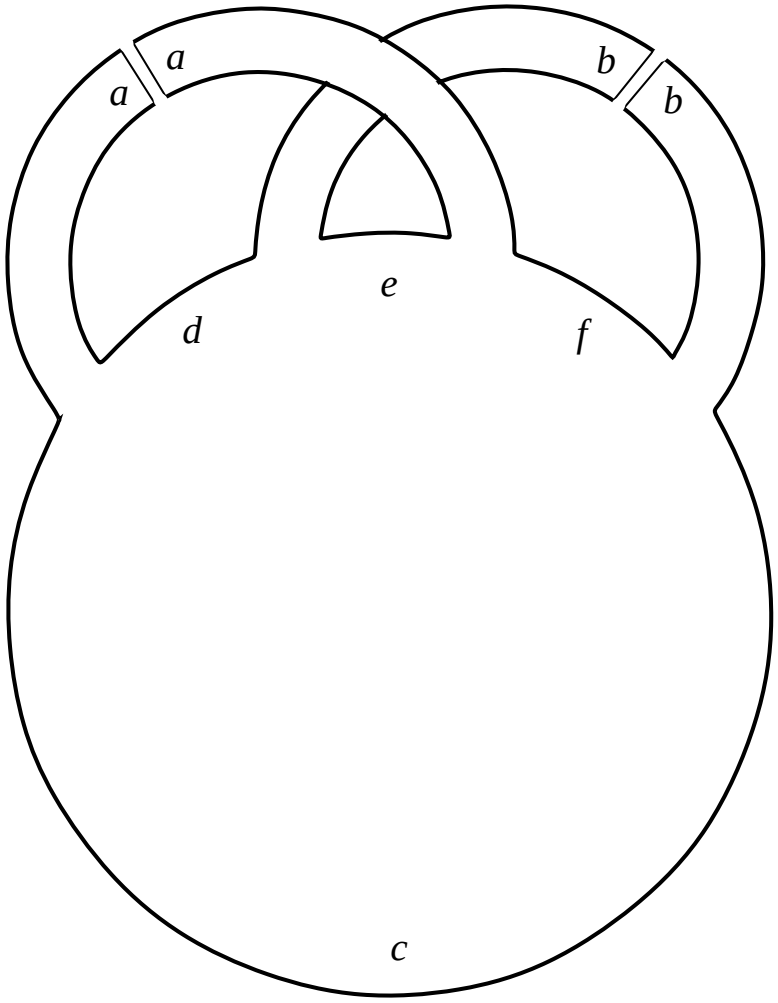

Figure 2.4: The punctured torus can be flattened. Although the punctured torus cannot be embedded in a plane, it has a planar immersion.

an image of a punctured torus embedded like this will always contain occluding contours, it can be readily verified that the essentially flat surface depicted in Figure 2.4 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 2.4; when the boundary of each panel is traversed in the same direction (e.g. clockwise) the edges are encountered in the same order: $adbeafbc$. It is also clear that the identifications of edges a and b indicated in Figure 2.4 correspond exactly with the identifications of edges a and b in Figure 2.3. Furthermore, the cycle of free edges, $cfed$, forms the boundary of both surfaces. From a topological standpoint, the two surfaces are equivalent, even though the image of the boundary is very different in the respective figures.

2.2 Labeled Knot Diagrams

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. A topological circle embedded in three-space is called a *knot*. A generic 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*.

In this thesis, knot diagrams are used to represent surface boundaries. 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 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 anterior scene is *generic*, then the crossings will be the only points of multiplicity two in the projection of the boundary onto the plane:

Defn. *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 viewing direction.*

We observe that boundary depth 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 Figure 2.5. 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. The two possible crossing labelings illustrated in Figure 2.5 have opposite values of writhe. Crossings with opposite writhe are mirror images.

It can be easily verified that the depth labels of the different edges in the labeling scheme accurately describes the effect of occlusion on the boundary depths 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 represent an anterior scene? Is the labeling scheme necessary *and* sufficient?

2.3 Sufficiency Proof

We now prove that a set of closed contours satisfying the labeling scheme illustrated in Figure 2.5 always defines an anterior 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 an anterior scene which projects generically as the labeled knot diagram.

Theorem 2.1 *Every knot diagram satisfying the labeling scheme illustrated below represents a generic view of an anterior scene.*

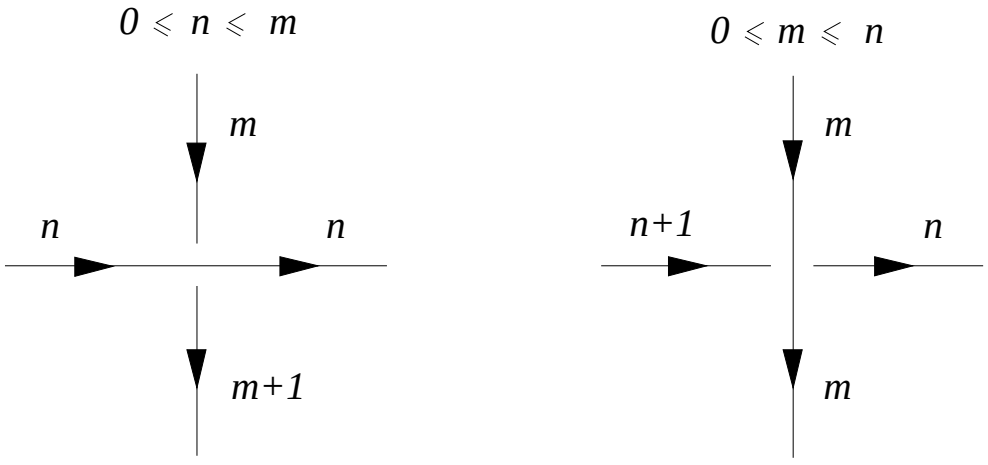

Figure 2.5: A boundary labeling scheme.

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 an anterior scene, then the multiplicity of the projection of interior surface points onto 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 onto image points will be constant within a planar region.

Let A and B be neighboring regions in an oriented knot diagram and let A lie to the right of B. If the labeled knot diagram represents an anterior scene, and if γ_A and γ_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 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

An example of a closed plane curve satisfying the labeling scheme (i.e. a *labeled knot diagram*) is illustrated in Figure 2.6. This plane curve partitions the plane into regions where the number of surface points which project onto a single image point differs by one across region boundaries. Together, these regions define a network representing a system of difference equations (see Figure 2.7). The weights of edges in

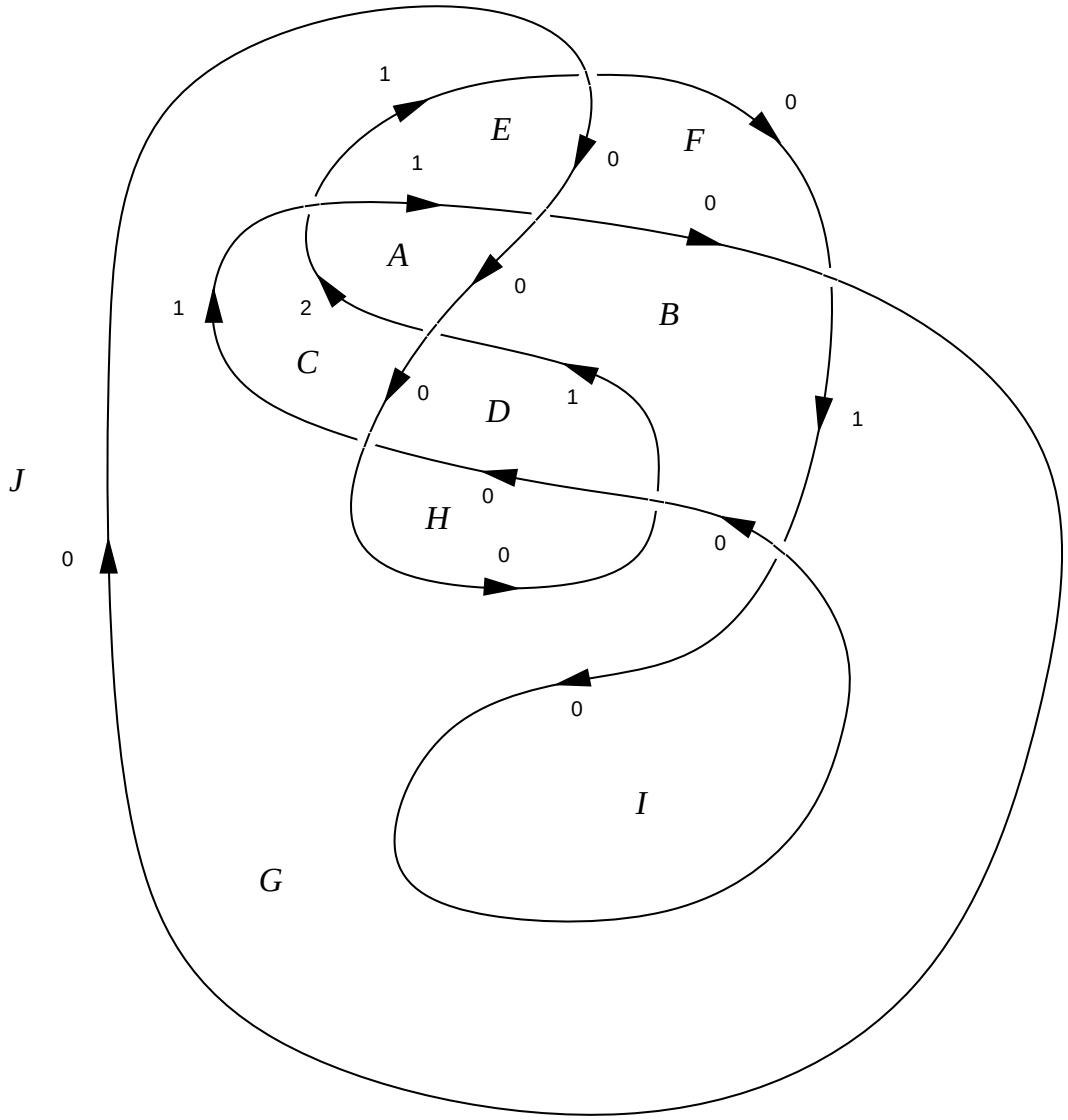

Figure 2.6: A labeled knot diagram. A labeled knot diagram is a set of closed plane curves satisfying the labeling scheme. Because it consists of a set of closed plane curves, a labeled knot diagram partitions the plane into regions.

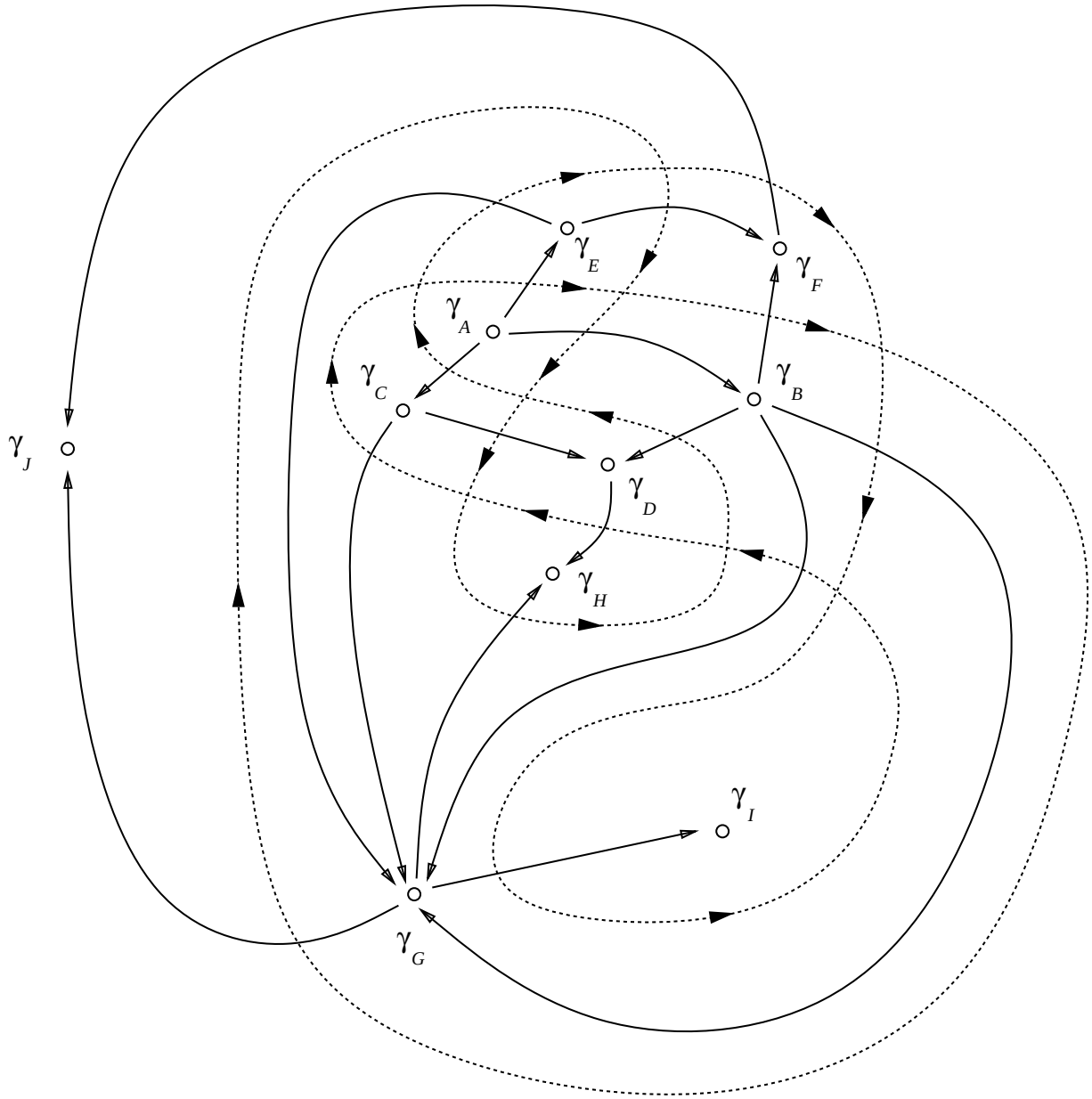

Figure 2.7: Multiplicity network. This network represents a system of difference equations involving the multiplicity of the projection of interior surface points onto adjacent planar regions. The weights of edges in the network are 1 when traversed in the direction of the arrows and -1 when traversed in the opposite direction. The closed plane curve corresponding to the labeled knot diagram of the previous figure is shown dotted.

the network are 1 when traversed in the direction of the arrows and -1 when traversed in the opposite direction. The system of difference equations appears below:

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

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 1 or -1 depending on the direction of traversal)[56]. 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 γ 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 (and there are no other solutions). Since a sufficiently large c can always be found[4] it is sufficient to prove that the sums of the weights around every closed cycle in a network constructed as described equal zero.

[4]Clearly, the simplest solution will use the smallest possible value of c.

We begin by proving the following lemma:

Lemma 2.1 *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 $\sum_{i=0}^{m-1} \mathrm{sgn}(\vec{j}_i \times \vec{c}_i) = 0$.*

Proof A Jordan curve divides the plane into two disjoint regions which we call the *black region* and the *white region*. We adopt the convention that the black region lies to the right as the Jordan curve is traversed in the direction of its orientation while the white region 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 black region to the white region or from the white region to the black region. In the first case, $\mathrm{sgn}(\vec{j}_i \times \vec{c}_i) = 1$ while in the second case $\mathrm{sgn}(\vec{j}_i \times \vec{c}_i) = -1$. Since successive crossings, i and $i+1$, must occur in opposite directions:

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

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

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

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 equals the network edge weight when traversed in the direction of its 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 weight of the network edge. It follows that when a simple cycle is traversed in a given direction, the signs of the cross products of vectors tangent to the cycle and the edges of the knot diagram equal the weights of the network edges. 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 an oriented Jordan curve in the plane. Therefore by Lemma 2.1, the sum of the weights for simple cycles is zero. Complex cycles, in turn, are the sums of one or more simple cycles, each of which is an oriented Jordan curve. 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 2.5 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 cut out from a sheet of paper. For each region, R, create γ_R copies of the paper panel, where γ_R is a so-

lution to the above system of difference equations. 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. $A(1) \rightleftharpoons B(1), ..., A(n) \rightleftharpoons B(n)$). Let the side of $A(n+1)$ adjacent to B remain non-identified. Now, unless γ_A equals $n+1$, identify the side (bordering B) of each panel (above region A) numbered $n+2$ through γ_A with the adjacent side of the copy of region B numbered $n+1$ through $\gamma_A - 1$ (i.e. $A(n+2) \rightleftharpoons B(n+1), ..., A(\gamma_A) \rightleftharpoons B(\gamma_A - 1)$). 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:[5]

(a) $A(1) \rightleftharpoons B(1), ..., A(n) \rightleftharpoons B(n)$

(b) $A(n+1) \rightharpoonup \emptyset$

(c) $A(n+2) \rightleftharpoons B(n+1), ..., A(\gamma_A) \rightleftharpoons B(\gamma_A - 1)$

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 an anterior 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

[5]If the final index of a subrange is less than the initial index, then that subrange is empty.

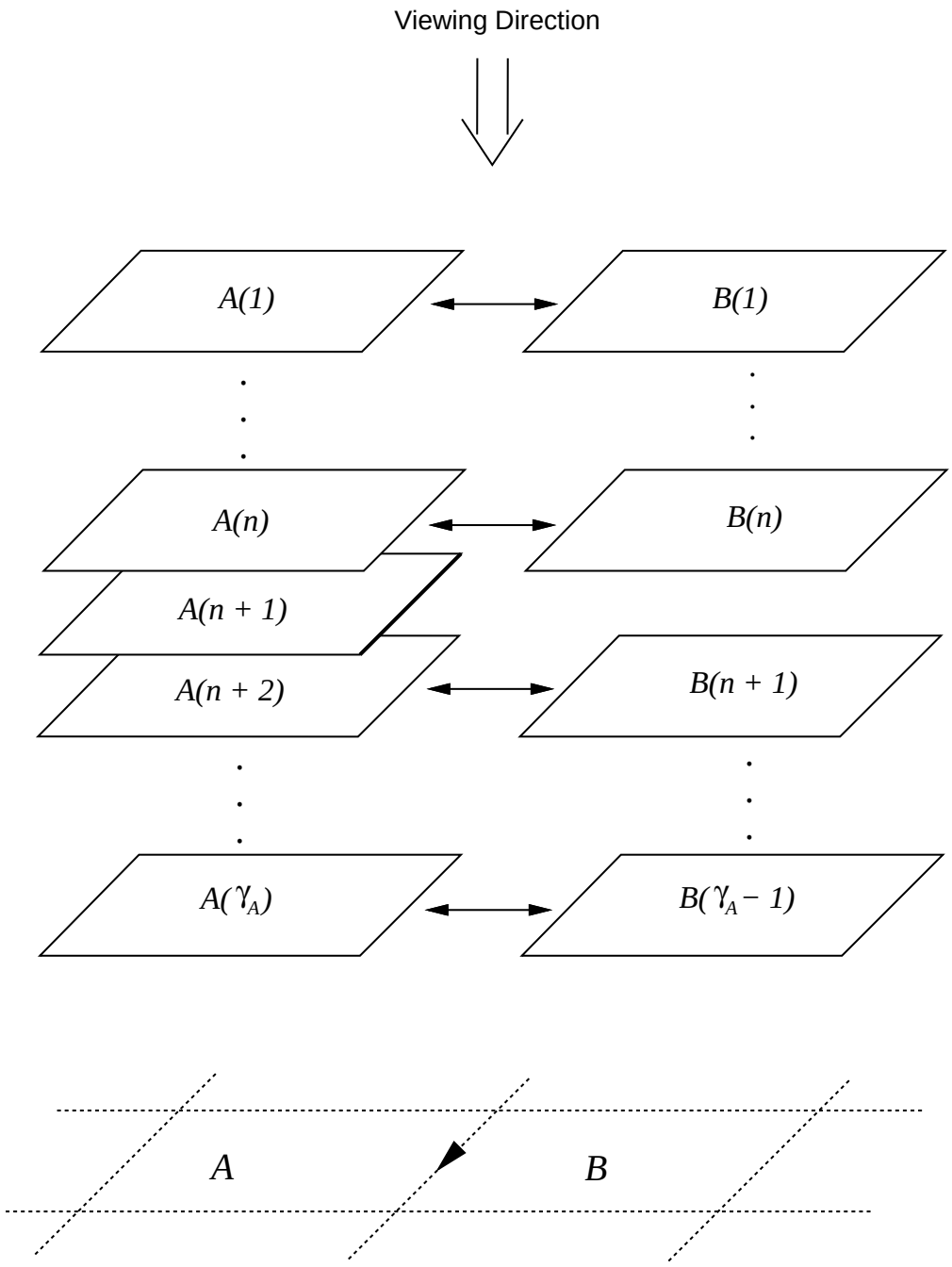

Figure 2.8: Paneling construction. Paper panels stacked above regions A and B in the plane. Following the *identification scheme*, all copies of regions A and B except $A(n+1)$ are glued along their adjacent sides. The free side of $A(n+1)$ becomes part of the boundary of the surface.

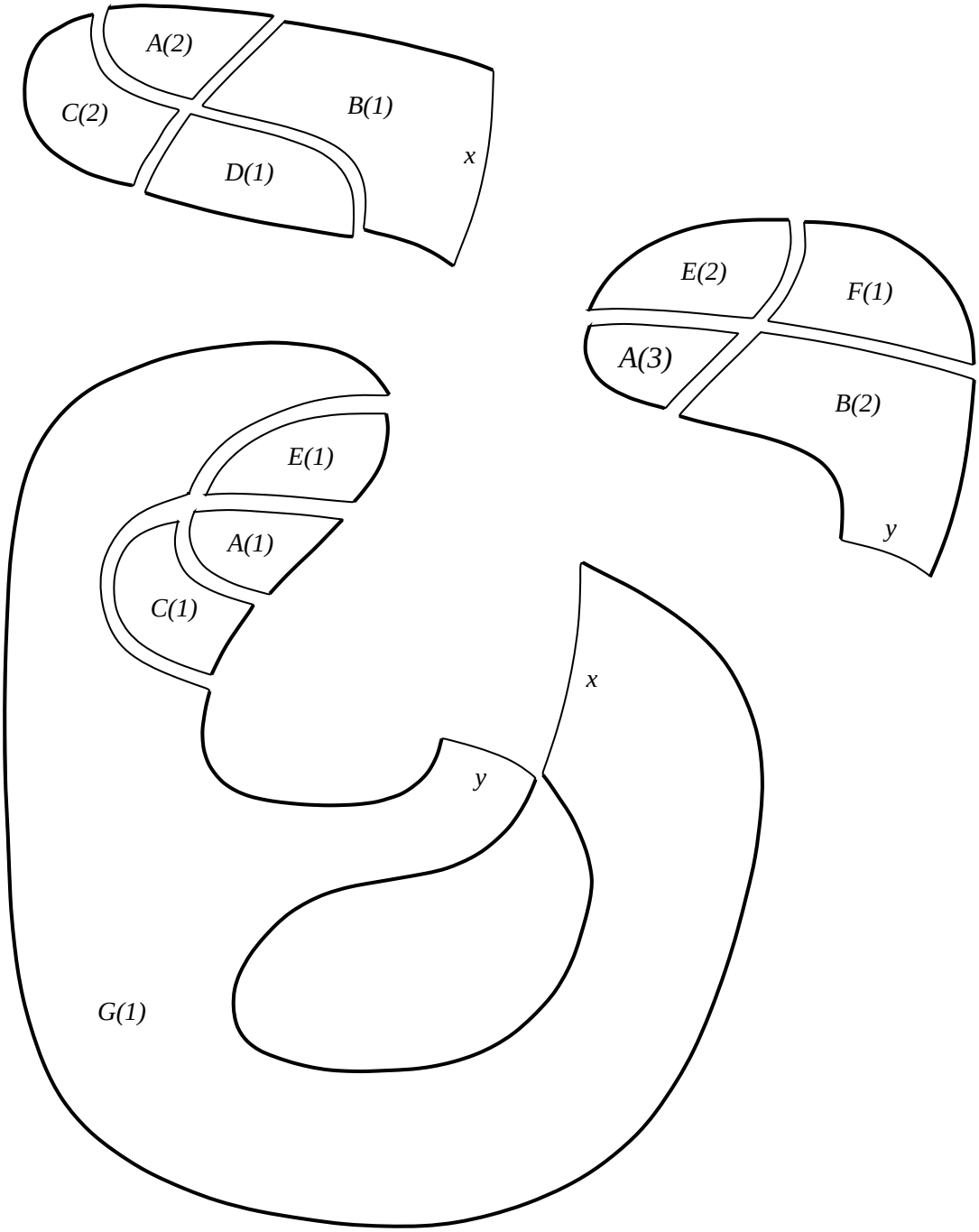

Figure 2.9: The paneling resulting from the construction. Bold edges remain free, and form the boundary. Additional identifications are indicated by x and y.

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 ensures 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 meet and 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 either discs or half-discs, it will be useful to study a negative example. The knot diagram shown in Figure 2.10 satisfies all conditions 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 2.11. 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

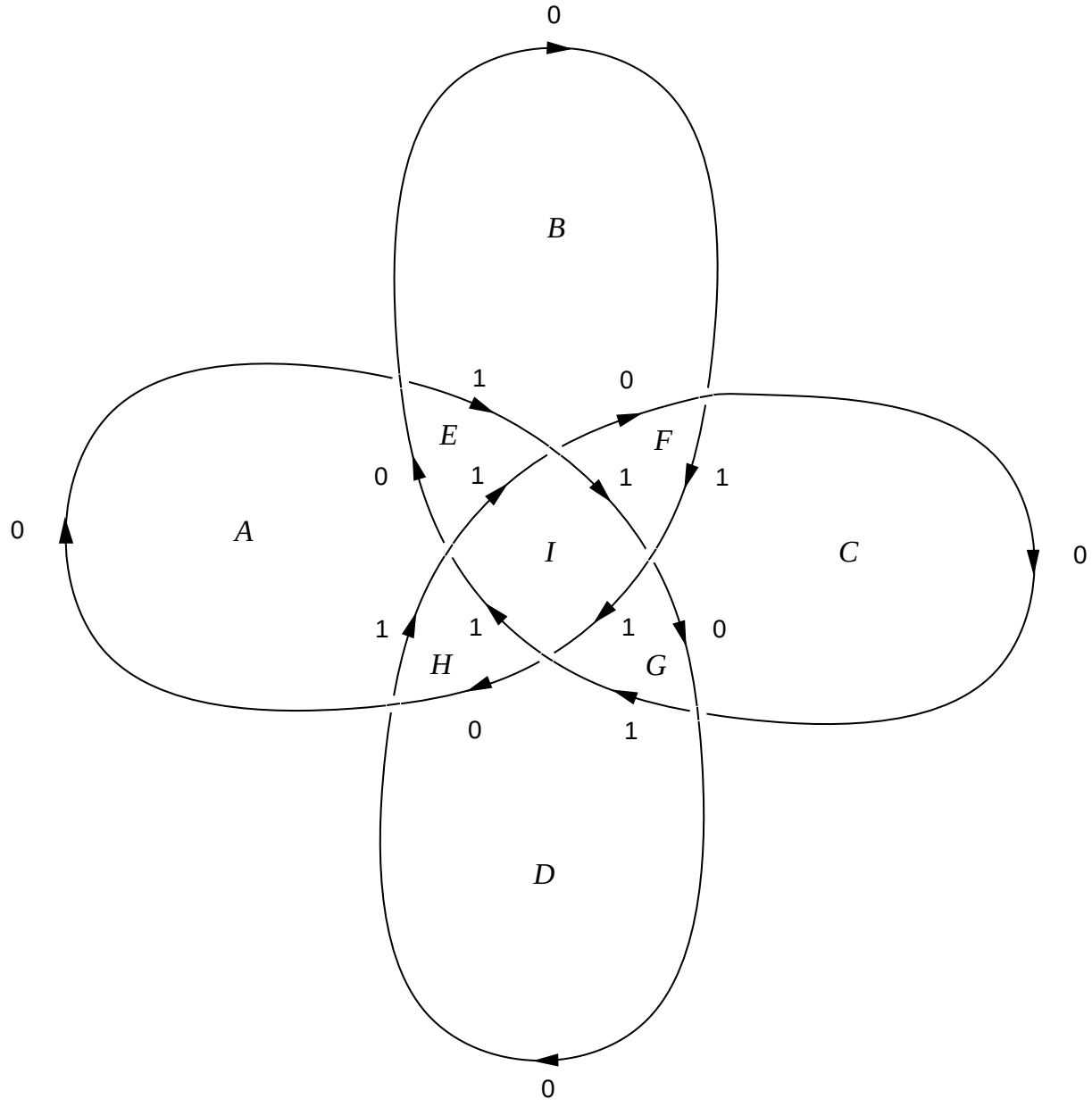

Figure 2.10: A knot diagram violating the labeling scheme. The labeling of the four crossings on the boundary of region I is in violation of the depth order requirement.

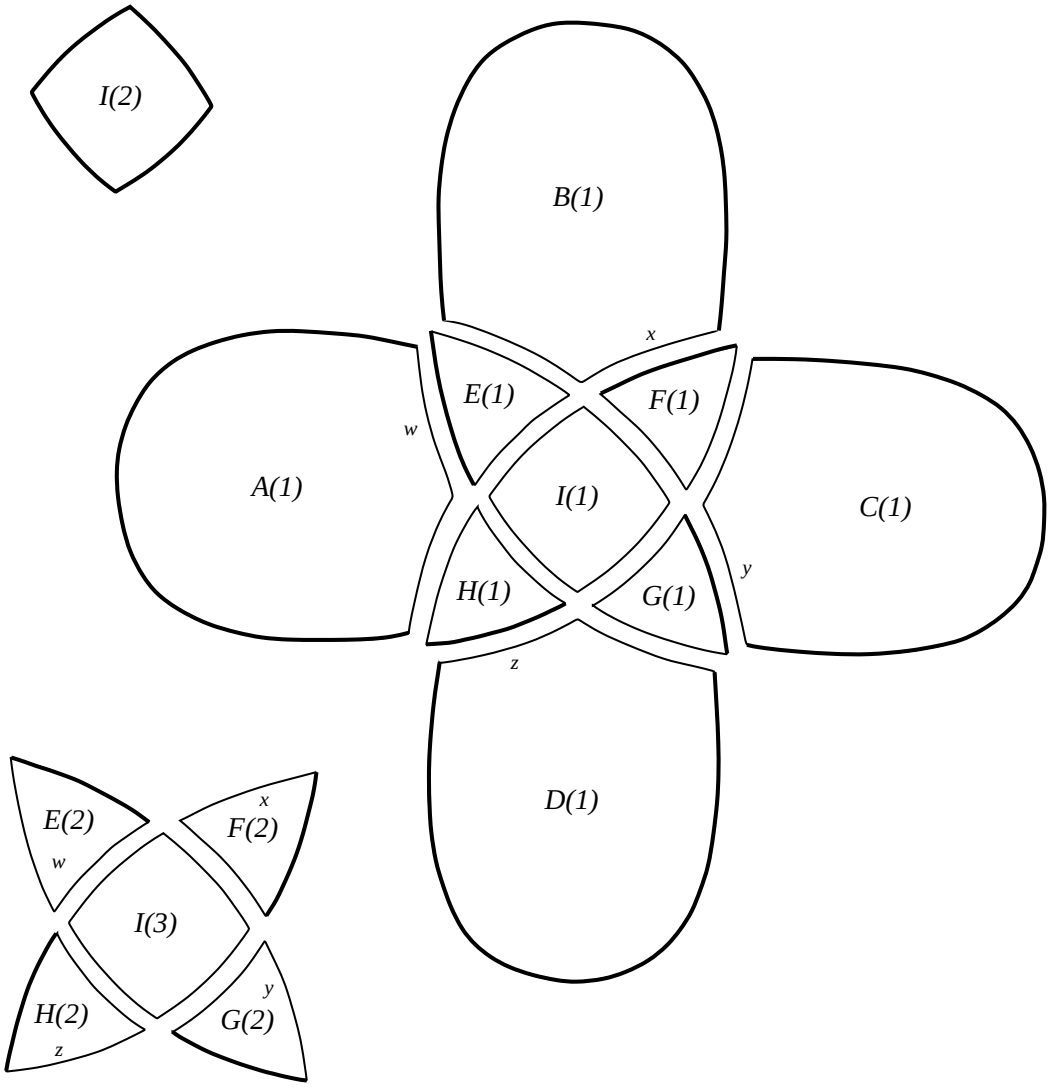

Figure 2.11: A pathological paneling. This paneling is produced when the construction is applied to the knot diagram in the previous figure. Bold edges remain free, and form the boundary. Additional identifications are required along edges labeled w, x, y and z. Although the set of boundary points form a single connected component, the set of interior points does not.

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.

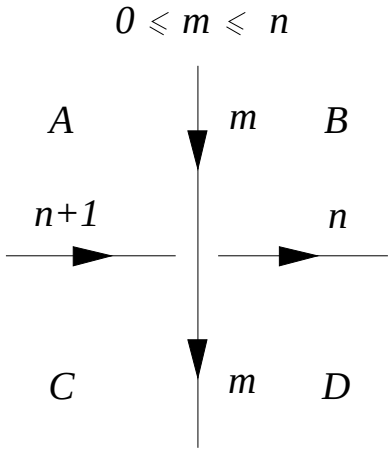

Figure 2.12: Four regions incident at a crossing with writhe equal to +1.

Let the four regions incident at a crossing with writhe equal to +1 be A, B, C and D as illustrated in Figure 2.12. 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 the image of both boundaries, 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 γ_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.

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) $A(1) \rightleftharpoons B(1), ..., A(m) \rightleftharpoons B(m)$

   (b) $A(m+1) \rightharpoonup \emptyset$

   (c) $A(m+2) \rightleftharpoons B(m+1), ..., A(\gamma_A) \rightleftharpoons B(\gamma_A - 1)$

2. Identifications between copies of C and D.

   (a) $C(1) \rightleftharpoons D(1), ..., C(m) \rightleftharpoons D(m)$

   (b) $C(m+1) \rightharpoonup \emptyset$

   (c) $C(m+2) \rightleftharpoons D(m+1), ..., C(\gamma_C) \rightleftharpoons D(\gamma_C - 1)$

3. Identifications between copies of C and A.

   (a) $C(1) \rightleftharpoons A(1), ..., C(n+1) \rightleftharpoons A(n+1)$

   (b) $C(n+2) \rightharpoonup \emptyset$

   (c) $C(n+3) \rightleftharpoons A(n+2), ..., C(\gamma_C) \rightleftharpoons A(\gamma_C - 1)$

4. Identifications between copies of D and B.

   (a) $D(1) \rightleftharpoons B(1), ..., D(n) \rightleftharpoons B(n)$

   (b) $D(n+1) \rightharpoonup \emptyset$

(c) $D(n+2) \rightleftharpoons B(n+1), ..., D(\gamma_D) \rightleftharpoons B(\gamma_D - 1)$

The identifications can be grouped into five consecutive subranges instead of three by exploiting the fact that $\gamma_C = \gamma_A + 1 = \gamma_D + 1 = \gamma_B + 2$ and $0 \leq m \leq n \leq \gamma_C$:

1. Identifications between copies of A and B.

   (a) $A(1) \rightleftharpoons B(1), ..., A(m) \rightleftharpoons B(m)$

   (b) $A(m+1) \rightharpoonup \emptyset$

   (c) $A(m+2) \rightleftharpoons B(m+1), ..., A(n+1) \rightleftharpoons B(n)$

   (d) None.

   (e) $A(n+2) \rightleftharpoons B(n+1), ..., A(\gamma_C - 1) \rightleftharpoons B(\gamma_C - 2)$

2. Identifications between copies of C and D.

   (a) $C(1) \rightleftharpoons D(1), ..., C(m) \rightleftharpoons D(m)$

   (b) $C(m+1) \rightharpoonup \emptyset$

   (c) $C(m+2) \rightleftharpoons D(m+1), ..., C(n+1) \rightleftharpoons D(n)$

   (d) $C(n+2) \rightleftharpoons D(n+1)$

   (e) $C(n+3) \rightleftharpoons D(n+1), ..., C(\gamma_C) \rightleftharpoons D(\gamma_C - 1)$

3. Identifications between copies of C and A.

   (a) $C(1) \rightleftharpoons A(1), ..., C(m) \rightleftharpoons A(m)$

   (b) $C(m+1) \rightleftharpoons A(m+1)$

   (c) $C(m+2) \rightleftharpoons A(m+2), ..., C(n+1) \rightleftharpoons A(n+1)$

   (d) $C(n+2) \rightharpoonup \emptyset$

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

4. Identifications between copies of D and B.

   (a) $D(1) \rightleftharpoons B(1), ..., D(m) \rightleftharpoons B(m)$

   (b) None.

   (c) $D(m+1) \rightleftharpoons B(m+1), ..., D(n) \rightleftharpoons B(n)$

   (d) $D(n+1) \rightharpoonup \emptyset$

   (e) $D(n+2) \rightleftharpoons B(n+1), ..., D(\gamma_C - 1) \rightleftharpoons B(\gamma_C - 2)$

The effect of gluing the panels according to the prescribed identifications is best illustrated by means of a diagram such as Figure 2.13. 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.

The effect of identifications 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 identifications 1-4 (a),(c) and (e) is: 1) To create m interior vertices above the upper boundary vertex; 2) To create $n - m$ interior vertices between the upper and lower boundary vertices; and 3) To create $\gamma_C - n - 2$ interior vertices beneath the lower boundary vertex. Clearly, if $m = 0$ no interior vertices are created above the upper boundary vertex. Similiarly, if $m = n$ then no interior vertices are created between the upper and lower boundary vertices. Finally, if $n = \gamma_c - 2$, no interior vertices are created beneath the lower boundary vertex. Inspection of the diagram confirms that exactly four panels meet at each interior vertex, and that the neighborhood structure of each interior vertex resembles a disc.

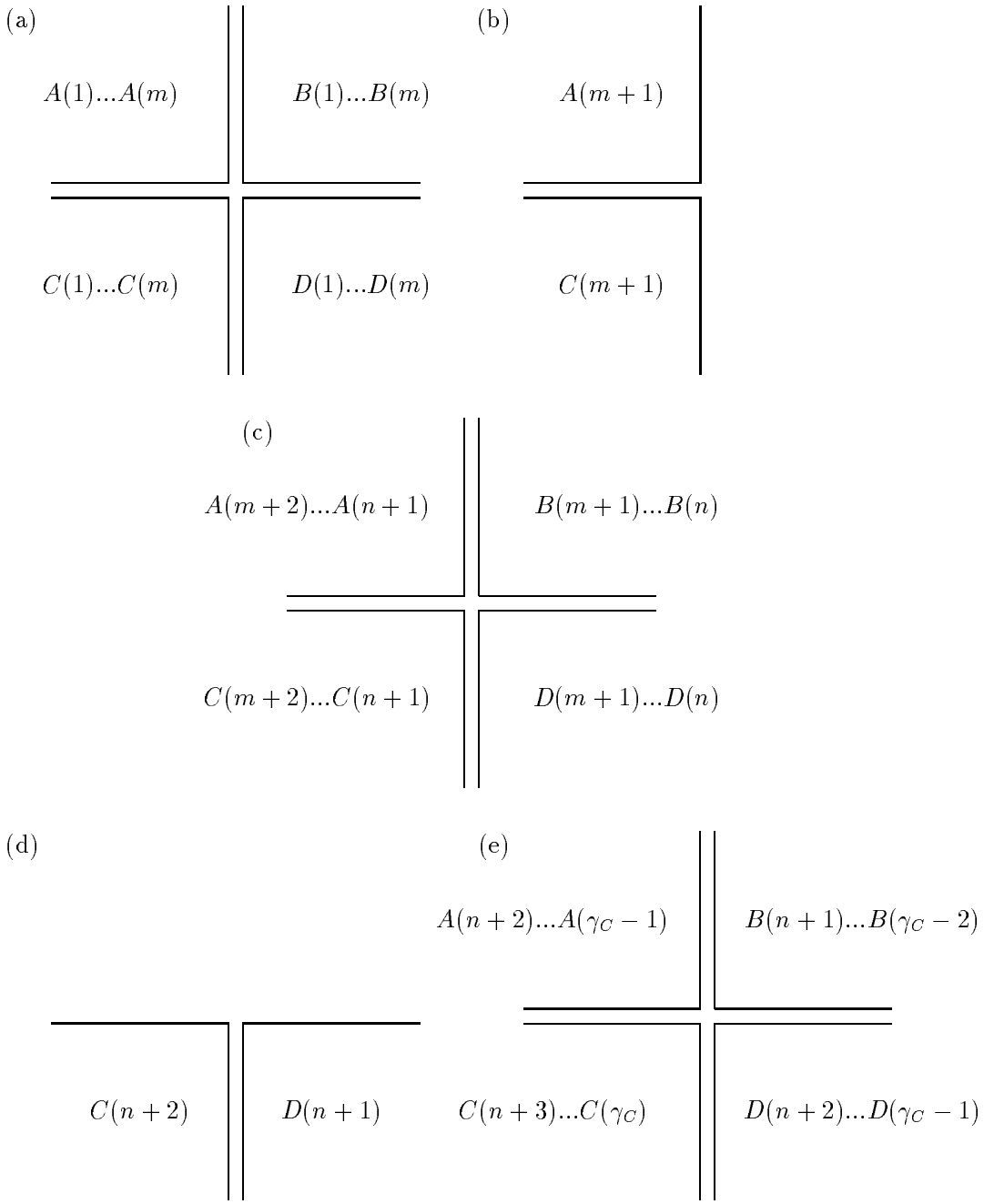

Figure 2.13: Paneling vertices produced by the construction. These vertices (and no others) are produced by the construction when applied to edges incident at a crossing. Thick lines are boundary edges.

We now show that the image of the boundary of the anterior scene produced by the construction corresponds to the knot diagram in every respect and that the view is generic. First, the definition of the construction 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 view 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 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.□

2.4 Smooth Manifold-Solids

An orientable surface divides three-dimensional space into two disjoint sets which can be interpreted as the interior and exterior of a manifold-solid. In addition, if the first three derivatives of surface orientation are continuous everywhere, then the manifold-solid is said to be *smooth*. We conclude this chapter by conjecturing that the anterior surfaces of scenes composed of smooth manifold-solids can be represented by labeled knot diagrams. Developing these ideas more fully is a topic for future work.

The *occluding contour* is the projection onto the image plane of the *contour generator*, which is the locus of points on the surface of a smooth manifold-solid which are tangent to the viewing direction. The contour generator consists of one or more

simple closed curves which divide the surface into two disjoint sets called the *anterior* and *posterior*. The anterior, which possibly consists of multiple components, is the portion of the surface potentially visible to the viewer. The surface orientation of every anterior point has a positive component in the viewing direction.

Conjecture 2.1 *Let $\mathcal{A}$ be the anterior scene formed by slicing a manifold-solid along the contour generator and discarding the backward facing surfaces. There is a topologically equivalent anterior scene $\mathcal{A}'$, arbitrarily close to $\mathcal{A}$, which can be represented by a labeled knot diagram.*

Rationale If the bounding surface of a manifold-solid is cut everywhere along the contour generator, then each of the frontward facing components forms an orientable surface with boundary. Although these surfaces define an anterior scene (i.e. $\mathcal{A}$), in general, the embedding of this scene will violate the definition of generic view, and so cannot be represented by a labeled knot diagram.

First, since the surface normal everywhere along the boundary of the anterior scene (i.e. $\delta\mathcal{A}$) is perpendicular to the viewing direction (the boundary consists of points which previously formed the contour generator), an arbitrarily small rotation will result in self-occlusion of some portion of $\mathcal{A}$. Although non-generic in the broad sense of the word, this is not where the problem lies.

The problem lies in the fact that the occluding contour of the smooth-manifold solid can contain *cusps*, which are the images of points where the direction of the contour generator coincides with the the viewing direction[33]. Because $\delta\mathcal{A}$ is described by the same space curve(s) as the contour generator, $\delta\mathcal{A}$ is asymptotic to the viewing direction at precisely the same points. Consequently, what were cusps in the occluding contour (i.e. prior to slicing) are now multiplicity two points (i.e. singularities) in the image of $\delta\mathcal{A}$. However, unlike crossings, these singularities are not stable: when the viewing direction is changed slightly, $\delta\mathcal{A}$ will no longer be asymptotic at these points and the singularities disappear. This violates the definition of generic view.

Proof of the conjecture would require showing that there exists a transformation, T, which maps anterior scene $\mathcal{A}$ to a topologically equivalent anterior scene $\mathcal{A}'$ arbitrarily close to $\mathcal{A}$. Furthermore, transformation T must be defined so that the images of both $\mathcal{A}'$ and $\delta\mathcal{A}'$ are non-singular. This would establish that the view of $\mathcal{A}'$ is generic, so that the image of $\delta A'$ is a labeled knot diagram.

Figure 2.14(a) depicts a smooth manifold-solid and its occluding contour. The occluding contour contains two cusps which form a "swallowtail." Figure 2.14(b) shows the anterior scene formed by slicing the manifold-solid along the contour generator, discarding the backward facing surfaces, and applying transformation T. Transformation T, which can be thought of as "erosion" of $\mathcal{A}$ inward from its boundary, changes one of the two cusps into a small loop. The image of $\delta A'$ is a labeled knot diagram. This labeled knot diagram represents an anterior scene which is an arbitrarily close to the anterior surfaces of the smooth manifold-solid.

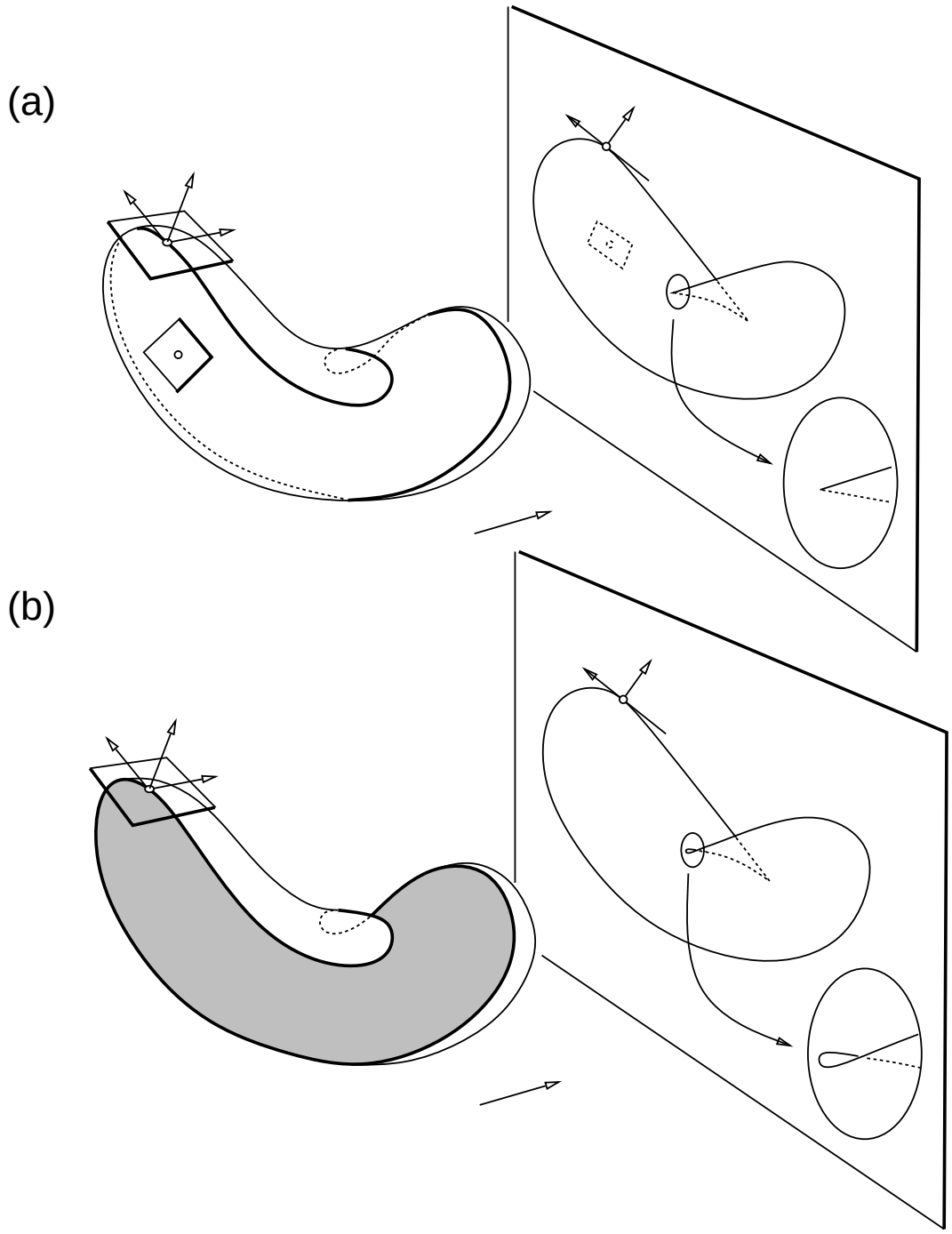

Figure 2.14: Anterior scene constructed from smooth manifold-solid. (a) A smooth manifold-solid and its image (after Callahan[8]). The image of the contour generator contains two cusps which form a "swallowtail." (b) Anterior scene $\mathcal{A}'$ is formed by slicing the manifold-solid along the contour generator, discarding the backward facing surfaces, and applying transformation T. Transformation T, which can be thought of as "erosion" of $\mathcal{A}$ inward from its boundary, changes one of the two cusps into a small loop. We conjecture that the image of $\delta\mathcal{A}'$ is a labeled knot diagram.

CHAPTER 3

FIGURAL COMPLETION: A PROBLEM-LEVEL FORMULATION

In this chapter, 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.

This thesis' treatment of the figural completion problem is quite a bit different than many treatments of perceptual organization problems in computer vision. Too few grouping problems are explored at the level of computational theory. Often there is no clear statement of a computational goal. This quickly leads to the paradoxical situation of algorithms which compute functions which have no definitions independent of the algorithms themselves. Quite often, a clearly developed computational theory leads directly to a formulation as an optimization problem. In this thesis, this optimization problem takes the form of an *integer linear program.*

3.1 Natural Constraints

This chapter begins with the somewhat pessimistic observation that the natural constraints which apply to this problem are few in number and not nearly sufficient to determine a unique solution. That having been said, we observe that 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* requirement. Basically, only closed plane-curves satisfying the labeling scheme define generic views of anterior scenes. Because this was discussed in detail in the previous chapter, we consider the second source of constraints, which will be termed the *stimulus conformity* requirement. The importance of this requirement has been stressed by Irvin Rock[48], who 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. Consider, for example, the stimulus depicted in Figure 3.1(a). Although Figure 3.1(b) and (c) both depict generic views of topologically valid anterior scenes, the anterior scene depicted in Figure 3.1(c) contradicts the stimulus in two different ways. First, contours plainly visible in the stimulus (i.e. luminance boundaries) are hypothesized to have depths greater than zero in the solution. Second, although the depth of the black surface is hypothesized to be zero everywhere, its boundary is only intermittently defined by a change in image brightness. In particular, there is no luminance boundary where it nominally overlaps the white triangle. The importance of stimulus conformity becomes obvious when one realizes that Figure 3.1(c) is actually Figure 3.1(b) (the correct inference) viewed from underneath.

3.2 Inherent Ambiguities

Given only these constraints, the problem of computing a labeled knot diagram from an image of a 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 Figure 3.2. Figure 3.2(a) apparently depicts a square opaque surface occluding a second surface of indeterminate shape.

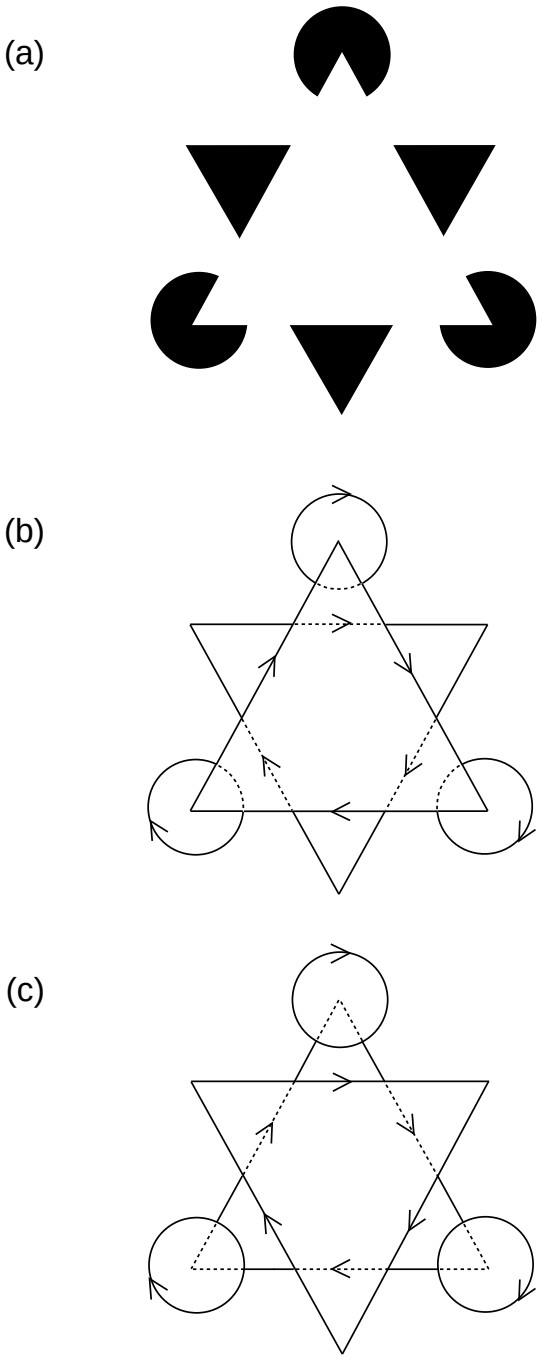

Figure 3.1: Stimulus conformity. (a) A variation of the Kanizsa triangle. (b) The anterior scene generally perceived by human observers. (c) A second anterior scene. Although topologically valid, this interpretation is not supported by the image evidence. In fact (c) depicts how (b) would appear if it were viewed from underneath.

Without "X-ray" vision, this problem can not 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 Figure 3.2(b). The other completions can be imagined, but they are 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. The word "unit" is used here in the manner it is used by the Gestalt psychologists (see for example Schumann[54]), and in this thesis refers to the boundary components which are the products of the grouping process. Unit ambiguity is illustrated in Figure 3.3. Rock[48] calls the interpretation in Figure 3.3(b) the "literal solution." Both the literal solution and the solution depicted in Figure 3.3(c) are topologically valid and conform to the image evidence. Yet these two interpretations consist of very different sets of "units," or boundary components.[1]

Second, there is *depth ambiguity*, which is ambiguity in the signs of occlusion (i.e. figure-ground sense) of different boundary components and their relative depths at crossings. Consider Figure 3.4(a), a variation of the Kanizsa triangle. Figure 3.4(b) illustrates the organization experienced by most observers: An illusory triangle appears to partially occlude three black discs and a second black triangle. Yet the other organizations depicted in the figure (i.e. (c), (d), (e) and (f)) also represent topologically valid anterior scenes conforming to the image evidence. These differ from the preferred interpretation in the signs of occlusion assigned to the different boundary components, and in the relative depths of different portions of boundary.

[1]One might ask why boundary components are considered to be the elementary "units" when complete surfaces might be more natural. There are several reasons for this choice, which will become clear in Chapter 5.

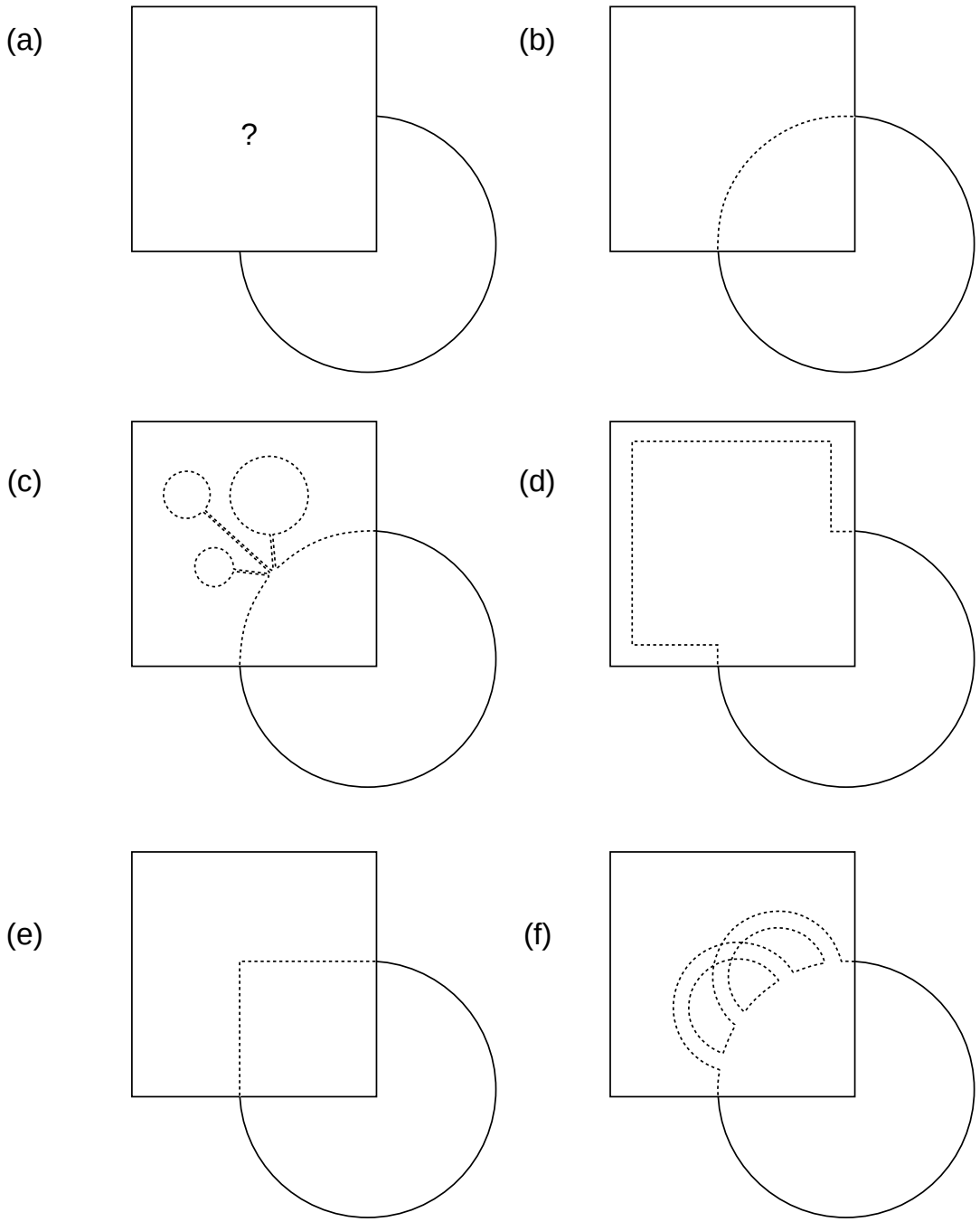

Figure 3.2: Shape ambiguity. (a) A square opaque surface occluding a second surface of indeterminate shape. (b) The shape which properly functioning human visual systems infer. (c), (d), (e) and (f) These completions can be imagined, but they are not perceived preattentively.

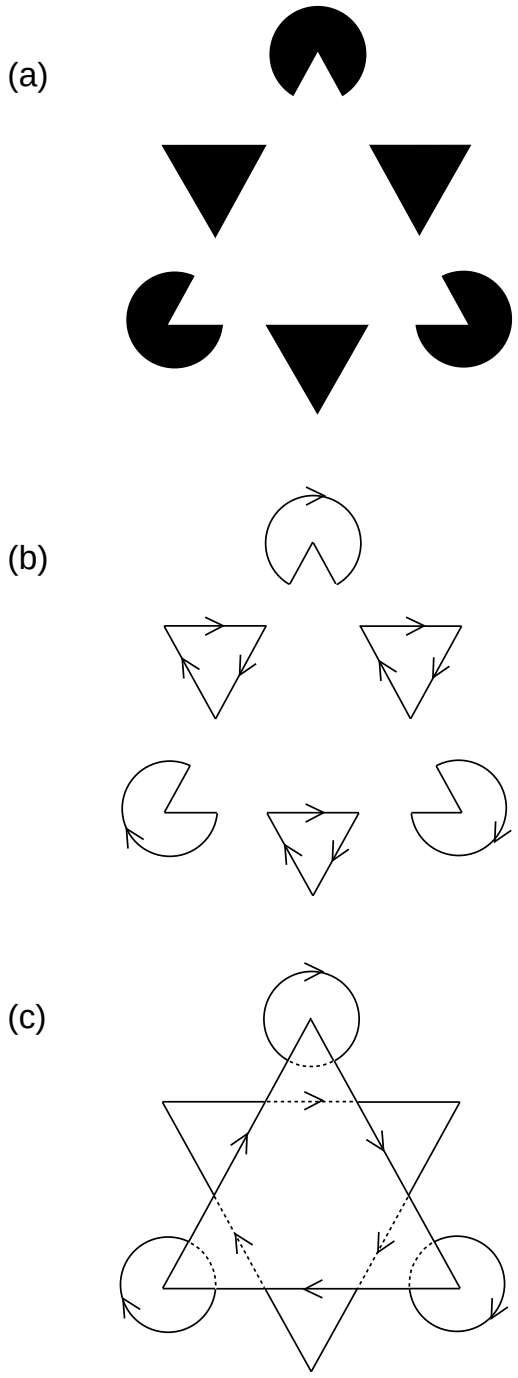

Figure 3.3: Unit ambiguity. (a) A variation of the Kanizsa triangle. (b) The organization Rock calls the "literal solution." (c) The organization experienced by most observers. Although both are topologically valid anterior scenes conforming to the image evidence, they contain different sets of closed boundaries.

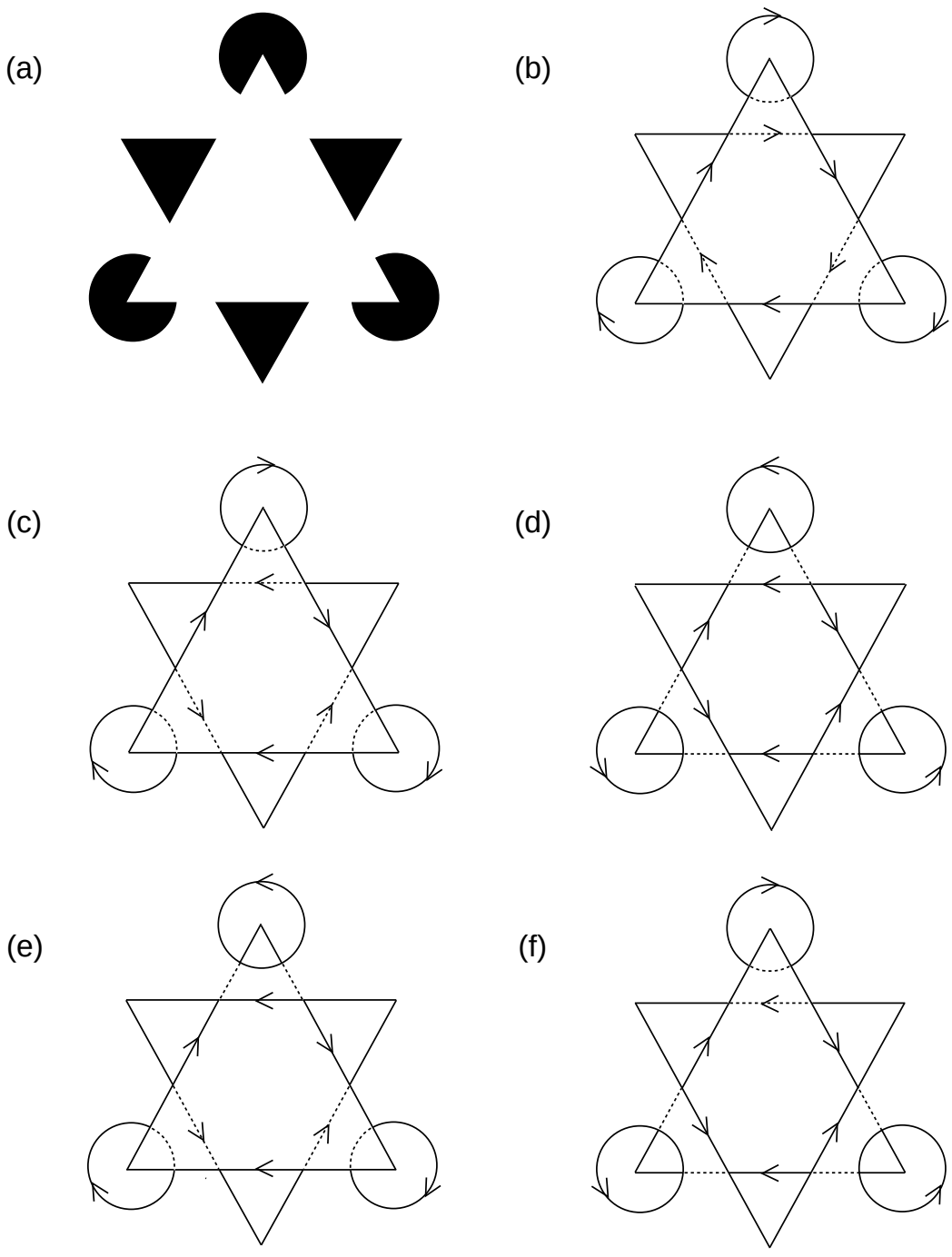

Figure 3.4: Depth ambiguity. (a) A variation of the Kanizsa triangle. (b) The organization experienced by most observers. (c), (d), (e) and (f) Other topologically valid anterior scenes conforming to the image evidence. Although, each organization contains the same set of closed boundaries, individual boundaries differ in sign of occlusion and/or relative depth.

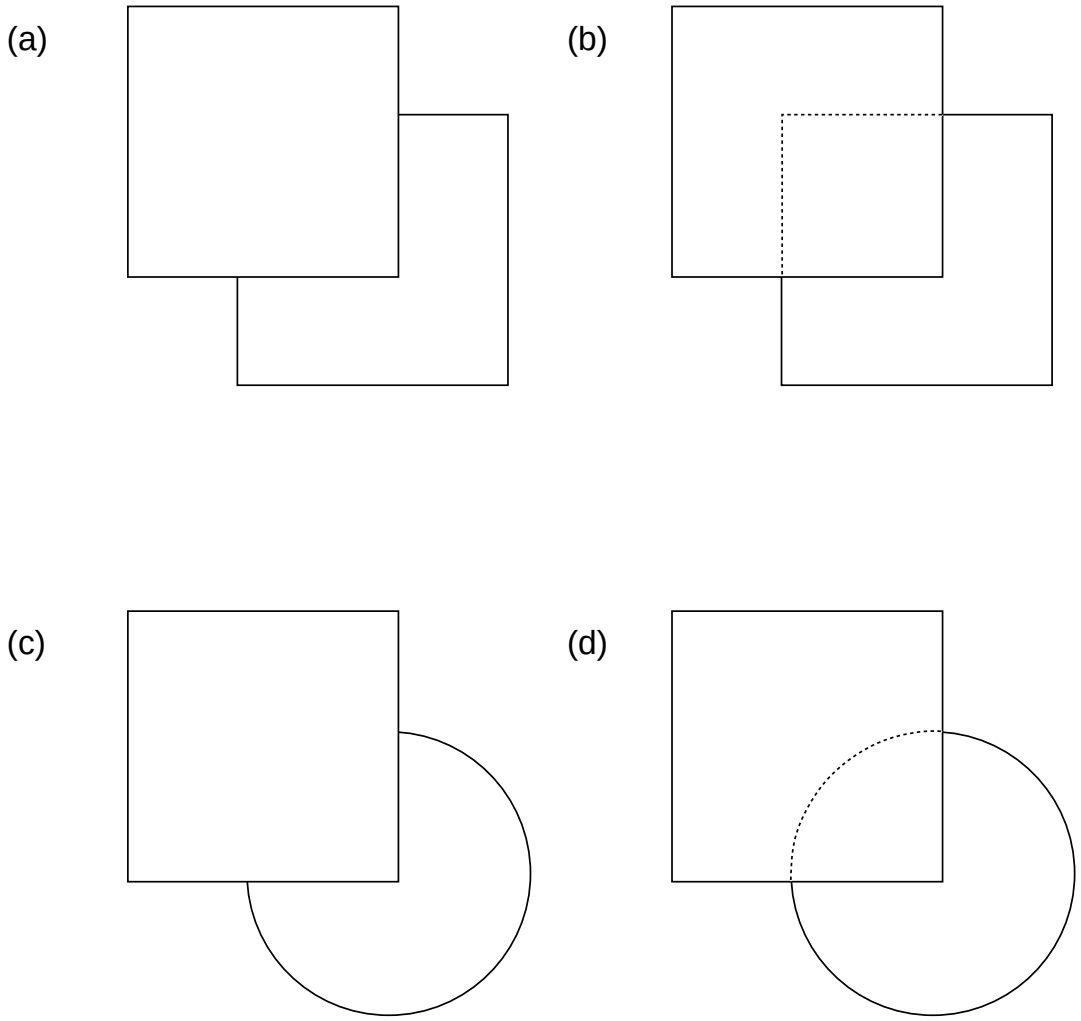

Figure 3.5: Completion shapes. Amodal completion of a partially occluded circle and square (redrawn from Kanizsa[29]). In both cases, completion is accomplished in a manner which preserves tangent and curvature continuity at the terminal ends of the occluded boundaries.

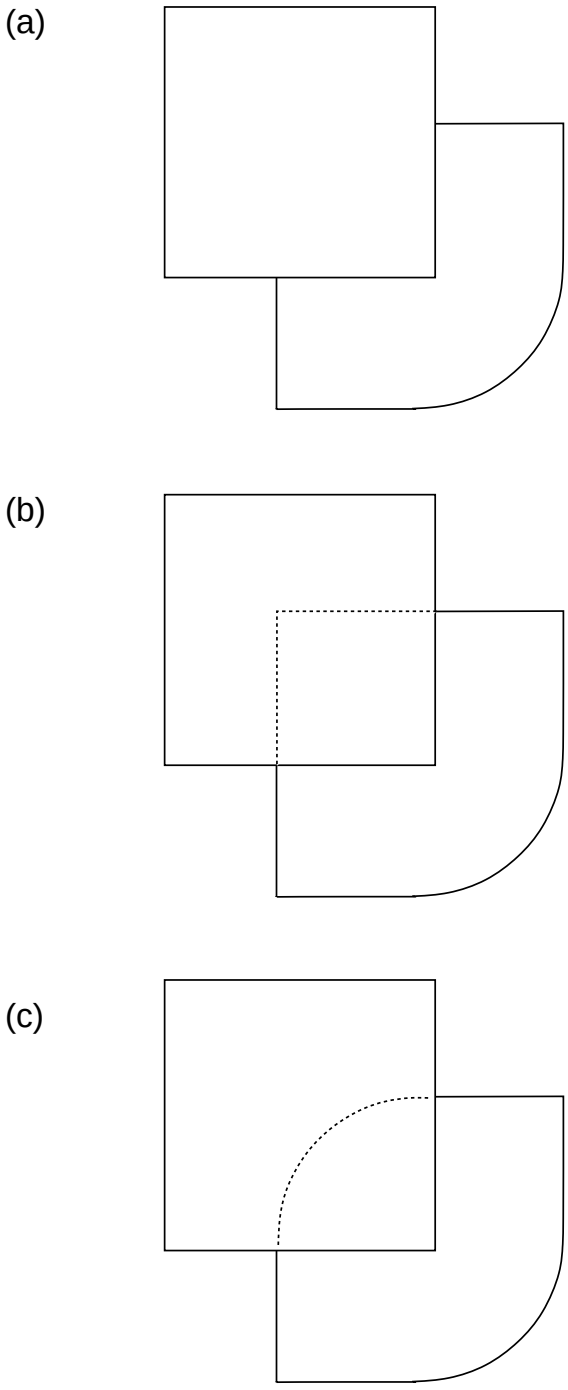

Figure 3.6: Symmetry is not a factor. This figure, redrawn from Kanizsa[29], demonstrates that symmetry is not a factor in determining completion shape. (a) The stimulus. (b) The shape reported by test subjects. (c) This shape can be imagined, but is not preattentively inferred.

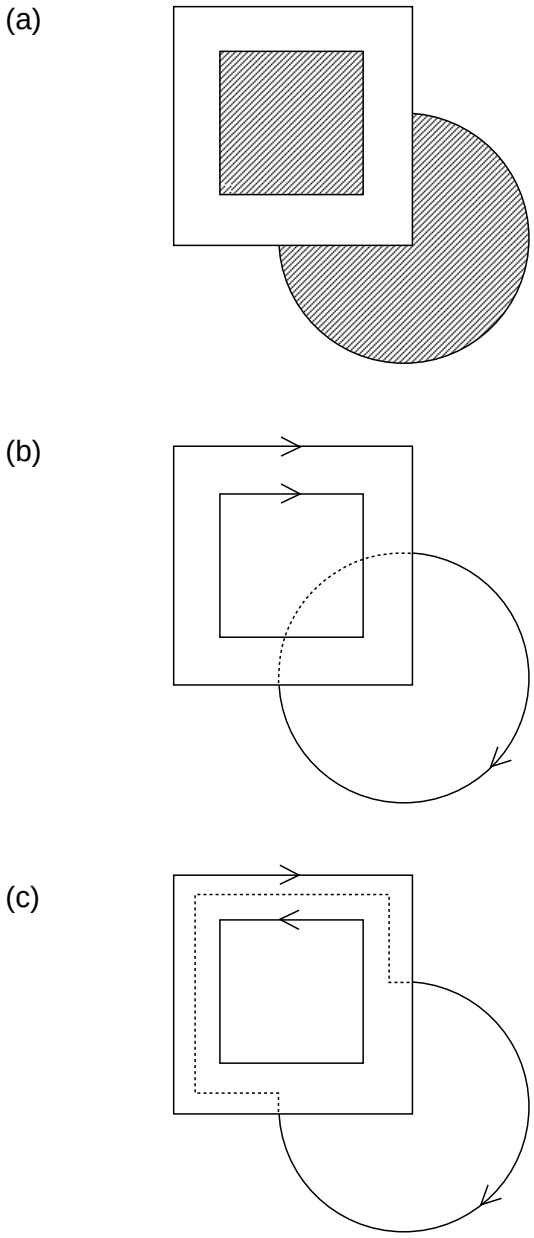

Figure 3.7: Shape is independent of role. Completion shape is determined by the position, orientation and curvature of the boundary fragments which the completion joins, not by the role the completion might play in a larger organization. The arrows represent hypothetical signs of occlusion of the different boundary components. In (b) the smaller square is nominally interpreted as a solid, while in (c) this square is nominally interpreted as a hole. If the smaller square were a hole, and the completion were of the shape depicted in (b), then the completion would be visible, which it very clearly is not. If the smaller square were a hole, then the shape of the completion must be similiar to that depicted in (c), which is not very compelling.

3.3 Completion Shape

In human vision, a relatively small number of factors seem to determine the shape of perceptual completions. These include the position, orientation and curvature of the boundary fragments which the completion joins. Although few in number, the interplay of these factors can be quite subtle. Consider Figure 3.5 (redrawn from Kanizsa[29]), which illustrates the amodal completion of a partially occluded circle and square. In both cases, completion is accomplished in a manner which preserves tangent and curvature continuity at the terminal ends of the occluded boundaries. Significantly, in the case of the square, curvature continuity at these "distinguished" points is achieved at the expense of introducing a tangent and curvature discontinuity (i.e. a corner) at a more "generic" location underneath the occluding surface. The human visual system seems unwilling to accept the spatial coincidence inherent in the superposition of a tangent and curvature discontinuity and a point belonging to a second boundary.

Significant for not being a factor is symmetry. Consider Figure 3.6 (also redrawn from Kanizsa[29]). Figure 3.6(b) depicts the shape of the completion reported by test subjects when presented with the stimulus depicted in Figure 3.6(a). Kanizsa maintains that the shape depicted in Figure 3.6(c) can be imagined, but is not preattentively perceived. This suggests that completion shape is determined without regard to the role the completion might play in a larger organization. Additional evidence that this is in fact the case is provided by Figure 3.7. The arrows represent hypothetical signs of occlusion of the different boundary components. In Figure 3.7(b), the smaller square is nominally interpreted as a solid, while in Figure 3.7(c), this square is nominally interpreted as a hole. If the smaller square were a hole, and the completion were of the shape depicted in Figure 3.7(b), then the completion would be visible, which it very clearly is not. If the smaller square were a hole, then the shape

of the completion must be similiar to that depicted in Figure 3.7(c), which is not very compelling. In fact, it seems quite contrived. This suggests that while the shape and location of completions can determine topological aspects of the organization (i.e. unit and depth), topological aspects do not determine completion shape. A still more radical notion is advanced by Kellman and Loukides[32], who theorize that "unit formation" occurs prior to and independently of "depth placement." This possibility will be considered in detail in Chapter 5.

3.3.1 Curves of Least Energy

Even if we assume that completion shape is solely a function of the position, orientation and curvature of the boundary fragments which the completion joins, tangent and curvature continuity, do not, by themselves, uniquely determine completion shape. Although no comprehensive theory exists, computational models have been proposed for the shape of illusory contours joining boundary fragments with orientation difference significantly less than $\frac{\pi}{2}$. Ullman[57] originally hypothesized that the curve used by the human visual system to join two contour fragments is constructed from two circular arcs. Each circular arc is tangent to its sponsoring contour at one end and to the other arc at their point of intersection (the curve is continuous in tangent, but not curvature). According to Ullman, from the family of possible curves of this form, the pair of circular arcs which minimizes total bending energy (i.e. $E = \int \kappa(s)^2 \, ds$, where κ is curvature) models the shape of the illusory contour. Horn[22] subsequently proposed that the shape of the contour joining two boundary fragments is described by the true curve of least energy, to which Ullman's curve is a two arc approximation.

More recently, Kass, Witkin and Terzopolous[30] have demonstrated the utility of active minimum energy seeking contours called "snakes" in a variety of computer

vision applications. The "internal" energy in the snake active contour model is defined to be:

$$E_{int} = \frac{1}{2} \int \alpha |\mathbf{v}_{xx}(s)|^2 + \beta |\mathbf{v}_x(s)|^2 \, ds \tag{3.1}$$

Here $\mathbf{v}_x$ and $\mathbf{v}_{xx}$ are finite difference approximations of the first and second derivatives. The $|\mathbf{v}_{xx}(s)|^2$ therefore plays the role of $\kappa(s)^2$. Kass et. al suggest that the curve minimizing this functional models the shape of an illusory contour with the same end conditions and demonstrate this agreement on a number of figures which elicit illusory contours in human vision.

Nitzberg and Mumford[39] point out that the problem of computing the shape of the curve of least energy was first studied by Euler, and have used "elastica" to represent potential completions in a combinatorial optimization approach to figural completion.[2] The functional they propose is of the form:

$$E = \int (\alpha \, \kappa(s)^2 + \beta) \, ds \tag{3.2}$$

The elastica and snake energy functionals differ primarily in their second terms: the snake functional minimizes $|\mathbf{v}_x(s)|^2$ while the classical functional minimizes arc length. The significance of this difference is unclear.

Of course, even if one assumes that the human visual system uses the curve of least energy (or a related curve) to represent the shape of perceptual completions,

[2] As in this thesis, Nitzberg and Mumford[39] have studied figural completion in illusory contour figures such as the Kanizsa triangle. 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 this thesis, they assume that environmental surfaces can be represented by sets of closed, non-self-intersecting plane curves (i.e. Jordan curves) of constant depth. Consequently, surfaces with boundaries which project as self-intersecting curves in the image plane can not be represented. However, there is no physical basis for this assumption and there is abundant evidence (e.g. Figure 1.10) that the human visual system does not use it.

then the question of why it does so remains. Ullman[57] listed criteria he believed illusory contours joining boundary fragments with small orientation difference must satisfy, including smoothness, isotropy and extensibility. Smoothness and isotropy imply tangent continuity and rotational invariance, which are clearly desirable properties and in accordance with observation. Extensibility is subtler, and simply stated, requires that any two points on a curve of least energy be joined by a curve of least energy for that interval. This is an example of the optimal substructure property characteristic of problems which can be solved by dynamic programming, and in this instance, makes local parallel computation in a network possible (see Figure 5.8(b)).

3.3.2 Inferential Leverage

Extensibility provides a compelling argument at the algorithmic level for using the curve of least energy to fill the gaps between boundary fragments. Is there an additional, computational theory level rationale? Witkin and Tenenbaum[60] examine this question at length and reject those commonly cited. They point out that Occam's razor provides no philosophical justification for "smooth as possible" solutions in computer vision since the definition of smooth (or simple) inevitably depends on the choice of representation. They also point out that a minimum energy solution is, at best, only marginally more likely than any other, and can not be adequately justified on the basis of prior probabilities. Although it is possible that given the ambiguity in completion shape, algorithmic level considerations (i.e. extensibility) and marginally higher prior probabilities become the determining factors, Witkin and Tenenbaum ultimately offer a different rationale for the use of minimum energy solutions. They suggest that although these solutions are not significantly more likely than others, the simple fact that a smooth low energy interpolating curve exists is itself a reliable indicator of a *non-accidental* 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.

Witkin and Tenenbaum[60] argue that perceptually significant geometric relations like collinearity and parallelism are actually the zero distortion limiting cases of more general "fuzzy identity" measures.[3] The measures are defined to be the total energy in least-distortion transformations mapping one token's position into another's. They suggest that there is a correlation between the degree of distortion and the likelihood that the tokens are products of a single physical process (i.e. they have a common origin). In their opinion, whether the prior probability of a particular measurement is high or low is far less important than the ratio of the probability density functions in the non-accidental and accidental cases (i.e. $Pr(energy \mid \neg accidental)/Pr(energy \mid accidental)$). Distortion measures offering the most "inferential leverage" maximize this ratio.

Although maximizing likelihood can serve as a rationale for choosing one unit organization over another, there has been no serious attempt in this thesis to identify the probability density functions, which by Bayes' Rule, relate energy (and other elements of the shape of the curve of least energy) to completion likelihood. Whether or not this is actually feasible is another matter. It is Witkin and Tenenbaum's[60] opinion that identifying the precise functional form of the probability densities underlying perceptual organization is unrealistic given the complexity of the real world. They are only prepared to say that these densities achieve their maximum at zero

[3]This idea has been further developed by Richards, Feldman and Jepson[46] who propose that perceptually significant categories correspond to the "non-tranverse modes" of feature probability densities. We will have more to say about this in the next chapter, where the features and categories which proved useful in an experimental system are discussed.

and decrease monotonically with increasing distortion (e.g. from strict collinearity, strict parallelism, etc.).

As part of a study of the reduction of search made possible by prior grouping in visual object recognition, Jacobs[28] computed the probability density of the size of boundary gaps due to occlusion in random juxtapositions of a set of flat polygonal surfaces.[4] Jacobs concluded that small gaps predominate and that incident frequency rapidly drops off with increasing size. Since the probability density of distances between the endpoints of boundary fragments belonging to separate objects is uniform (assuming positional independence), gap size, like bending energy, provides inferential leverage in identifying fragments which form consecutive segments of the same boundary. However, the specific shape of the probability densities Jacobs computed varied considerably for different sets of polygons, casting doubt on the idea that a universal distribution exists or can be quantified in a principled way.

3.4 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 (even if arbitrarily) 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:

[4]It should be noted that Jacobs defined gap size as the Euclidean distance between fragment endpoints. It seems equally natural to define gap size as the length of the curve of least energy joining the two boundary fragments, and this is the definition adopted here.

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*.[5] The boundary fragments form the edges of a graph, $G_{input} = (V_{endpoints}, E_{fragments})$. 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 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 $G_{non-planar} = (V_{endpoints}, E_{fragments} \cup E_{completions})$ by adding edges representing potential completions. As with each element of $E_{fragments}$, each element of $E_{completions}$ is a contour joining two elements of $V_{endpoints}$. Finally,

[5]This is done without regard to the sign of contrast, so that no distinction is made between positive curvature maxima and negative curvature minima. Although Hoffman and Richards[20] propose that silhouettes are divided into parts at negative minima of curvature, we observe that this does not correctly predict part decomposition in illusory contour figures like the Kanizsa triangle. In general, negative minima of curvature will correspond to part boundaries only in the case of black surfaces occluding black surfaces against a white background. In the Kanizsa triangle, a presumably white triangle occludes black discs against a white background, so that positive curvature maxima are also perceived as part boundaries.

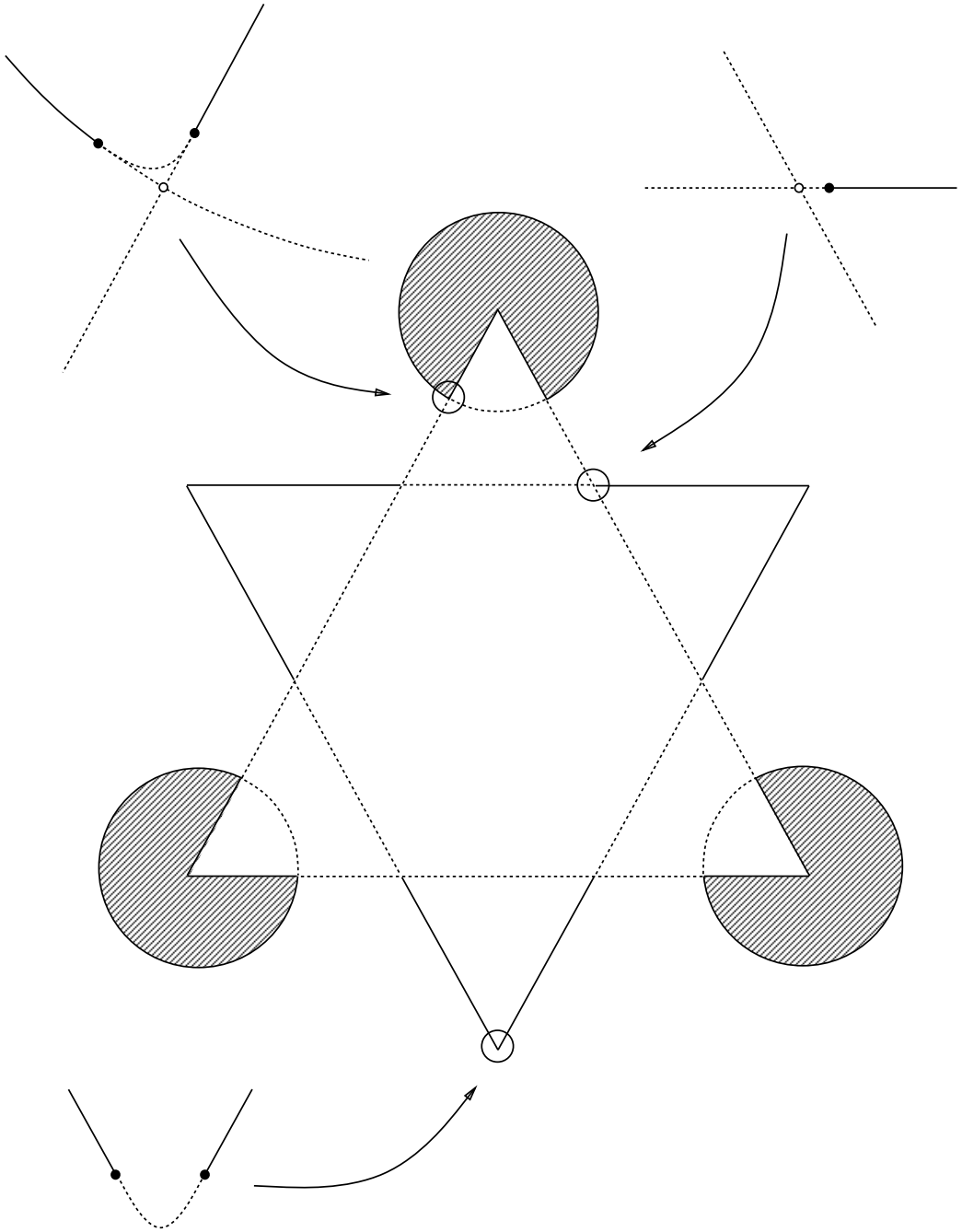

Figure 3.8: A graph labeling problem. This figure illustrates what $G_{planar} = (V_{endpoints} \cup V_{crossings}, E'_{fragments} \cup E'_{completions})$ looks like for the Kanizsa Triangle. The endpoints (i.e. $V_{endpoints}$), are drawn as filled circles and the crossings, (i.e. $V_{crossings}$), as non-filled circles. The boundary fragments (i.e. $E'_{fragments}$), are drawn as solid lines and the set of potential completions (i.e. $E'_{completions}$), as dotted lines.

a planar graph, G_{planar} is created by splitting the edges of $G_{non-planar}$ wherever two intersect, and creating a vertex at that point called a *crossing*. If $V_{crossings}$ is the set of crossings, and $E'_{fragments} \cup E'_{completions}$ is the set of edges after the splitting operation, then $G_{planar} = (V_{endpoints} \cup V_{crossings}, E'_{fragments} \cup E'_{completions})$. Figure 3.8 illustrates what G_{planar} looks like in the case of the Kanizsa Triangle. Here the endpoints (i.e. $V_{endpoints}$) are drawn as filled circles and the crossings (i.e. $V_{crossings}$) as non-filled circles. The boundary fragments (i.e. $E'_{fragments}$) are drawn as solid lines and the set of potential completions (i.e. $E'_{completions}$), as dotted lines.

The problem of maximizing (or minimizing) a linear objective function subject to linear inequality (or equality) constraints 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 (See Figure 3.9). Although integer linear programming, in its full generality, is NP-complete, there are specific instances, such as bipartite matching and maximum network flow, which have polynomial time algorithms. 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*.

3.4.1 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 surface lies to

$$\max \quad \mathbf{c}^\top \mathbf{x}$$

$$\text{subject to} \quad A\mathbf{x} \leqslant \mathbf{b}$$

$$\mathbf{x} \geqslant \mathbf{0}$$

$$\mathbf{x} \quad \text{integer}$$

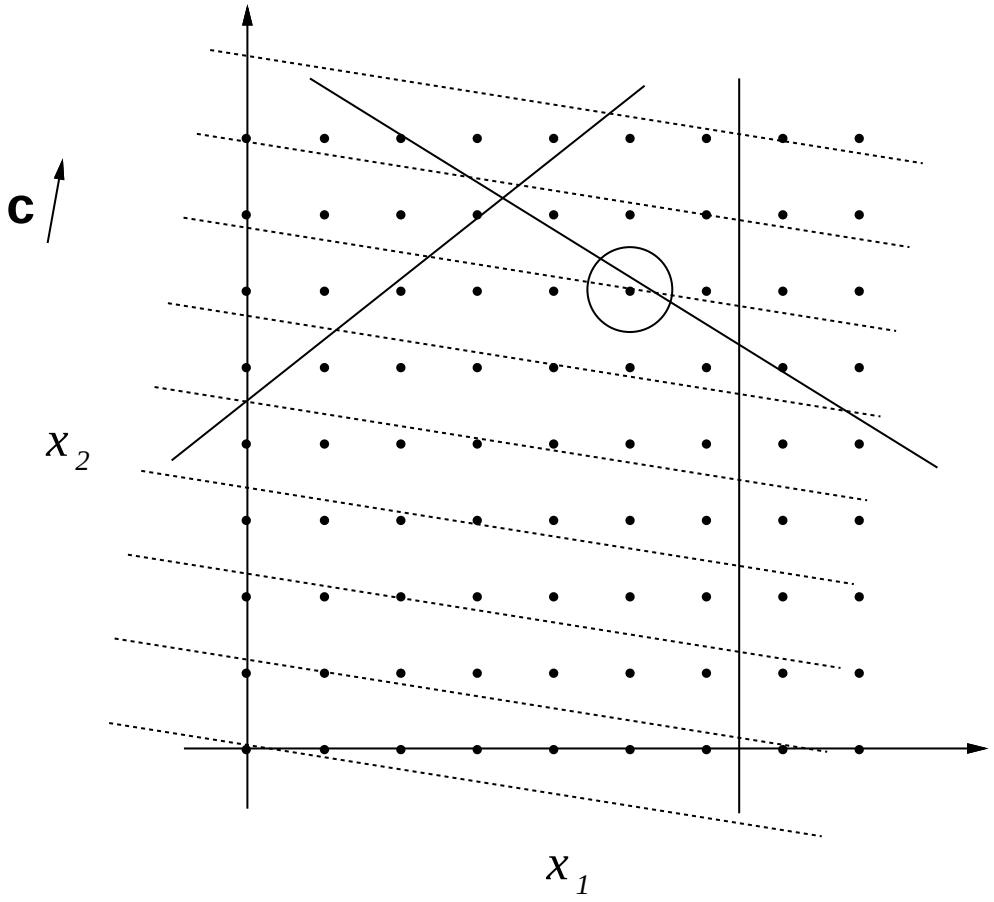

Figure 3.9: An integer linear program. The problem of maximizing (or minimizing) a linear objective function subject to linear inequality (or equality) constraints is termed a *linear program* (or *LP*). An *integer linear program* (or *ILP*) is a linear program where the solution is further constrained to have integer components.

the right as its boundary is traversed in the direction of its orientation.[6] 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, $p \in V_{endpoints}$:

$$x_p + x'_p \leq 1 \tag{3.3}$$

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, completions can not be instantiated independently of their "sponsoring" boundary fragments but can only join fragments to form closed boundaries. Finally, the sign of occlusion of the completion must be unique and compatible with the sign of occlusion of the sponsoring boundary fragments. This ensures that every cycle in G_{knot} has a unique sign of occlusion. Let $completions(p)$ be the potential completions of the boundary fragment through endpoint p (Figure 3.10). Two constraints per endpoint guarantee all of the above:

$$x_p = \sum_{j \in completions(p)} x_j \tag{3.4}$$

$$x'_p = \sum_{j \in completions(p)} x'_j \tag{3.5}$$

These constraints play a role analogous to those in a network flow problem. In this case, "sign of occlusion" is "conserved" at each endpoint. The right side of each

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

inequality is the sum of all completions whose sign of occlusion is consistent with the sign of the sponsoring boundary fragment on the left side. Since the left sides are 0-1 valued, the right sides are likewise bounded, guaranteeing a unique completion. When $x_p = x'_p = 0$, the right sides of both inequalities must also equal zero, which ensures that no completion is instantiated independently of a sponsoring boundary fragment. Conversely, a fragment can not form part of a surface boundary if it is not completed through both its endpoints.[7] This guarantees that all instantiated edges are part of cycles.

We now systematically present the linear constraints required to model the occlusion of one opaque surface by another. 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 (Figure 3.11). 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 needn't 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

[7] More formally, these constraints ensure that the neighborhood of every point of the completed boundary is homeomorphic to an open interval.

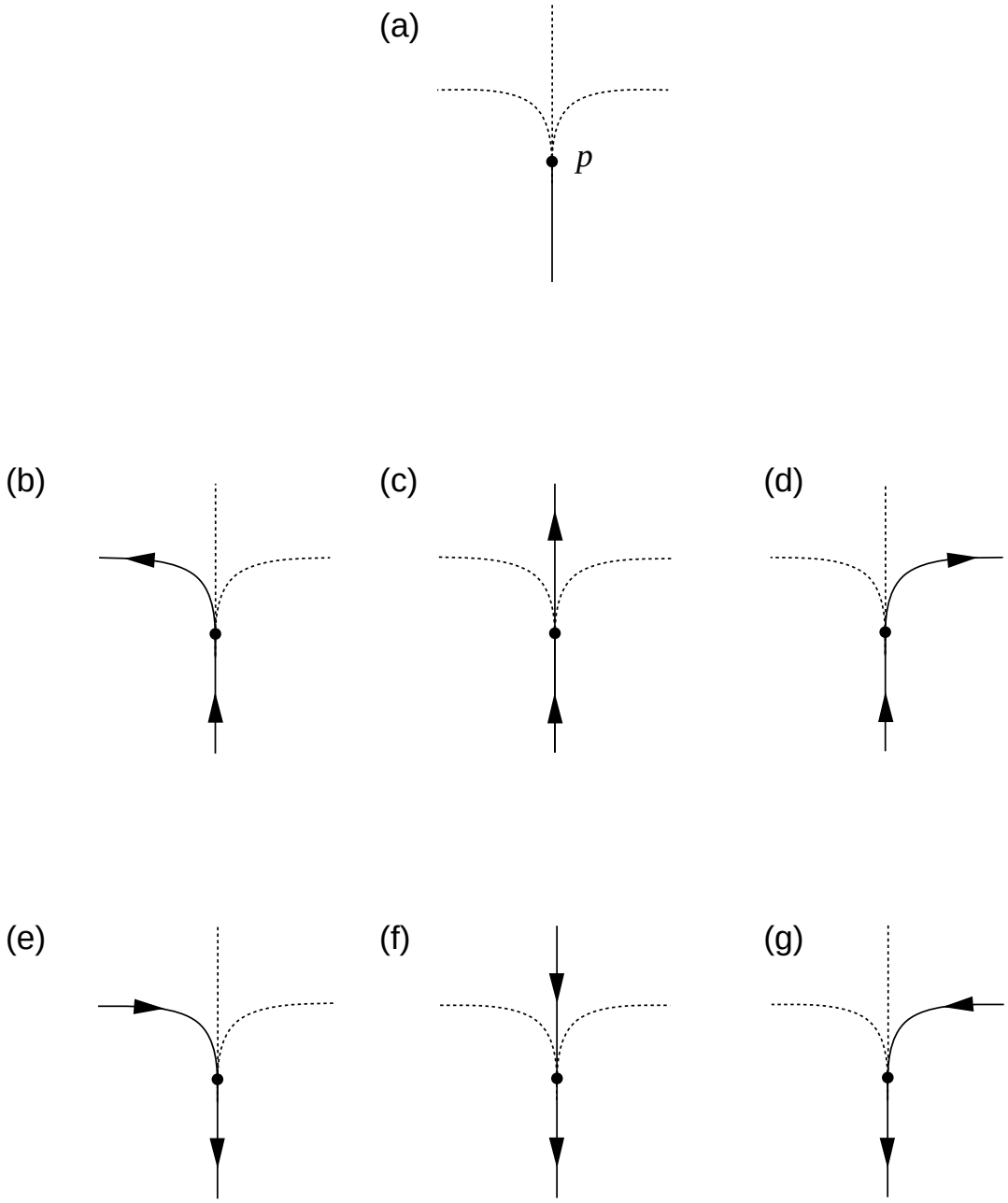

Figure 3.10: Potential completions. (a) Potential completions at endpoint p, of graph G_{planar}; (b),(c) and (d) Unique continuation in case where $x_p = 1$; (e),(f) and (g) Unique continuation in case where $x'_p = 1$.

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. $x_u = 1$).

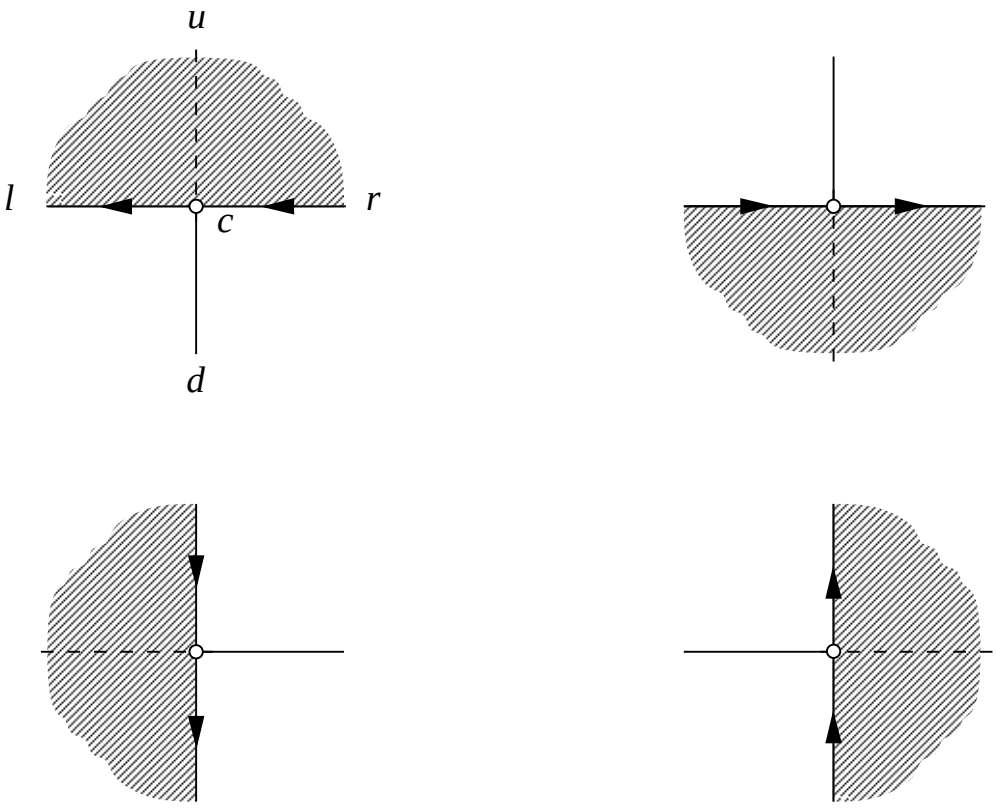

Figure 3.11: Four principal crossing states.

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 $c \in V_{crossing}$ (with adjacent edges, u, d, l and r) enforce the following:

$$x_u + x'_u + x_l + x'_l \leq x_c^{\top} + x_c^{\perp} + x_c^{\vdash} + x_c^{\dashv} + 1 \tag{3.6}$$

When u and l are instantiated, the left side of the inequality 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:

$$x_c^{\top} + x_c^{\perp} + x_c^{\vdash} + x_c^{\dashv} \leq 1 \tag{3.7}$$

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

$$x_c^{\top} \leq x'_l \tag{3.8}$$

$$x_c^{\perp} \leq x_l \tag{3.9}$$

$$x_c^{\vdash} \leq x'_u \tag{3.10}$$

$$x_c^{\dashv} \leq x_u \tag{3.11}$$

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:

$$n_u - n_d = x_c^{\top} - x_c^{\perp} \tag{3.12}$$

$$n_l - n_r = x_c^{\vdash} - x_c^{\dashv} \tag{3.13}$$

There is one additional requirement which must be explicitly enforced at every crossing vertex. This is the requirement that the depth of the occluding surface must be less than or equal to the depth of the occluded surface. In terms of the variables of the integer linear program, the following four conditions must be satisfied:

1. if $x_c^{\top} = 1$ then $n_u > n_l$
2. if $x_c^{\perp} = 1$ then $n_d > n_l$
3. if $x_c^{\vdash} = 1$ then $n_l > n_u$
4. if $x_c^{\dashv} = 1$ then $n_r > n_u$

The first of these conditions reflects the requirement that the depth of the surface occluding edge u must be less than the depth of edge u itself. The other three represent the same requirement for edges d, l and r. Unfortunately, there is no easy way to achieve this effect with integer linear inequalities. However, with a bit of creativity, the necessary constraint can be enforced for values of n less than some constant N:

$$x_c^{\top} - N(1 - x_c^{\top}) \leq n_u - n_l \tag{3.14}$$

$$x_c^{\perp} - N(1 - x_c^{\perp}) \leq n_d - n_l \tag{3.15}$$

$$x_c^{\vdash} - N(1 - x_c^{\vdash}) \leq n_l - n_u \tag{3.16}$$

$$x_c^{\dashv} - N(1 - x_c^{\dashv}) \leq n_r - n_u \tag{3.17}$$

Although these inequalities are not very elegant, it is easy to verify that in each of the four crossing states, they correctly enforce the depth order requirement. For example, when $x_c^\top = 1$, the left side of the first inequality equals one, and will only be satisfied by values of n_u greater than n_l.

3.4.2 Stimulus Conformity

Topological validity is a necessary but not a sufficient condition for feasibility of G_{knot}. For a solution to be feasible, it must also conform to the image evidence. Most importantly, since 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 $f \in E'_{fragments}$, we require that $n_f = 0$). This can be enforced by simply excluding all the n_f for these edges from the linear program, and need not increase the size of the constraint matrix.

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[48] words "the [perceptual] solution must not contradict the stimulus." Consider contour j which joins visible boundary fragments i and k (possibly through an arbitrary number of additional edges and crossings; see Figure 3.12). Let x_j be the sign of occlusion of edge j corresponding to the $i \rightarrow k$ direction and x'_j be the opposite sign. Depending on its sign of occlusion, edge j bounds either surface $\mathcal{S}_j$ or surface $\mathcal{S}'_j$. Associated with $\mathcal{S}_j$ is reflectance φ_j and with $\mathcal{S}'_j$ is reflectance φ'_j. Assuming roughly uniform illumination, φ_j and φ'_j can be approximated by the average brightness within narrow regions on either side of edge j. In a similiar manner, we can compute φ_i, φ'_i, φ_k and φ'_k associated with surfaces $\mathcal{S}_i$, $\mathcal{S}'_i$, $\mathcal{S}_k$ and $\mathcal{S}'_k$.

Following Rock[48], 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. Let $|\Delta\varphi_j|$ be the magnitude of the brightness gradient across contour j, then:

$$|\Delta\varphi_j| \;=\; |\varphi_j - \varphi'_j| \tag{3.18}$$

When $x_j = 1$ and $n_j = 0$ then j joins i and k to form the visible boundary of a single surface: $\mathcal{S}_i \equiv \mathcal{S}_j \equiv \mathcal{S}_k$. Assuming roughly uniform illumination, we conclude that $\varphi_i \approx \varphi_j \approx \varphi_k$. Conversely, if $x'_j = 1$ and $n_j = 0$, then j joins i and k to form the visible boundary of surface $\mathcal{S}'_i \equiv \mathcal{S}'_j \equiv \mathcal{S}'_k$. We conclude that $\varphi'_i \approx \varphi'_j \approx \varphi'_k$. Since $n_j = 0$, the assumption is that no surfaces lie between the boundary and the eye or camera. Normally, under these circumstances, the change in reflectance across the surface boundary produces a luminance boundary in the image. However, when a surface occludes another surface with similiar reflectance then $\varphi_j \approx \varphi'_j$ and $|\Delta\varphi_j| \approx 0$. This suggests that illusory contours should be permitted only where the observed image brightnesses are consistent with the existence of two overlapping surfaces of roughly constant and approximately equal reflectance. This effect can be achieved by adding Constraint 3.19 unless $\varphi'_j \approx \varphi_i \approx \varphi_j \approx \varphi_k$ and adding Constraint 3.20 unless $\varphi_j \approx \varphi'_i \approx \varphi'_j \approx \varphi'_k$:

$$x_j \;\leq\; n_j \tag{3.19}$$

$$x'_j \;\leq\; n_j \tag{3.20}$$

A "pacman" from the Kanizsa triangle can be used to illustrate the visibility constraints which are written in the three possible cases (see Figure 3.13). First, because contour i is a visible boundary fragment, its depth is constrained to be zero (i.e. $n_i = 0$). Second, because the completion represented by contour j (with the sign of occlusion indicated) satisfies the conditions described above for an illusory

contour, 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. $x_k \leq n_k$). If instantiated, the completion must be amodal.

3.4.3 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. *verdicality*), then it is important that the predictions be equally specific. In our analysis, events are represented by 0-1 valued expressions. 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. We propose that the prediction consist of an assignment to each element of the set of potential completions, the label "accidental" or "non-accidental." Since the set of potential completions is of fixed size, these predictions are equally specific.

3.4.3.1 Unit Preference

In any given surface organization, some subset of completions is instantiated while a complementary set remains uninstantiated. Recall that one of the motivations for using curves of least energy to represent potential completions lay in a presumed

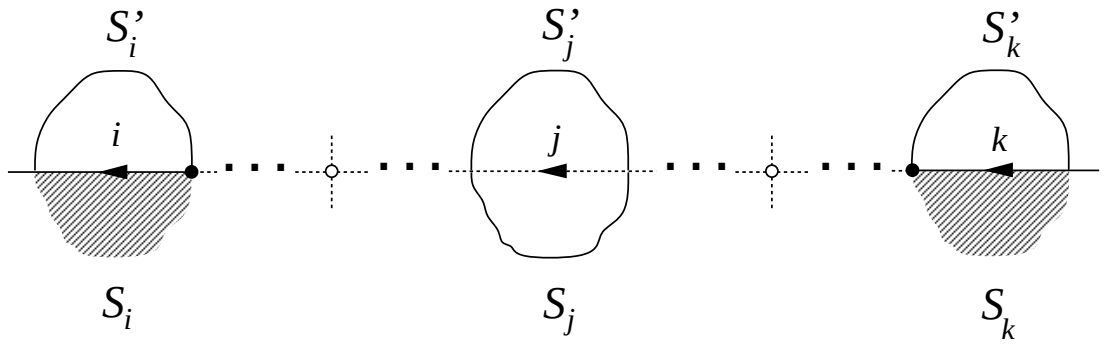

Figure 3.12: Brightnesses adjacent to visible boundary fragments and completion. This information, together with the sign of occlusion, determines whether a completion will be modal or amodal.

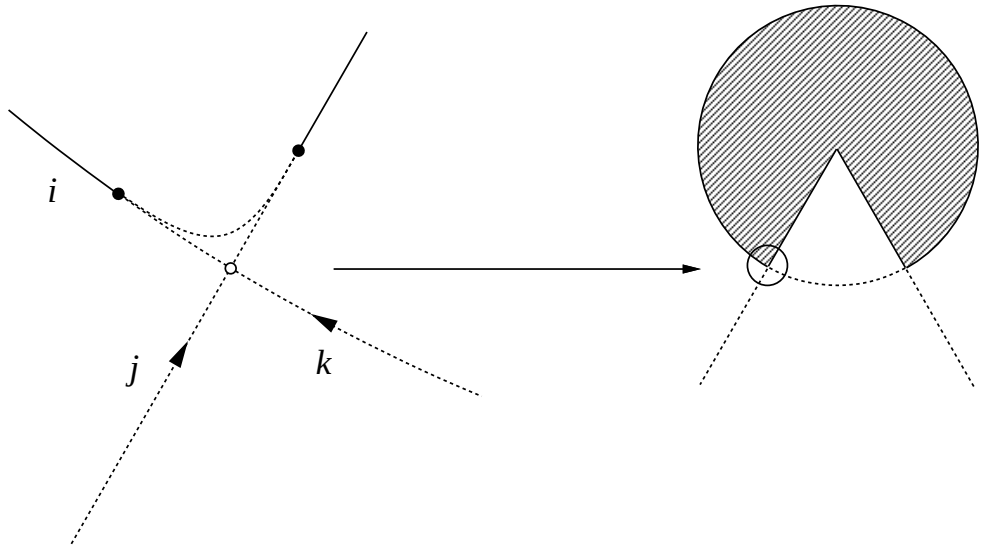

Figure 3.13: Examples of visibility contraints. A "pacman" from the Kanizsa triangle is used to illustrate the visibility constraints which are written in the three possible cases. First, because contour i is a visible boundary fragment, its depth is constrained to be zero (i.e. $n_i = 0$). Second, because the completion represented by contour j (with the sign of occlusion indicated) satisfies the conditions described in the text for illusory contours, 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. $x_k \leq n_k$). If instantiated, the completion must be amodal.

relationship between the shape of the curve of least energy and the likelihood that two contour fragments have a common non-accidental origin (i.e. the likelihood that they form consecutive segments of a single boundary). It was assumed that shape and completion likelihood were related through an unspecified probability density function.[8]

The instantiation of a completion is, in effect, an assertion that its origin is *non-accidental.* Since a completion is instantiated if and only if $x_i + x_i' = 1$, the probability that completion i's origin is non-accidental 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 non-accidental is independent of whether any other completion's origin is accidental or non-accidental. 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:

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

[8]For the purposes of the experimental implementation, specific shape features and probability densities were chosen. These are described in the next chapter.

This function is, in turn, a monotonically increasing function of the sum of the logarithms of the likelihoods, so that maximimizing 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:

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

In information theory, *surprise* is used as a measure of the amount of information the news of an event carries. For example, because it is less likely to be raining in Las Vegas than Seattle, there is more information in the statement "It is raining in Las Vegas" than in "It is raining in Seattle." If one expects to see a certain event, x, with probability, $Pr(x)$, then the degree of surprise when one doesn't see x is $-\ln(1 - Pr(x))$. The surprise at not seeing x is infinite when $Pr(x) = 1$ and zero when $Pr(x) = 0$.

The surprise metaphor appears frequently in theories of grouping. For example, Rock[48] suggests that: "The perceptual system detects continuity of direction among contours. Once doing so, not to accept two or more elements as part of a larger entity is to accept that continuity as the result of coincidental placement in space of these elements, that is, of elements that have no intrinsic relationship to one another." A similiar concept of "non-accidentalness" has been emphasized by Lowe[37]. By instantiating a completion, we are accounting for its coincidence of shape (i.e. we are attributing the origin of the two boundary fragments it joins to a single physical process). If a completion is not instantiated, we are asserting that the two fragments

originate in independent physical processes. The coincidence of shape represented by the potential completion is therefore attributed to chance. The $\ln(1 - Pr(x_i + x'_i = 1))(1 - x_i - x'_i)$ term in objective function Q_{unit} is a measure of the surprise implicit in the failure of a surface organization to account for the coincidence of shape represented by potential completion i.

We observe that objective function Q_{unit} simultaneously maximizes likelihood and minimizes surprise. Consequently, variable weights range from negative infinity to positive infinity. In comparison, the log-likelihood weighting function which Ullman[58] proposed for the correspondence problem in apparent motion uses exclusively negative weights (i.e. log-likelihoods). While maximizing a linear objective function with exclusively negative weights would be adequate under the assumption that every boundary fragment appears in the solution (i.e. for every fragment endpoint p, $x_p + x'_p = 1$), it is not adequate if every boundary fragment need not be instantiated (i.e. for every fragment endpoint p, $x_p + x'_p \leq 1$). In the latter case, the objective function is always minimized by the solution where no boundary fragments are instantiated, that is, by the zero vector, which is always in the feasible region. This is due to the fact there is no penalty for failing to account for observed spatial coincidence. Yet, minimizing surprise alone is also problematic, since this policy favors organizations which incorporate large numbers of completions, regardless of overall likelihood. Stated differently, it encourages highly unlikely explanations for what are potentially only negligible coincidences. For these reasons, it seems that the objective function employed here is better than an objective function based upon surprise or likelihood alone.

As it currently stands, this analysis is incomplete, since it does not adequately account for the phenomenon of illusory contours and the difference in human perception of contrast and outline stimuli[29]. Compare Figure 3.14(a), in which an

illusory triangle is readily perceived, to Figure 3.14(c). Although the closed contours in Figure 3.14(c) correspond to the boundaries of the black regions in Figure 3.14(a), they do not trigger the percept of an illusory triangle. A second demonstration of this phenomenon is provided by Figure 3.14(b) and Figure 3.14(d). While Figure 3.14(b) is perceived as two overlapping bars, Figure 3.14(d) is perceived as a cross in outline. Finally, Figure 3.14(e) and Figure 3.14(f) demonstrate that figural completion does play a role in the organization of outline stimuli, but seems to governed by different logic depending on whether contrast or outline stimuli are involved.

It has been known since Hubel and Wiesel[25] that simple cells in the human visual cortex have oriented receptive fields of either odd or even symmetry. Hubel and Wiesel conjectured that these neurons are feature detectors which respond to edges and bars (depending on their symmetry) of different orientations. More recently, it has been hypothesized that figural completion phenomena are the result of integrating simple cell responses over extended spatial areas (see for example [15]). A possible explanation for the qualitatively different completion effects in contrast and outline stimuli is that the output of cortical simple cells with odd and even receptive fields are integrated separately and in different ways. Although this may represent an explanation at the "implementation level" it clearly begs the question of "why it should be so."

Kanizsa[29] attributed the discrepancy to the (apparent) "incompleteness" of elements in the contrast stimulus (e.g. discs with missing sectors) and the "completeness" of the same elements in the outline stimulus (e.g. closed contours). However, in this instance, Kanizsa seems to be guilty of what he himself calls the "stimulus error," that is, he confuses stimulus and percept.

To identify a possible explanation for the difference between contrast and outline stimuli, let us again consider Figure 3.3, this time for the purpose of determining the

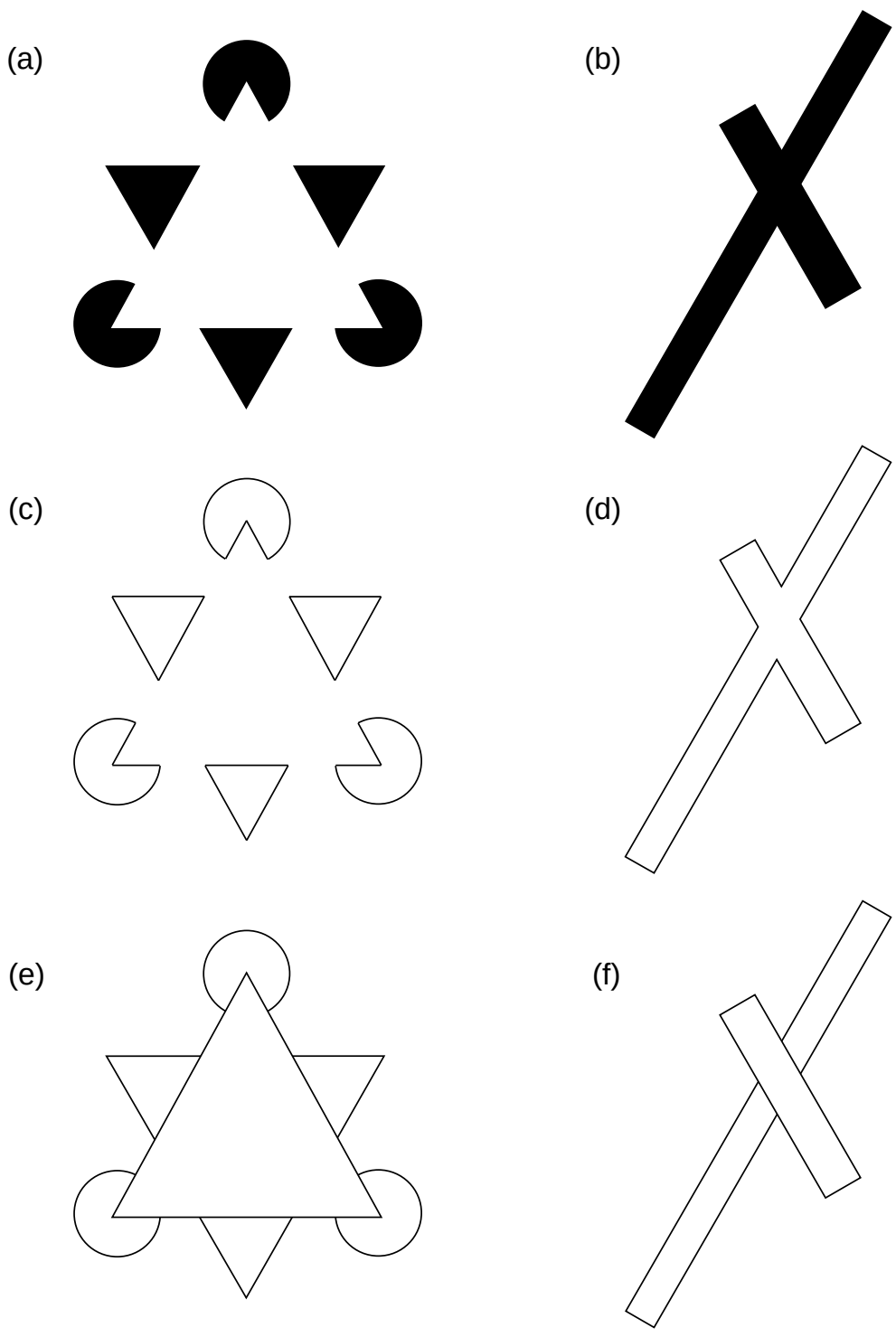

Figure 3.14: Figural completion in contrast and outline stimuli. Although figural completion does play a role in the organization of outline stimuli, it seems to be governed by different logic depending on whether contrast or outline stimuli are involved.

different degrees of surprise implied by the two *unit organizations*.[9] Figure 3.15(b) and (c) show magnified views of one neighborhood (i.e. the area within the circle in Figure 3.15(a)) in the two unit organizations. In the "literal solution" (i.e. Figure 3.15(b)) the only completion instantiated is the "corner" (i.e. 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:

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

Now compare this to the "illusory contour solution" where only completions j and k are instantiated (Figure 3.15(b)). If we assume that because completion i is not instantiated, the spatial coincidence it represents is attributed to chance[10] 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 surprise 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:

[9]The *unit organization* consists of the subset of potential completions which are instantiated in the *surface organization*.

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

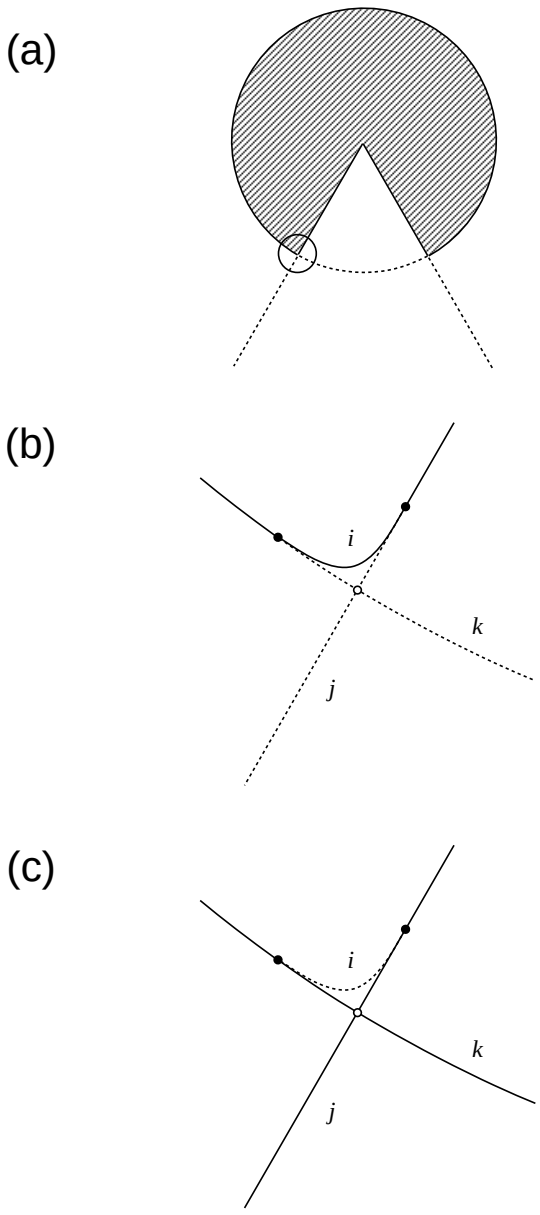

Figure 3.15: Two possible unit organizations shown in magnified view. (b) Rock's "literal solution." The spatial coincidence represented by completions j and k is attributed to chance, so that the degree of surprise (i.e. $S_{literal}$) is $-\ln(1-Pr(x_j+x'_j=1))-\ln(1-Pr(x_k+x'_k=1))$. (c) The "illusory contour solution" offers a unified explanation for all of the completions, even though the corner is not instantiated. The degree of surprise (i.e. $S_{illusory}$) is therefore zero.

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

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 $Pr(1 - x_i - x'_i)$. This assumption apparently does not hold for outline stimuli. Presumably, in the physical model underlying outline stimuli, overlapping boundaries image as "T-junctions," and never as corners. Accordingly, the spatial coincidence represented by a corner can only be accounted for by the "literal solution" (i.e. the solution in which the corner is instantiated). This suggests that objective function Q_{unit} should be used for outline stimuli and that Q'_{unit} should be used for contrast stimuli.

3.4.3.2 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 equivalence 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 was demonstrated by Figure 3.4, a given unit organization will usually admit many different depth labelings, so that the majority of these equivalence classes will contain many distinct (athough 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[48] 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 figure. Similiarly, 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.

Another very interesting phenomenon, important in preference, is known as the *Petter effect*[43]. The Petter effect occurs when two surfaces of equal reflectance overlap. Because the reflectances of the two surfaces are the same, their relative depth can not be determined from figural information alone. Even so, there is a strong tendency to see the broader of the two surfaces in front of the narrower (i.e. the longer completion is perceived amodally).

A dramatic example of the Petter effect, designed by Kanizsa[29], is shown in Figure 3.16. 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."

At first glance, it appears to be a simple matter to incorporate both the additional figure-ground preferences and the preference underlying the Petter effect into a new objective function:

$$Q'_{surface} = \alpha\, Q'_{unit} + \beta\, Q_{depth} \tag{3.21}$$

Clearly, the variables required to form 0-1 valued expressions representing the competing depth labelings exist in the integer linear program. Specifically, $x_i = 1$ and

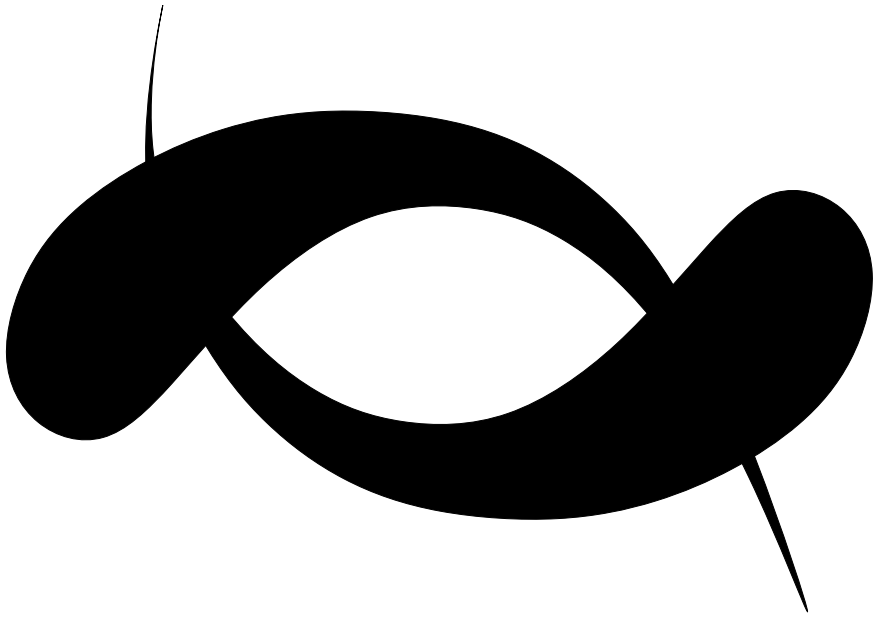

Figure 3.16: Kanizsa's "Paisleys." Although interpretations as surfaces embedded in planes at constant depth are plausible, they are not perceived. As Kanizsa points out, the slender "tail" of each surface seems to pass under the "head." This is an example of the *Petter effect.*

$x_i' = 1$ represent the two possible signs of occlusion, and $x_c^{\top} + x_c^{\perp} = 1$ and $x_c^{\vdash} + x_c^{\dashv} = 1$ represent the two possible depth orderings at a crossing. 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 its sign of contrast is presumably straightforward. Nor would it be difficult (in principle) to characterize the relative likelihood of a crossing's two possible depth orderings as a function of the lengths of the intersecting completions. The problem, once again, is ensuring that the objective function interprets all points in the feasible region as equally specific predictions.

Consider the two predictions: $x_i + x_i' = 0$ and $x_i + x_i = 1$. The first prediction is that the spatial coincidence represented by contour i is accidental while the second prediction is that the spatial coincidence represented by contour i is non-accidental. Let us assume that they are equally likely (each is correct with probability 0.5). Because these two predictions are equally specific, their likelihoods can be compared. Now consider the two predictions: $x_i + x_i' = 0$ and $x_i = 1$. If both signs of occlusion are also equally likely, then the probability that $x_i + x_i' = 0$ is still 0.5, but the probability that $x_i = 1$ is only 0.25. Maximizing the sum of the logarithms of the likelihoods would be inappropriate because the second prediction is less likely only because it is more specific. Because they are not equally specific, the likelihoods of these two predictions cannot be compared.

In summary, although it would be a simple matter to assign weights to variables of the integer linear program, and in doing so, mimic the figure-ground preferences of the human visual system, it is difficult to do this in a way that does not introduce a bias for (or against) organizations with larger numbers of completions. Solving this problem is a subject for future work.

CHAPTER 4

EXPERIMENTAL SYSTEM

Here we depart from the natural computation[45] methodology outlined in the first chapter. Although algorithms and representations implementing the computational theory *are* the primary focus of this chapter, we make no claims as to their biological plausibility. Instead, our intention is to validate the computational theory by demonstrating that a well defined procedure for computing the mapping does in fact exist.[1]

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[6] 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 Figure 4.1.

4.1 Minimum Energy Cubic Bezier Splines

In the design of the experimental system, practical considerations like simplicity, economy and efficiency were stressed. For simplicity's sake, cubic Bezier splines of

[1] In particular, the experimental system is not intended to be a theory at what Marr[38] termed "the level of algorithm and representation."

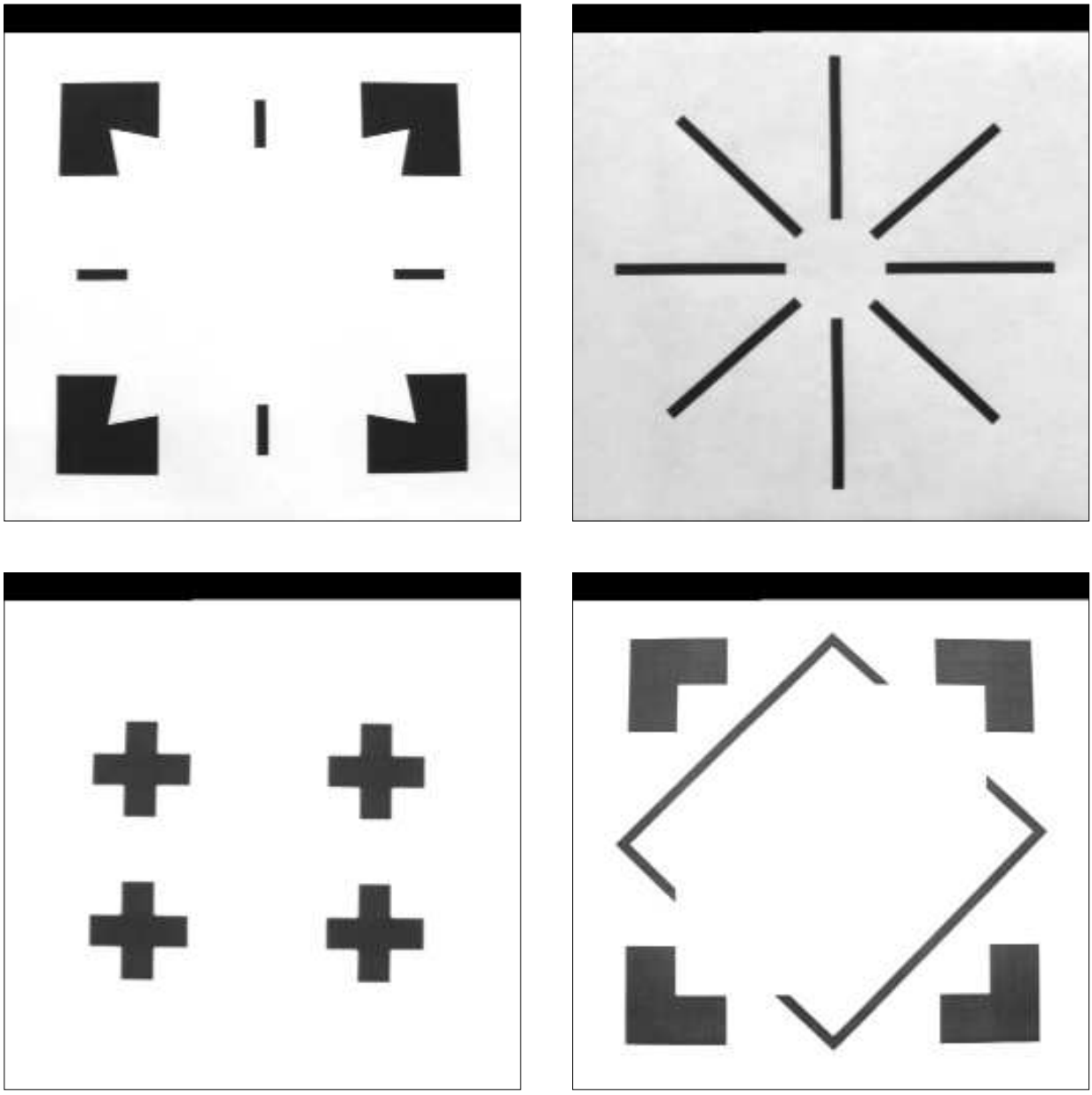

Figure 4.1: Four test figures from the illusory contour literature. Clockwise from upper left: *Warped Square*, *Ehrenstein*, *Kanizsa Plusses* and *Woven Square*.

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 non-accidental origin.

Since the shape of a cubic Bezier spline is determined by only four points, the use of splines also satisfied the need for an economical representation. Although the minimum energy cubic Bezier spline does not possess the optimal substructure property which makes efficient parallel computation possible, computing its control polygon on a non-parallel computer is considerably faster than computing a point vector approximating the true curve of least energy. Finally, using cubic Bezier splines to represent all contours in the system (both completions and boundary fragments) allowed procedures for display, spatial querying and intersection to treat contours in a uniform fashion, which simplified system implementation.

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 Figure 4.2). 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. The first and fourth control points, b_0 and b_3 are simply the ends of the fragments (i.e. $b_0 = p$ and $b_3 = p'$). The locations of the second and third control points control the spline's orientation at either end. If the orientation of the spline and the boundary fragment are to match at point p, then the second control point, b_1, must lie along the line passing through p and tangent to $\hat{t}$. Similiarly, the third control point, b_2, is constrained to lie along the line passing through p' with tangent $\hat{t}'$. Because the locations of b_1 and b_2 are not completely determined, there will be (in general) a two parameter family of smooth cubic splines passing through

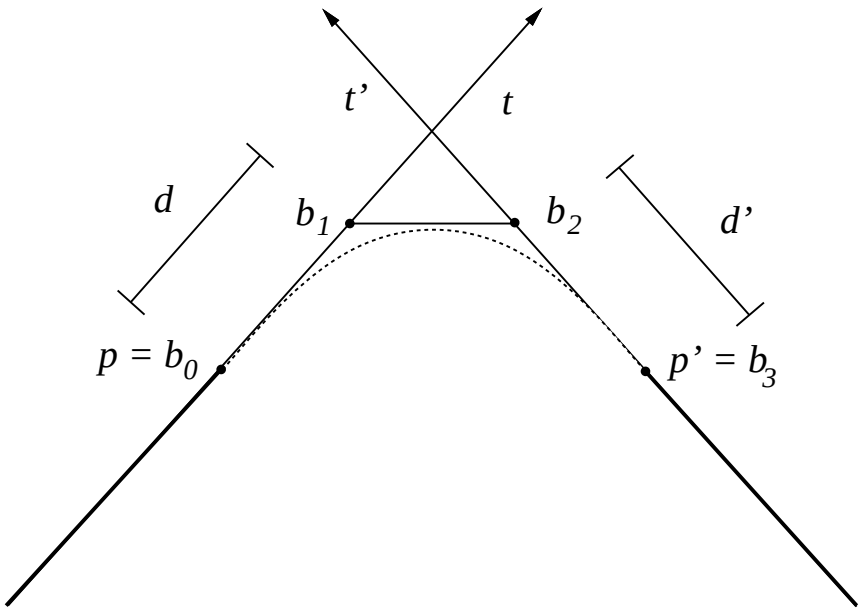

Figure 4.2: A cubic Bezier spline. In general, there is a two parameter family of smooth cubic splines passing through any two points p and p' with specified tangents. The free parameters are the distance between b_1 and b_0 in the direction of $\hat{t}$ (i.e. $d = |b_1 - b_0|$) and the distance between b_3 and b_2 in the direction of $\hat{t'}$ (i.e. $d' = |b_3 - b_2|$).

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, $E = \int \kappa(s)^2 \, ds$, 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 [44]. Because the shape of the minimum energy cubic spline is invariant to translation, rotation and scale change, a table of d and d' values minimizing the bending energy of a cubic Bezier spline passing through any two points in the plane with arbitrary tangents can be precomputed and indexed by two angles (see Figure 4.3). This results in significant run-time savings.

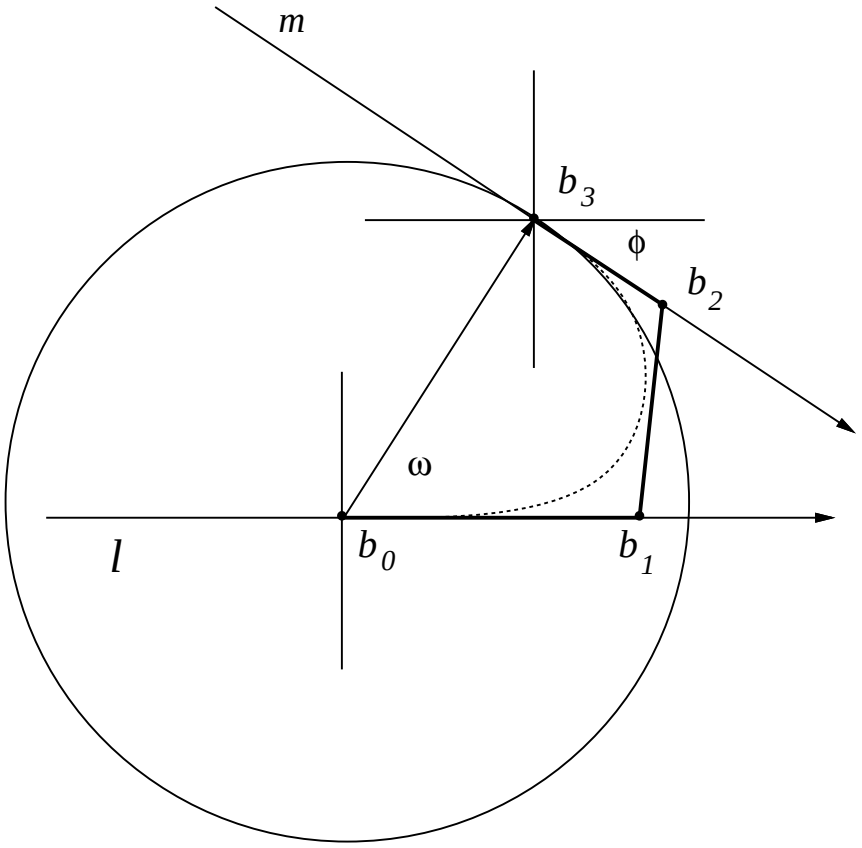

Figure 4.3: Two angle parameterization. A table of d and d' values minimizing the bending energy of a cubic Bezier spline passing through two points in the plane with arbitrary tangents can be precomputed and indexed by two angles. This is a consequence of the invariance of the shape of the minimum energy cubic spline to translation, rotation and scale change.

4.2 Completion Features and Categories

If the world only contained objects with smooth boundaries, then the figural completion problem would be greatly simplified, since all orientation discontinuities (i.e. "corners") would be artifacts of occlusion. In this world, the probability density of completion shape features could potentially be quite simple (e.g. unimodal). Unfortunately, in the real world, the problem is more complicated because there is no way of knowing whether a corner is simply an artifact of occlusion or whether it is

the image of an orientation discontinuity in a surface boundary. Consequently, 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 of completion shape features is even more complex.

The approach adopted here is consistent with a recent theory of perceptual categories proposed by Richards, Feldman, and Jepson[46]. Richards *et al.* argue that the categories used by the human visual system to classify geometric configurations of a given type (e.g. triples of points, pairs of line segments, etc.) form a natural hierarchy. Assuming that the geometric configuration can be parameterized in the generic case by k parameters, then the hierarchy is formed by systematically removing degrees of freedom from the generic (i.e. *tranverse*) configuration to create categories representing degenerate (i.e. *non-tranverse*) configurations of increasing *codimension*. The usefulness of their theory follows from an assumption that, within a given context, a subset of this hierarchy models the physical processes underlying image structure. Under this assumption, there will be peaks (or *modes*) in the feature probability densities corresponding to one or more of the perceptual categories. Figure 4.4 illustrates the perceptual categories predicted by Richards *et al.* for planar configurations of two line segments together with the subset of categories used in the experimental system described here.

Consider the conditional probability densities of completion shape features given gaps caused by:

- Orientation discontinuity (i.e. "corner")
- Partial occlusion of a straight boundary (i.e. "straight")
- Generic occlusion (i.e. "generic")

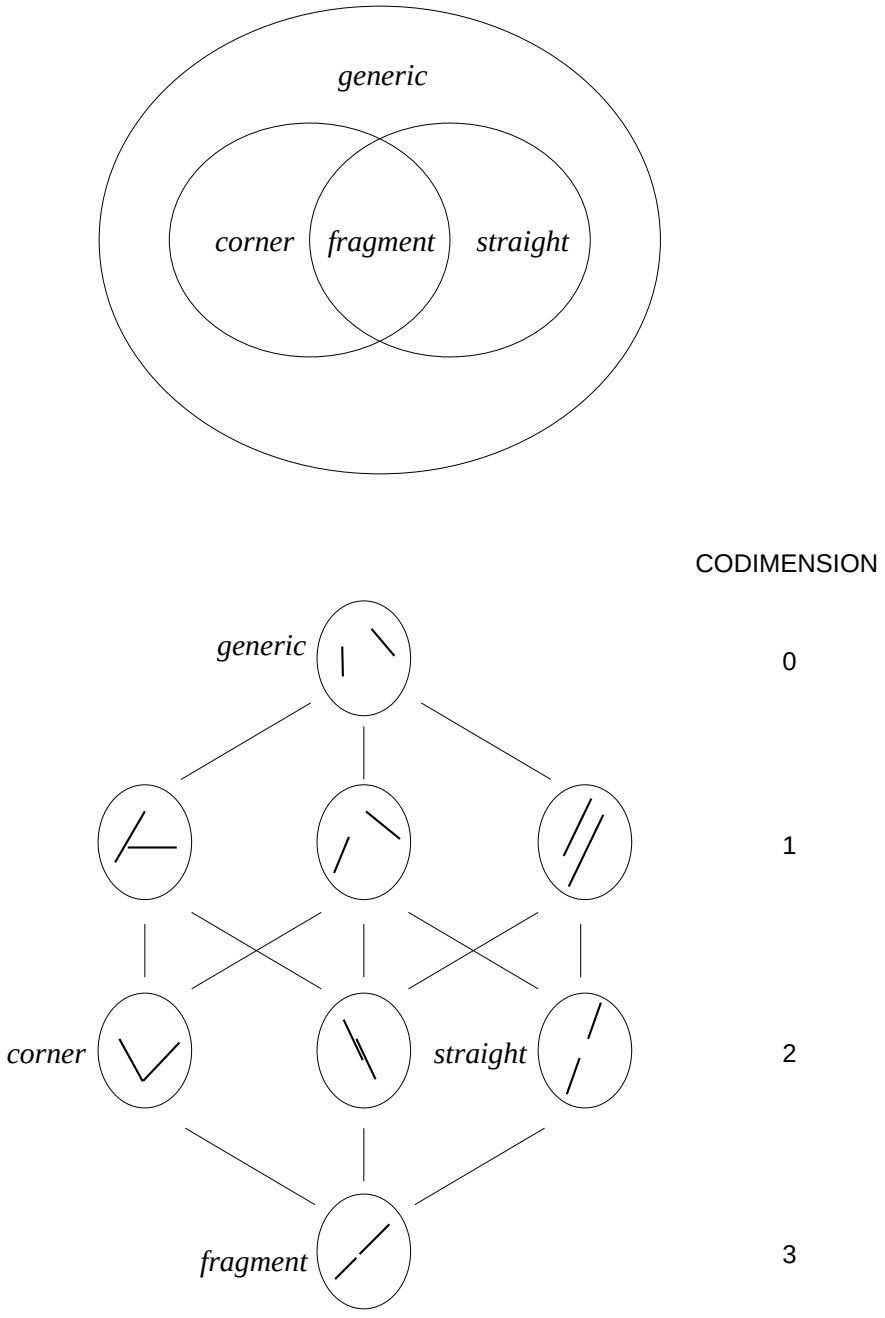

Figure 4.4: Perceptual categories. The perceptual categories predicted by Richards *et al.* for planar configurations of two line segments together with the subset of categories used in the experimental system described here. In the case of two straight line segments, the geometric configuration can be parameterized in the generic case by three parameters. The hierarchy is formed by systematically removing degrees of freedom from the generic (i.e. *tranverse*) configuration to create categories representing degenerate (i.e. *non-tranverse*) configurations of increasing *codimension*.

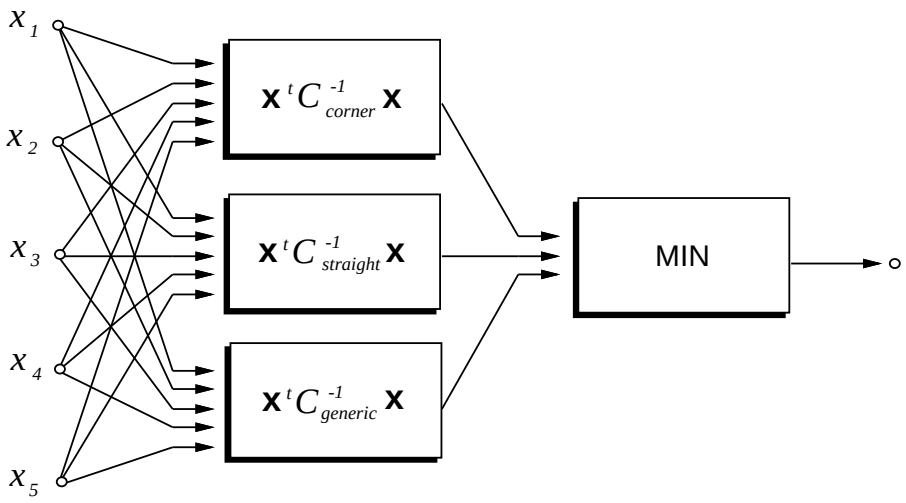

Figure 4.5: A Bayes classifier. The category minimizing squared Mahalanobis distance between observed and mean shape feature values (with zero mean) is used to compute a heuristic measure of likelihood.

Although the likelihood that a gap between two fragments is non-accidental 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. Similiarly, 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,[2] it proved useful to divide the set of completions into three categories and employ a multicategory *Bayes classifier*[10] to select among them (see Figure 4.5). The category minimizing squared *Mahalanobis distance*[10] (i.e. covari-

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

ance weighted Euclidean distance) between observed and mean shape feature values (with zero mean) is used to compute a heuristic measure of likelihood:

$$Pr(completion \mid \mathbf{x}) = \exp(-\min(\mathbf{x}^t C^{-1}_{corner}\mathbf{x}, \mathbf{x}^t C^{-1}_{straight}\mathbf{x}, \mathbf{x}^t C^{-1}_{generic}\mathbf{x})) \tag{4.1}$$

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

$$x_1 = \frac{\int \kappa(s)^2\, ds}{|b_0 - b_3|} \tag{4.2}$$

$$x_2 = \int ds \tag{4.3}$$

$$x_3 = \frac{l + \int ds + l'}{\int ds} \tag{4.4}$$

$$x_4 = |\theta - \theta'| \tag{4.5}$$

$$x_5 = \begin{cases} 0 & \text{if } \kappa(0)\kappa(\int ds) \geq 0 \\ 1 & \text{otherwise} \end{cases} \tag{4.6}$$

Here x_1 is the total bending energy normalized by the distance between the fragment endpoints, x_2 is gap size,[3] 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:

$$C_{corner} = \begin{bmatrix} 1.0 \times 10^9 & 0 & 0 & 0 & 0 \\ 0 & 5.0 & 0 & 0 & 0 \\ 0 & 0 & 0.05 & 0 & 0 \\ 0 & 0 & 0 & 1.0 \times 10^9 & 0 \\ 0 & 0 & 0 & 0 & 1.0 \times 10^{-9} \end{bmatrix} \tag{4.7}$$

[3]Elder and Zucker[11] investigate the relationship between number and size of boundary gaps and the degree of perceptual closure. They measure latency time in a visual search task where the target and distractor elements are incomplete outlines figures. The total length of missing boundary is distributed either uniformly or in a single gap. They conclude that the sum of the squares of the gap lengths models the relationship between the number and size of boundary gaps and the degree of perceptual closure. This is consistent with the use of squared Mahalanobis distance as a measure of log-likelihood.

$$C_{straight} = \begin{bmatrix} 0.001 & 0 & 0 & 0 & 0 \\ 0 & 2500.0 & 0 & 0 & 0 \\ 0 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0.001 & 0 \\ 0 & 0 & 0 & 0 & 1.0 \times 10^{9} \end{bmatrix} \tag{4.8}$$

$$C_{generic} = \begin{bmatrix} 5.0 & 0 & 0 & 0 & 0 \\ 0 & 1000.0 & 0 & 0 & 0 \\ 0 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & \frac{\pi}{5} & 0 \\ 0 & 0 & 0 & 0 & 1.0 \times 10^{-9} \end{bmatrix} \tag{4.9}$$

In the heuristic likelihood function, C_{corner}, $C_{straight}$ and $C_{generic}$ play the role of 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 all or nothing "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.

Similiarly, 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.

4.3 Building the Graphs

Although the graphs G_{input}, $G_{non-planar}$, G_{planar} and G_{knot} are important parts of the "in principle" problem formulation, until now there has been no mention of specific representations or methods of construction. Indeed, descriptions at this level are not properly part of a computational theory. However, they are important for understanding its implementation in the experimental system.

4.3.1 Boundary Fragments

The straight line segments which serve as input to the experimental system are produced by Boldt's zero crossing grouping algorithm[6]. This algorithm is quite elegant and definitely worth studying, if only because its recursive formulation addresses the difficult problem of multiple spatial scales in an intelligent manner. Although more costly to run than simpler algorithms, its value is appreciated by the many members of the computer vision group at the University of Massachusetts who have applied

it to a large number of problems requiring stable descriptions and accurate spatial localization[7, 36, 53].[4]

Boldt's algorithm begins by creating an initial set of unit length line segments with orientation normal to the direction of maximum gradient along zero crossing contours of the Laplacian convolved image. The endpoints of these initial line segments form the vertices of a graph, which Boldt calls the *link graph*. The edges of the link graph are the line segments themselves. To the link graph, a second set of edges which Boldt calls *links* are added. Each link represents the potential merging of two line segments through the endpoints it connects. Some of the geometric relations used as *linking criteria* are endpoint proximity, orientation difference, lateral distance, segment overlap and contrast difference. All paths in the *link graph* within a fixed *replacement radius* around each line segment are enumerated. The segments along the path minimizing the mean-squared-error of a straight line fit are replaced by a new line segment. The algorithm is then invoked recursively on the new set of line segments, using a larger replacement radius, resulting in ever smaller sets of increasingly longer lines. Boldt's algorithm typically requires about fifteen link and replace cycles to produce straight lines spanning an image of size 256×256. The line segments are then filtered on length and contrast.

4.3.2 Adding Potential Completions

After correctness, the primary consideration in building a practical system is computational complexity. 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

[4] In his forthcoming dissertation, Dolan[9] studies the problem of generalizing Boldt's algorithm to compute image curves other than straight lines.

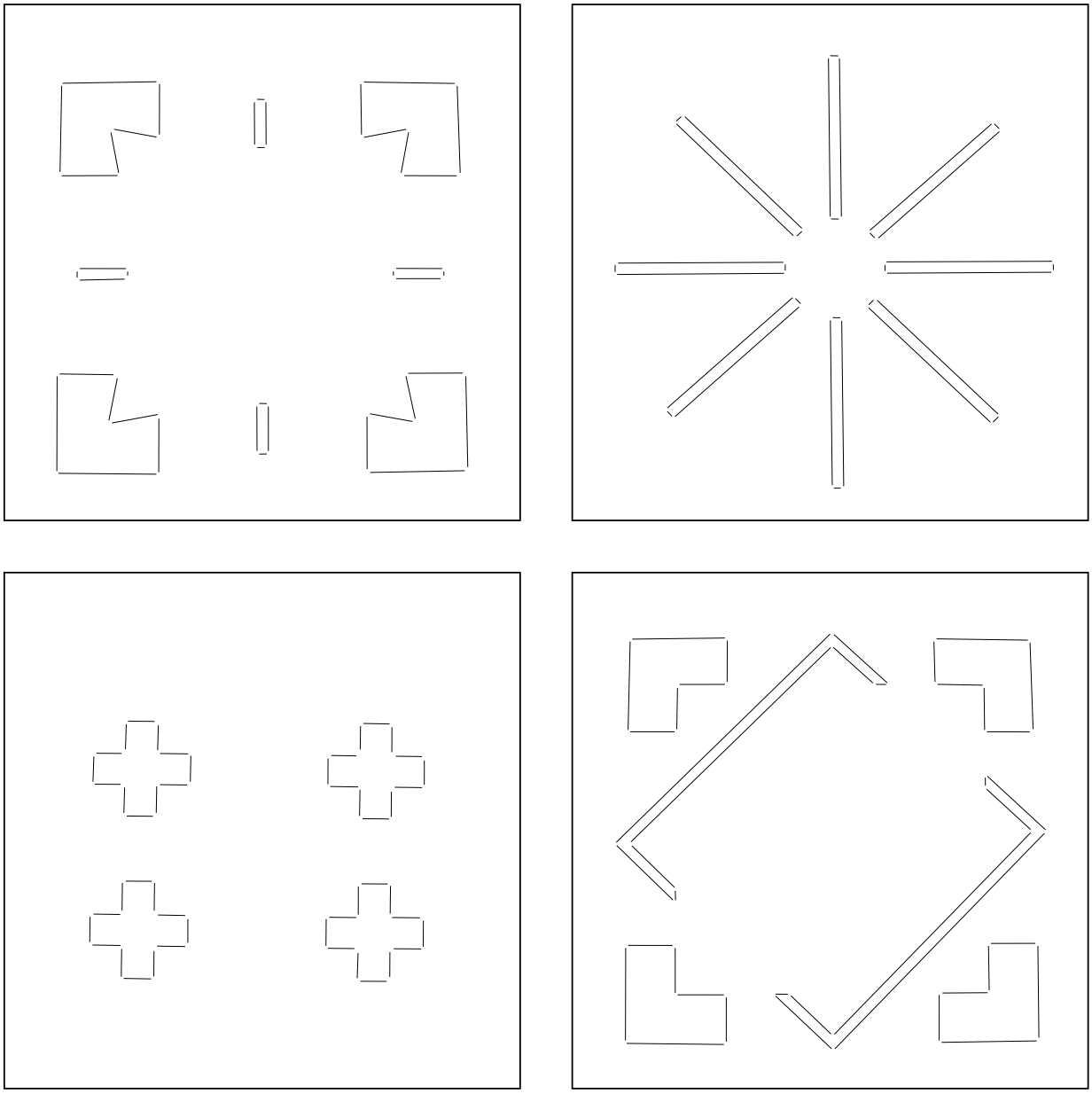

Figure 4.6: G_{input} for the four test figures. The line segments are produced by Boldt's algorithm.

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 $O(n^8)$.[5] 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 $G_{non-planar}$. Since a factor of two reduction in the number of edges in $G_{non-planar}$ 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 first of these strategies is self explanatory: only completions with a likelihood above some fixed value are explicitly represented by edges in $G_{non-planar}$. Although for any two points in the plane and most combinations of tangent orientations there exists a well defined cubic spline of minimum energy, the majority of these are too long to be plausible completions. There is little point in wasting effort explicitly representing these edges and trying to figure out if and where they intersect one another.

[5] Use of sparse matrix representations (not exploited here) would bring this back down to $O(n^4)$.

Of course the problem of determining a satisfactory threshold remains. The threshold needs to be high enough so that totally implausible completions are pruned, but low enough so that potentially viable completions are retained. The problem of choosing this value becomes much simpler if the threshold strategy is combined with the second strategy of explicitly representing only the k-most likely completions at each endpoint (where k is a small integer constant). Since the k-most likely strategy limits the number of potential completions to $O(n)$, there is a great deal of latitude in choosing the value for the minimum allowable likelihood. For all of the experimental results in this thesis, this value was 1.0×10^{-9}.

A further factor of two reduction is possible by only allowing completions which join boundary fragments with compatible signs of contrast.[6] Unfortunately, this constraint is only valid for two tone stimuli without extra markings, but when used in combination with the k-most likely strategy effectively doubles the allowable size of k. The problem of computing the k-most likely completions is then solved in the obvious way: the $n^2/2$ allowable completions are enumerated and the k-most likely (above the threshold) are stored for each endpoint. For all of the experimental results in this thesis, $k = 3$.

The final strategy, overlap pruning, is a bit more involved. The intent is to eliminate completions which are unlikely to be instantiated in the surface organization because more compelling paths incorporating larger sets of boundary fragments exist. This happens most often when a surface with a smooth boundary lies in front of a cluttered background. Where the surface and its background are of similiar orientation and reflectance, there will be little or no brightness difference. Elsewhere, fragments

[6]The use of the contrast sign constraint in the experimental system is a simple expedient to control the number of potential completions and is not consistent with human vision. Indeed, Grossberg and Mingolla[15] demonstrate that illusory contours can join luminance boundaries with opposite signs of contrast.

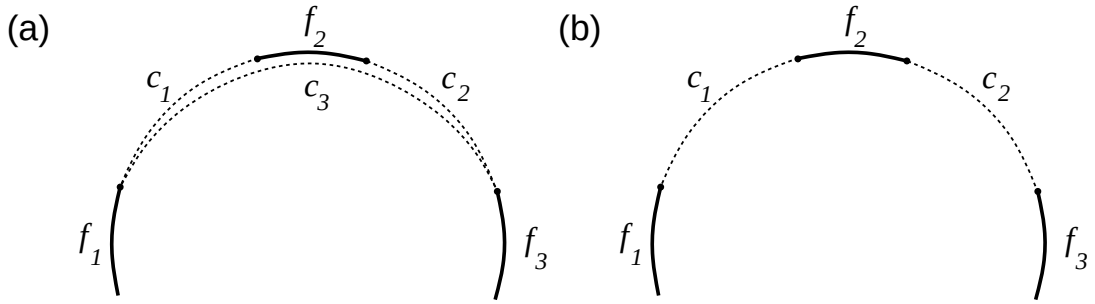

Figure 4.7: Overlap pruning. (a) Fragments f_1, f_2 and f_3 form consecutive segments of a single surface boundary. Completion c_3 is redundant, since only completions c_1 and c_2 will be instantiated in the surface organization. (b) Overlap pruning eliminates redundant completions.

aligned along the trace of the occluding surface boundary will be detectable. Since the fragments all lie along a single smooth curve, overlapping completions joining many different fragment pairs will be added to $G_{non-planar}$. Yet, only completions joining fragments consecutive on the boundary will be instantiated in the surface organization. The others are "redundant." A completion can be considered redundant when its likelihood is lower than the product of the likelihoods along an alternative path joining the same fragments.

A hypothetical example is depicted in Figure 4.7(a). Here fragments f_1, f_2 and f_3 form consecutive segments of a single surface boundary. The reflectance of the background matches the surface in the interval between f_1 and f_2 and between f_2 and f_3, producing the gaps. Because the boundary is smooth, a completion joining any two of these fragments will appear, individually, very compelling. Yet only completions c_1

and c_2 will be instantiated in the surface organization. Completion c_3 is redundant. The more fragments aligned along a boundary, the greater the number of redundant completions.

Of course, the process of identifying redundant completions (which requires search for alternative paths) can itself be quite expensive. There will clearly be a point of diminishing return, since the redundant completions which are the most expensive to identify and eliminate (i.e. those with the longest alternative paths) will also be the rarest. A reasonable strategy, likely to eliminate the bulk of the redundant completions, is to eliminate those with alternative paths less than some fixed length. In the current implementation, this length is two. The process of identifying and eliminating redundant completions is fast and efficient, since the set of length two paths is easily enumerated, and the likelihood of each of these paths can be quickly checked against the likelihood of length one paths stored in a hash table. Edges forming length one paths with lower likelihood are deleted from $G_{non-planar}$. In practice, the average time complexity of this process is linear in the number of boundary fragments. The effect of overlap pruning is demonstrated in Figures 4.8 and 4.9, which show $G_{non-planar}$ for the *Ehrenstein* test figure before and after overlap pruning.

4.3.3 Identifying Crossings

Because all contours, whether boundary fragments or completions, are represented by cubic Bezier splines, it was possible to use a standard algorithm to identify and localize points of contour intersection. The method which proved to be the fastest and most stable was *recursive subdivision.*[7] The recursive subdivision method for locating crossings exploits two properties of the Bezier representation. First, a Bezier

[7]For descriptions and comparisons of different methods for finding points where two Bezier curves intersect, see Sederberg[55].

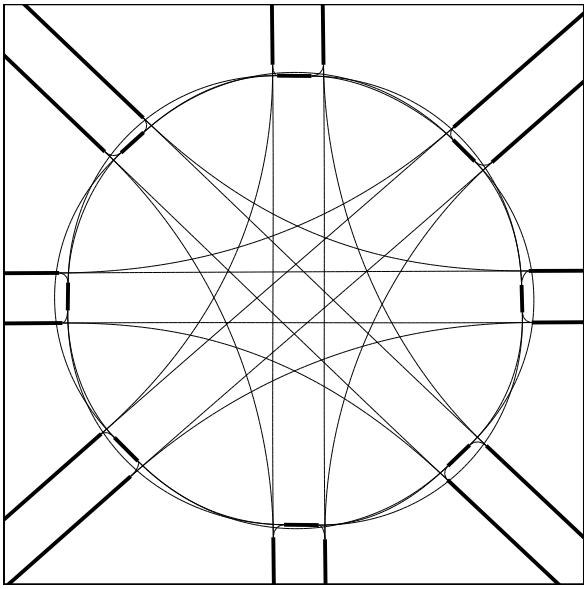

Figure 4.8: Before overlap pruning. A magnified view of a portion of $G_{non-planar}$ ($k = 3$) for the *Ehrenstein* test figure is shown prior to overlap pruning. Boundary fragments are drawn thick, potential completions (including redundant completions) are drawn thin.

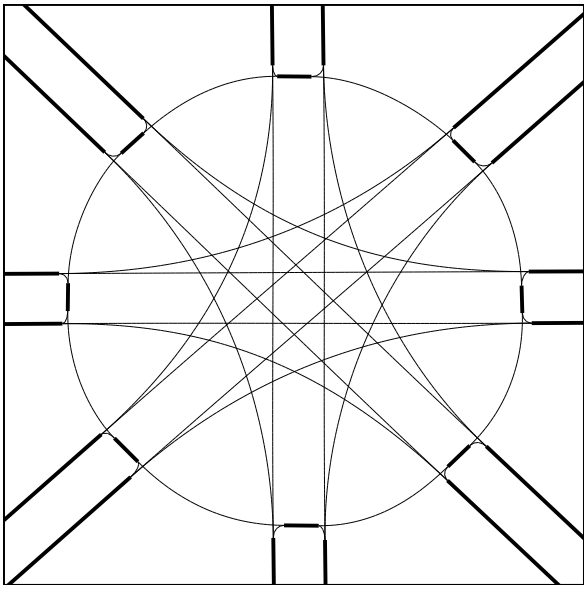

Figure 4.9: After overlap pruning. The same view of $G_{non-planar}$ ($k = 3$) after overlap pruning. Redundant completions have been eliminated.

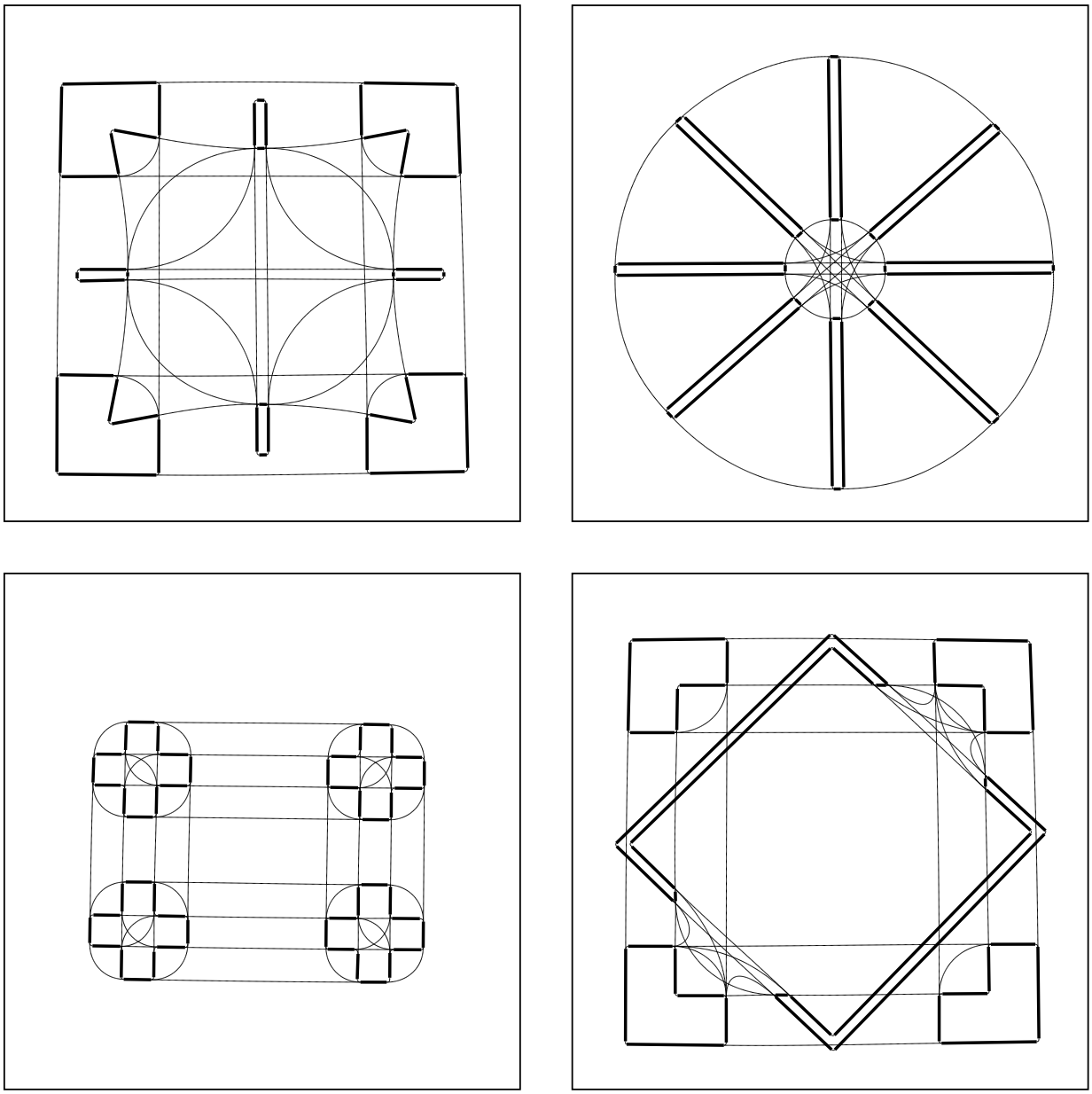

Figure 4.10: G_{planar} for the four test figures. This graph represents the k-most-likely completions ($k = 3$) for every boundary fragment as cubic Bezier splines of least energy. Boundary fragments are drawn thick and potential completions are drawn thin. Graph vertices (endpoints and crossings) are omitted for clarity's sake.

curve of a given degree, parameterized on an interval 0 to s, can always be subdivided into two Bezier curves of the same degree, one parameterized on the interval 0 to $\frac{s}{2}$, and the other on the interval $\frac{s}{2}$ to s. The second property is that unlike other spline representations, a Bezier curve lies completely within the convex-hull of its control points. Together, these two properties allow a straightforward recursive search to locate crossings. In order to build G_{planar} from $G_{non-planar}$, edges of $G_{non-planar}$ are compared pairwise to locate points of intersection. When a point of intersection is found, the two Bezier curves are subdivided, and a crossing vertex (i.e. an element of $V_{crossing}$) is "spliced in." The G_{planar} computed for the four contrast test figures are shown in Figure 4.10.

4.3.4 Solving the ILP

The integer linear program is generated by writing a fixed set of integer linear inequalities, as described in the last chapter, for each of the vertices and edges of G_{planar}. The ILP is solved by the method of *branch and bound* search[40, 51]. In branch and bound search, all integer points within the feasible region are enumerated (either explicitly or implicitly), and the optimal feasible solution is selected. This process is greatly facilitated by the fact that the LP formed by "relaxing" the requirement for an integer solution provides an upperbound on the solution of the "harder" ILP. A standard algorithm for solving LP's, such as the simplex algorithm, can be used to compute these upperbounds. In the course of enumerating the feasible solutions, the upperbounds computed by simplex can be compared to the current lowerbound on the solution to the ILP. Where the current lowerbound on the solution to the ILP exceeds the upperbound computed by the simplex algorithm, no improvement is possible, and the "search-tree" can be pruned. The search continues until either all solutions have been enumerated, in which case the optimal feasible solution is found, or the problem

is shown to be infeasible. 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 Figure 4.11. These organizations maximize the following objective function:

$$Q'_{surface} = \alpha\, Q'_{unit} + \beta\, Q_{depth} \tag{4.10}$$

Here, Q'_{unit} is as defined in Chapter 3 while Q_{depth} implements a preference for black figure against a white background. If we adopt the convention that x_i represents the sign of occlusion with orientation matching the sign of contrast (and let x'_i represent the opposite sign), then this can be accomplished by assigning a positive unit weight to x_i for $i \in E_{fragments}$:

$$Q_{depth} = \sum_{i \in E_{fragments}} x_i \tag{4.11}$$

As was noted in Chapter 3, without the Q_{depth} term, the ILP is underconstrained. In all experiments, $\alpha = \beta = 1$.

At least superficially, these organizations 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.

However, the printed output is misleading in at least one respect: although from Figure 4.11, it is clear that four "sticks" are completed underneath an illusory disc, what is not obvious is that the experimental system also commits to a particular

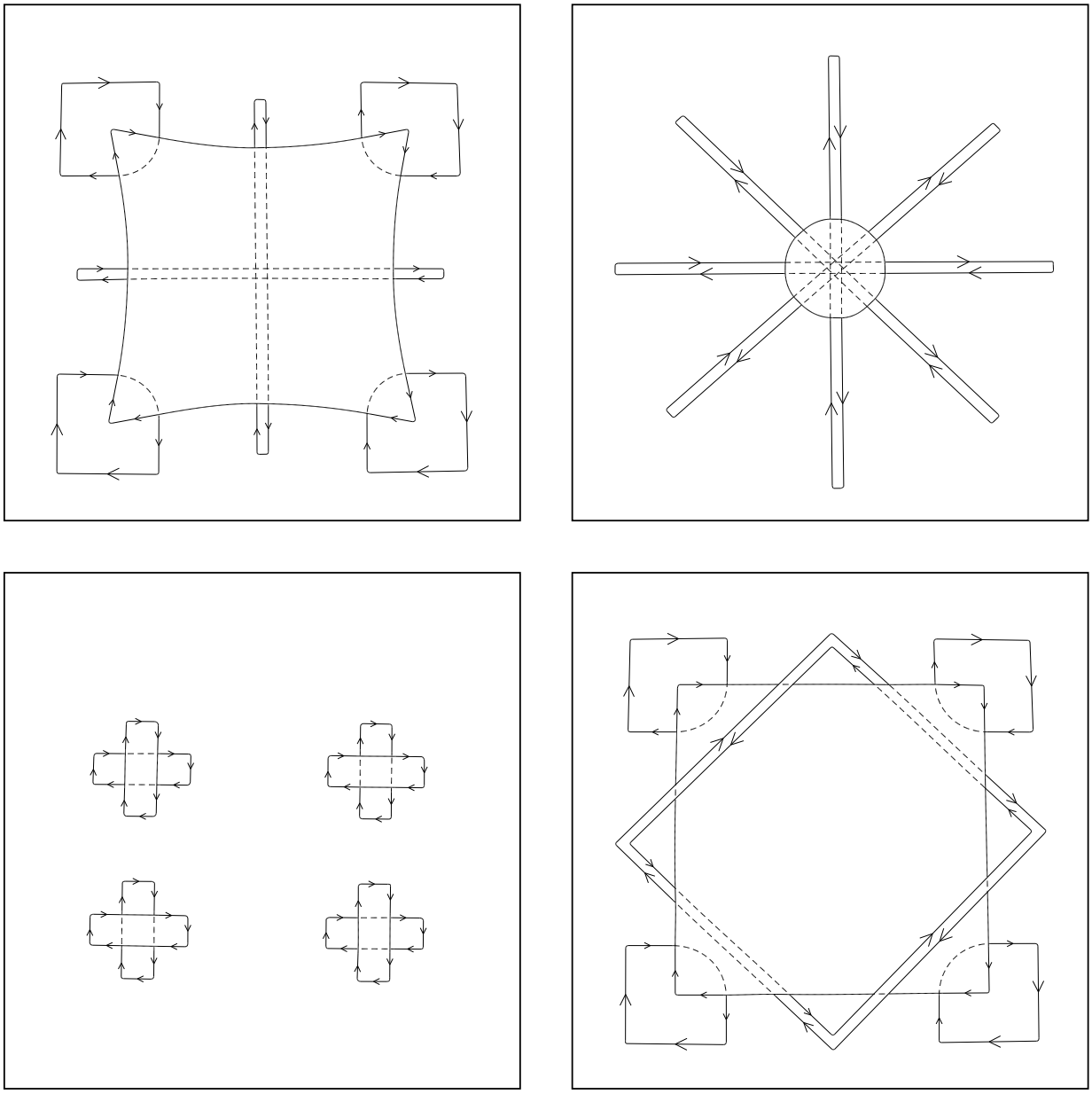

Figure 4.11: G_{knot} for the four test figures.

ordering in depth. That is, G_{knot} represents precisely one of the twenty-four (i.e. 4!) possible depth orderings. It seems unlikely that the human visual system, in this situation, commits to a unique depth ordering. Indeed, perceptual completion of the "sticks" is not even a precondition for formation of the illusory disc, since Ehrenstein figures with odd numbers of inducing elements also elicit illusory discs. This suggests that the computational goal may ultimately need to be modified.

4.4 Outline Stimuli

Kanizsa's line drawings of partially occluded cubes demonstrate that figural completion phenomena are not peculiar to perception of contrast stimuli but are also characteristic of perception of outline stimuli (see Figure 4.12). The contours which form a line drawing differ from the luminance boundaries of a natural image in that the former possess no sign of contrast. In the experiments conducted thus far, the sign of contrast was used in two different ways. First, it was used to limit the number of potential completions: only completions joining boundary fragments with compatible signs of contrast are explicitly represented as edges in $G_{non-planar}$. Second, by weighting the variables representing the sign of occlusion by an appropriate function of the sign of contrast, it was possible to implement the human preference for perceiving black figure against white ground.

Unfortunately, in line drawings, signs of contrast do not exist and therefore can not be used for either purpose. As far as limiting the number of potential completions is concerned, we note that although the size of $G_{non-planar}$ potentially doubles, this additional complexity is not insurmountable. However, because the sign of contrast plays a critical role in figure-ground preference, it was necessary to find a substitute for this purpose.

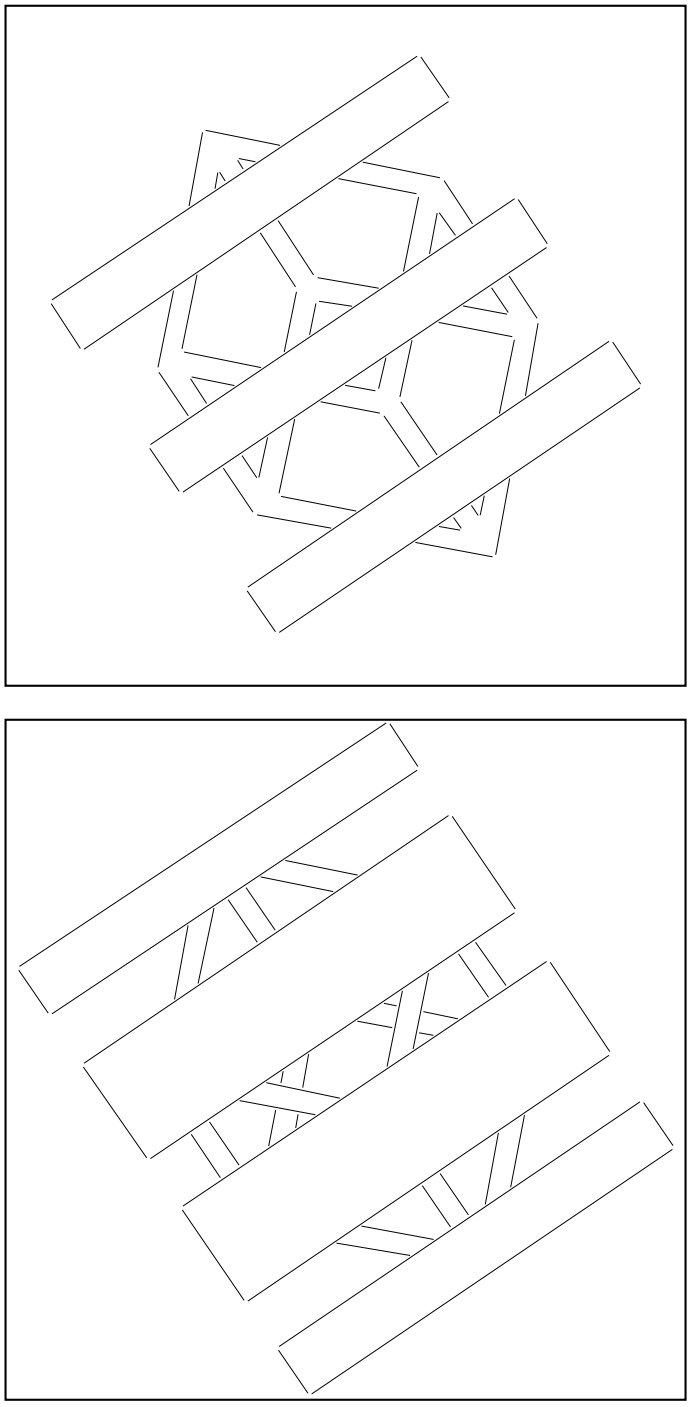

Figure 4.12: Two of Kanizsa's partially occluded cubes. For convenience sake, the upper and lower cubes will be referred to as the *Y's with Bars* and *X's with Bars* test figures.

In the last chapter, we noted that in human vision there is a tendency to perceive the space between closely spaced parallel lines as figure. Pairs of lines related in this way are called *bars*. In the experimental system, bars were identified by computing the following measure for all pairs of boundary fragments and discarding those pairs beneath a set value (i.e. 1.0×10^{-9}):

$$Pr(bar \mid \mathbf{x}) = \exp(-\mathbf{x}^t C_{bar}^{-1} \mathbf{x}) \tag{4.12}$$

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

$$x_1 = |\theta - \theta'| \tag{4.13}$$

$$x_2 = width \tag{4.14}$$

$$x_3 = \frac{width}{overlap} \tag{4.15}$$

and C_{bar} is:

$$C_{bar} = \begin{bmatrix} 0.001 & 0 & 0 \\ 0 & 50.0 & 0 \\ 0 & 0 & 0.5 \end{bmatrix} \tag{4.16}$$

Here x_1 is orientation difference, x_2 is *width* (i.e. the distance between an arbitrarily chosen endpoint of one boundary fragment and the second boundary fragment along a direction perpendicular to the first boundary fragment), and x_3 is "aspect ratio" (i.e. the ratio of *width* and a second quantity called *overlap*, which is the length of the region of overlap in the perpendicular projection of one boundary fragment onto the second). The meaning of these quantities is illustrated in Figure 4.13.

Every boundary fragment bordering a bar can be oriented so that the interior of the bar lies to the right as the fragment is traversed in the direction of its orientation. Fragments which do not border a bar remain unoriented. The orientation

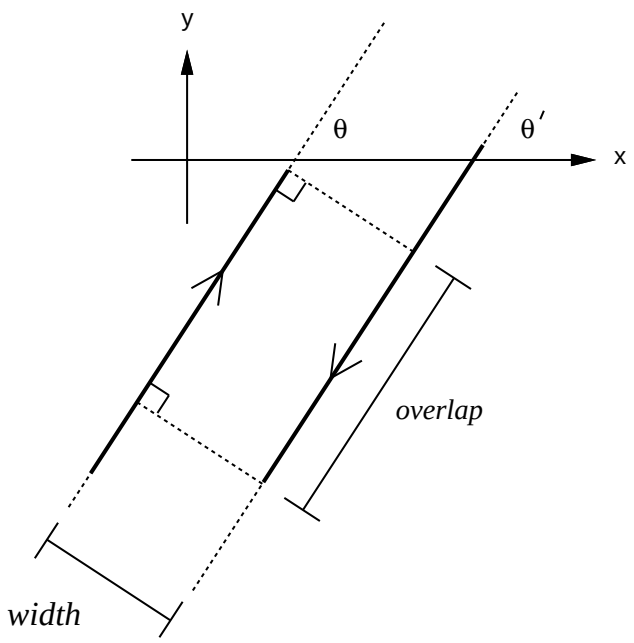

Figure 4.13: The meaning of *orientation*, *overlap* and *width.*

of a fragment which forms a bar to its left and right is ambiguous. Where defined and unambiguous, these orientations can be substituted for sign of contrast and can be used both to limit the number of potential completions and as a basis for figure-ground preference. The bars identified for the cube line drawings are shown in Figure 4.14. The G_{planar} computed for the cube line drawings, and constrained by the bar orientations wherever possible, are shown in Figure 4.15.

The optimal G_{knot} for the Kaniza cube line drawings are shown in Figure 4.16. Unlike G_{knot} for the contrast test figures, these organizations optimize $Q_{surface}$, which assigns non-zero weight to corners. It is tempting to describe these results by simply observing that in each case, the experimental system has successfully "completed the occluded cube." It is particularly worth noting that the experimental system has completed the difficult *X's with Bars* test figure in a consistent manner. This demonstrates the disambiguating power of topological constraints.

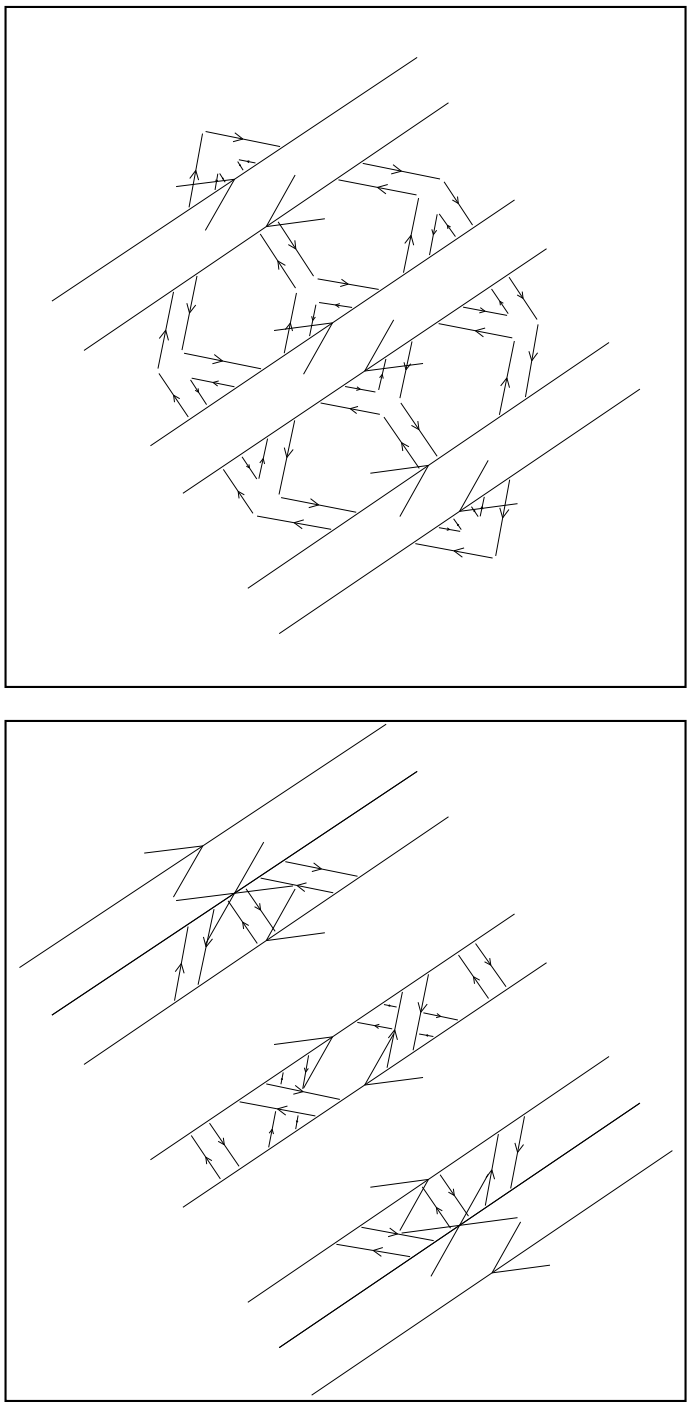

Figure 4.14: Orientations consistent with bars in cube line drawings. Every boundary fragment bordering a bar can be oriented so that the interior of the bar lies to the right as the fragment is traversed in the direction of its orientation. Fragments which do not border a bar remain unoriented. The orientation of a fragment which forms a bar to its left and right is ambiguous. Where defined and unambiguous, these orientations can be substituted for sign of contrast and can be used both to limit the number of potential completions and as a basis for figure-ground preference.

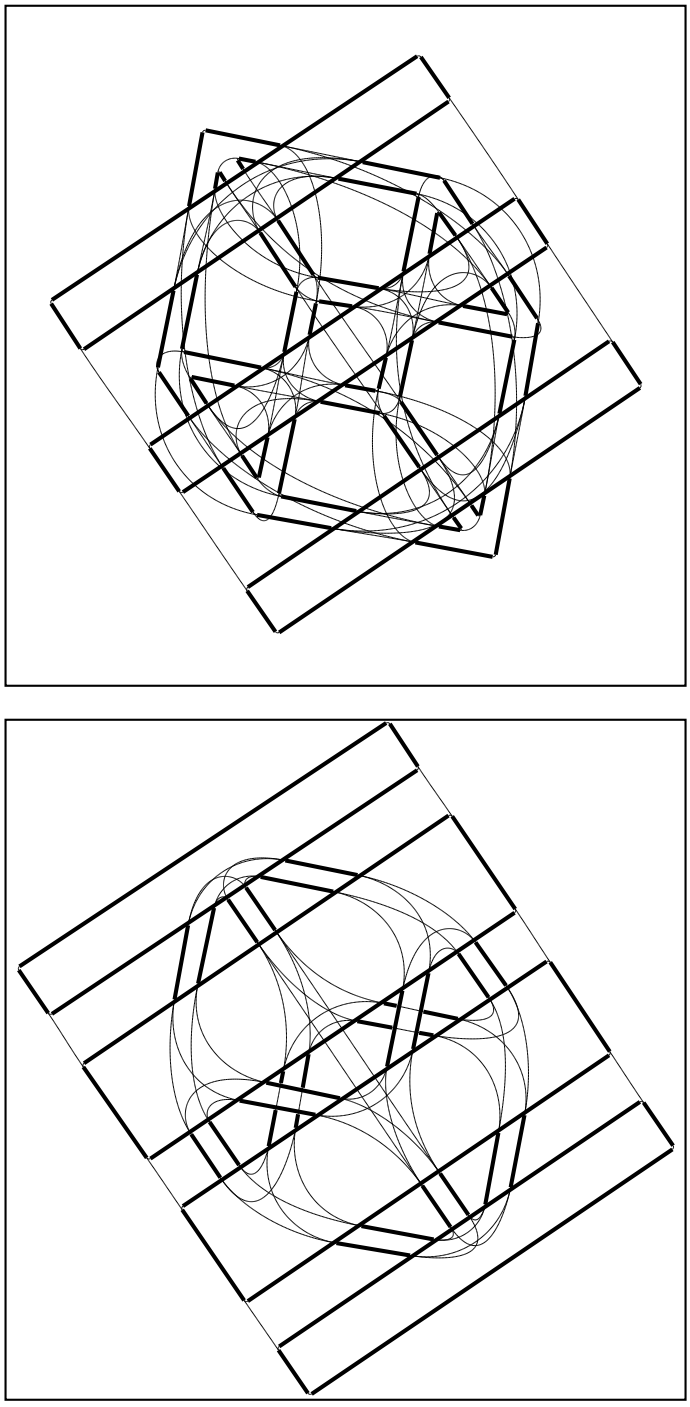

Figure 4.15: G_{planar} for Kanizsa's partially occluded cubes. Again, this graph represents the k-most-likely ($k = 3$) completions for every boundary fragment as cubic Bezier splines of least energy .

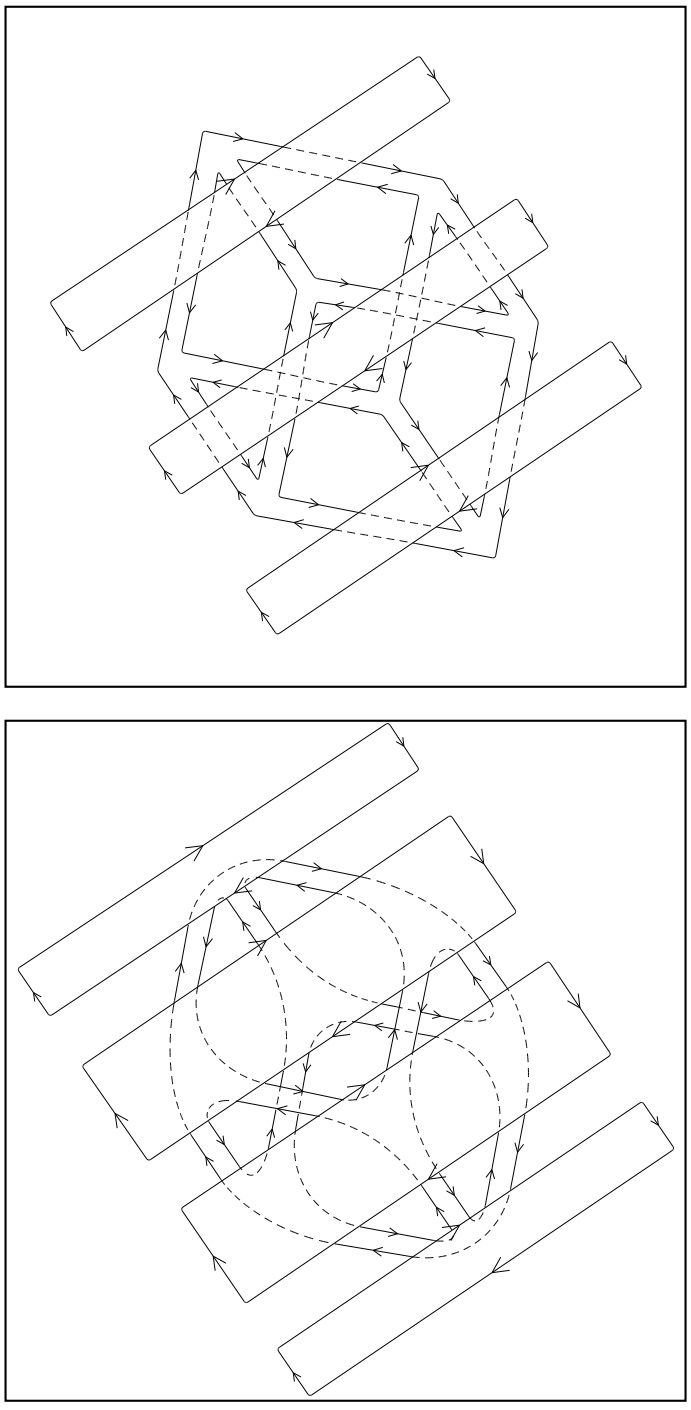

Figure 4.16: G_{knot} for Kanizsa's partially occluded cubes. It is particularly worth noting that the experimental system has completed the difficult *X's with Bars* test figure in a consistent manner. This demonstrates the disambiguating power of topological constraints.

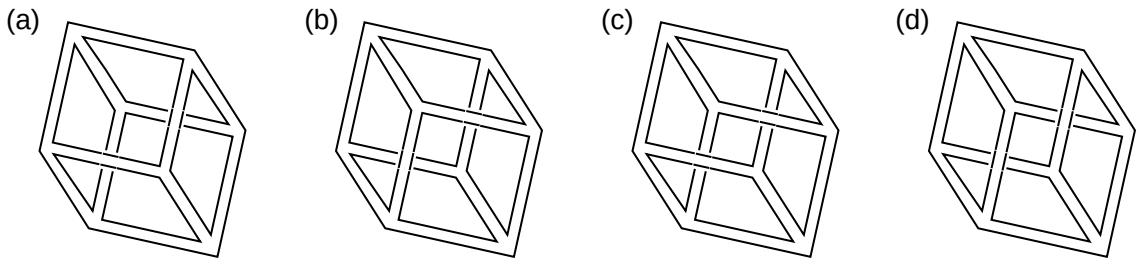

Figure 4.17: Four equally optimal organizations. As with the *Ehrenstein* test figure, the printed output for the *Y's with Bars* test figure is somewhat misleading because it does not indicate that the experimental system arbitrarily commits to the organization shown in (c). Although all four are topologically consistent orientable surfaces with boundary, (c) and (d) lead to inconsistent three-dimensional interpretations as cubes.

However, as with the *Ehrenstein* test figure, the printed output for the *Y's with Bars* test figure is somewhat misleading because it does not indicate that the experimental system arbitrarily commits to one of four equally optimal organizations (see Figure 4.17(c)). Furthermore, two of the four organizations, while topologically sound, are inconsistent in their three-dimensional embedding in space. Although the two consistent organizations could serve as the precursors of a Necker cube stimulus (see Figure 4.18), strictly speaking, what has been perceptually completed is not a cube, but an orientable surface with boundary.

4.5 Theory and Experiment

This section is included to illustrate the role the experimental system played in development of the computational theory which forms the major part of this thesis. The labeling scheme originated in a series of observations about edges in scenes formed

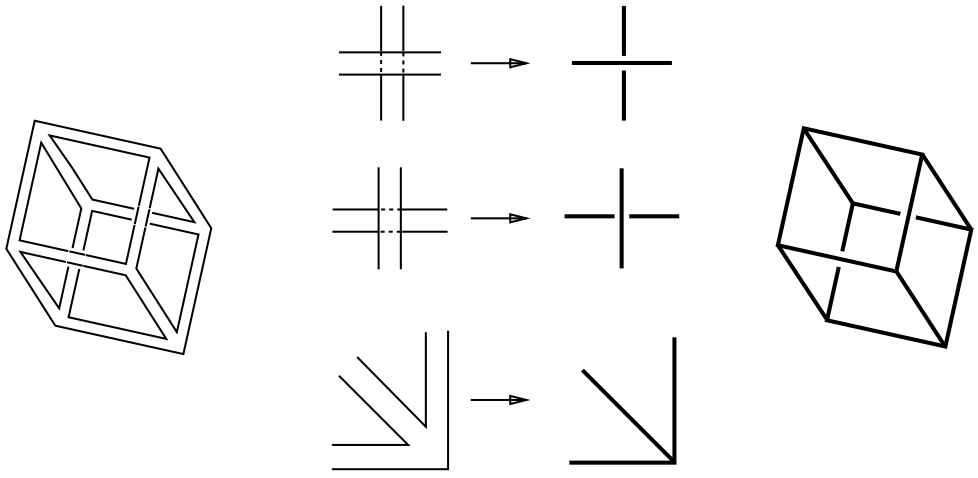

Figure 4.18: Labeled knot diagram as the precursor of a Necker cube stimulus. Although strictly speaking, the only thing the experimental system has perceptually completed is an orientable surface with boundary, a straightforward abstraction process could produce a Necker cube stimulus.

from flat vinyl cutout shapes called "Colorforms." These observations were incorporated into a set of necessary constraints on the appearance of edges in these simple scenes.[8] These constraints formed the basis of the integer linear program described in [59].

The requirement that the depth of the occluding surface be less than or equal to the depth of the occluded surface was among the constraints that were readily apparent. However, because this particular constraint is very ugly when expressed in terms of integer linear inequalities, it was omitted from the integer linear program. A certain amount of wishful thinking prevailed—it was hoped that this constraint somehow derived from the others. A not very energetic search failed to discover a "counterexample."

This changed with the discovery of a "counterexample" while experimenting with different objective functions. The anomolous G_{knot} is displayed in Figure 4.19 (this

[8]It is interesting to speculate that had stiff cardboard been used instead of flexible vinyl, a more restrictive set of necessary constraints might have resulted.

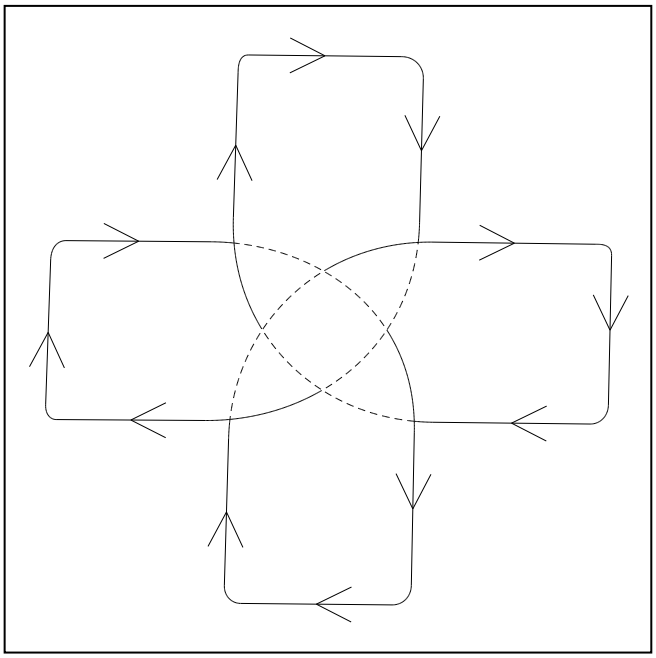

Figure 4.19: "Knotted" experimental output. This knot diagram was accidently produced in the course of experimenting with different objective functions. Because it can not be reduced to the unknot by a sequence of Reidemeister moves (see Kauffman[31]), it is knotted in the true mathematical sense.

was the source of Figure 2.10 in Chapter 2). This example shows that the contrast sign constraint alone does not ensure topological validity. It was later discovered that this figure is in fact "knotted" in the true mathematical sense (i.e. it can not be reduced to the unknot by a sequence of Reidemeister moves). It happens to be Knot 8_{18} in the table of knots included in Kauffman's[31] book "On Knots." This was a pretty momentous event. While integer linear inequalities enforcing the "depth order requirement" could be added to the integer linear program to fix the immediate problem (see Chapter 3), deeper questions emerged: how could one be certain that a second "counterexample" of a different kind would never appear? What, apart from the fact that it contradicts the depth order requirement, makes Figure 4.19 "bad?"

That is, what *independent* criterion could be used to judge the "sufficiency" of a given set of constraints?

This lead to a clear definition of a domain which until this point had only been defined as "scenes which can be constructed from Colorforms." The sufficiency criterion proved to be the mathematical definition of surface with boundary. These are objects which possess neighborhoods that are everywhere homeomorphic to either discs or half-discs.

CHAPTER 5

A REVISED PROBLEM-LEVEL FORMULATION

Figural completion was portrayed in Chapter 3 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 sub-problems, the first devoted to shape, the second devoted jointly to unit and depth. Decomposed in this way, it proved possible to formulate the second sub-problem as an integer linear program, which led to a unique solution.

In this chapter, a further decomposition is proposed. Again, this decomposition occurs along the lines of the inherent ambiguities (see Figure 5.1). 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 two-fold. First, evidence from human vision supports the hypothesis that unit organization occurs in advance and

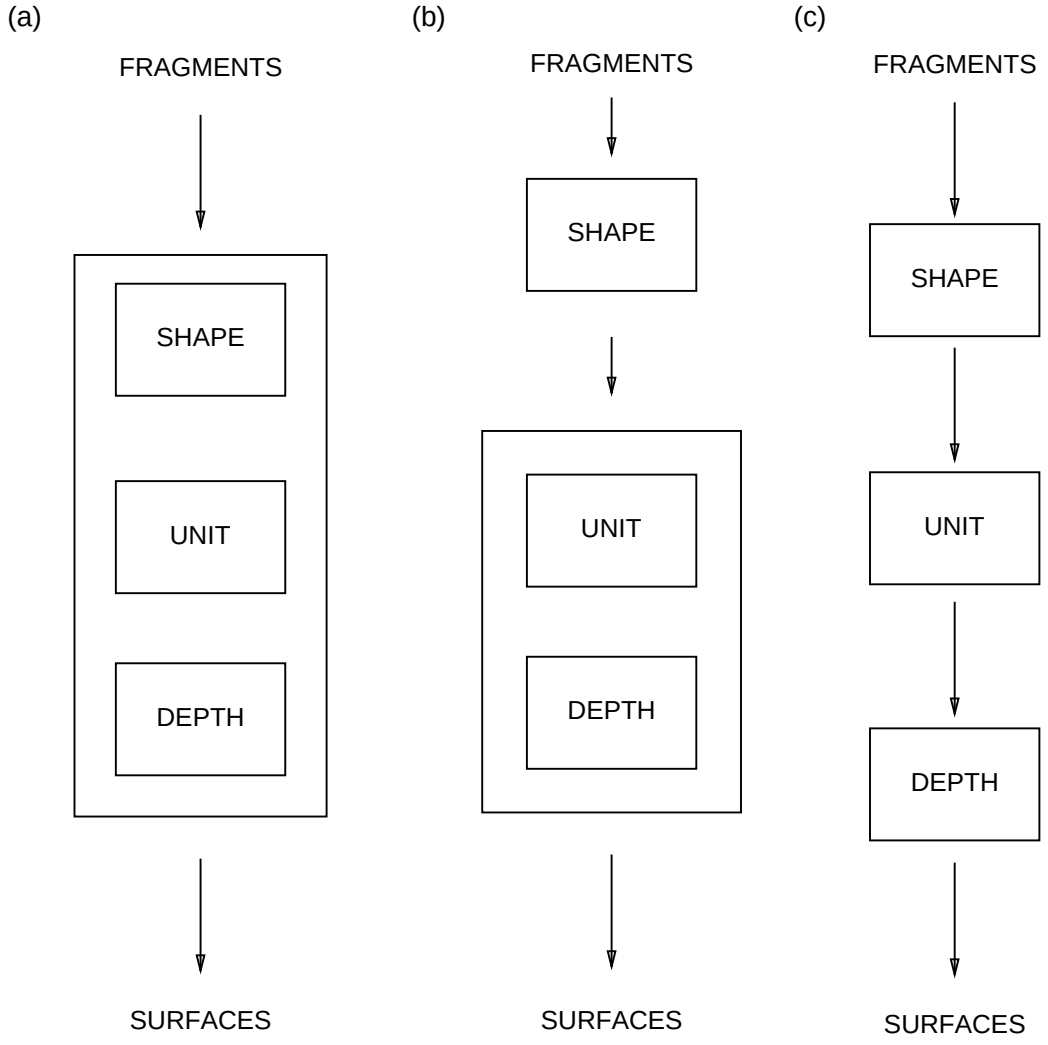

Figure 5.1: Alternate problem decompositions. Figural completion requires three different sources of ambiguity to be overcome (i.e. shape, unit and depth). In theory, all three can be resolved concurrently, which would result in the problem decomposition shown in (a). However, in Chapter 3, we proposed 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 chapter, 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.

independently of depth organization (see Figures 5.2 and 5.3). This lends support to the conjecture of Kellman and Loukides[32] mentioned in Chapter 3. Second, the integer linear program for the unit organization sub-problem is shown to possess a property called *total unimodularity*, which allows it to be solved by numerical relaxation in a locally connected network. We argue that this form of computation is consistent with the observation that figural completion is sensitive to non-local changes in image context.

5.1 Evidence from Human Vision

In Chapter 3, 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[32], who argue that unit organization is accomplished in advance and independently of depth organization. More specifically, they propose that visible contour 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's 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 Figure 5.2(a) which is a variation of a figure designed by Fahle and Palm[12]. Although only one illusory rectangle is physically

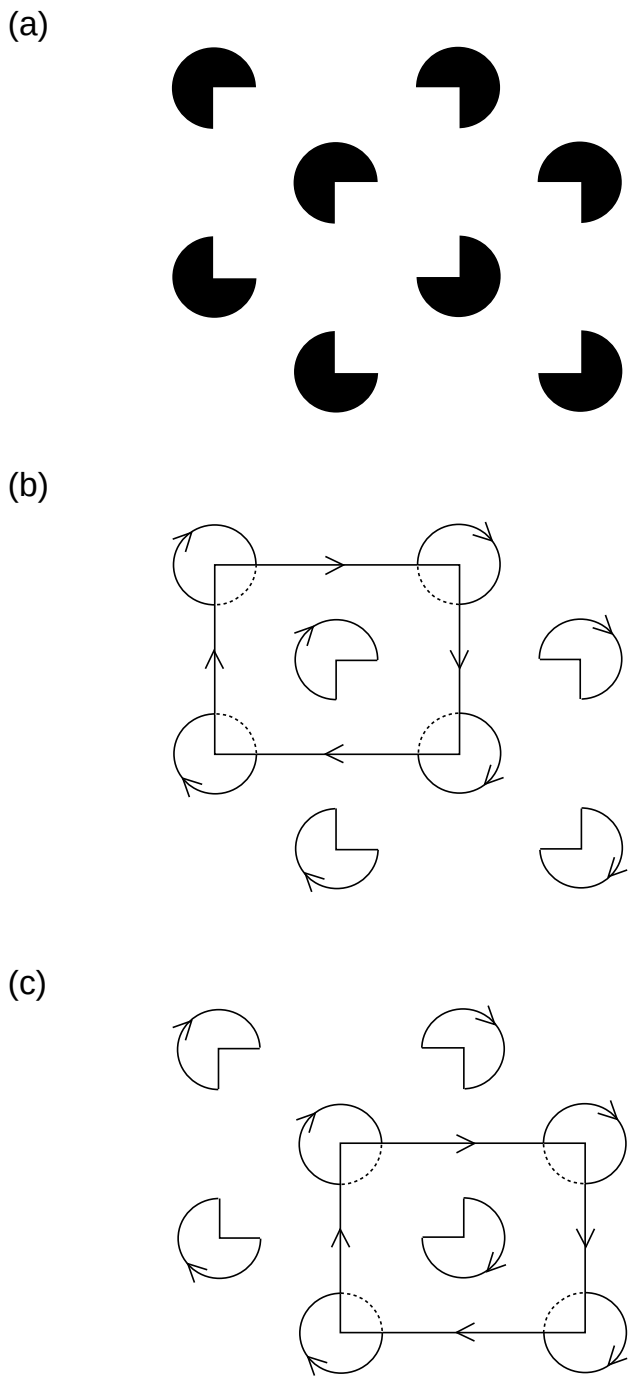

Figure 5.2: Prediction of surface organization model. (a) A variation of a figure designed by Fahle and Palm[12]. Although only one illusory rectangle is physically possible, the arrangement of the inducing elements and the overall symmetry of the figure seems to suggest that two are present. (b) and (c) Bistable interpretation predicted by surface organization model.

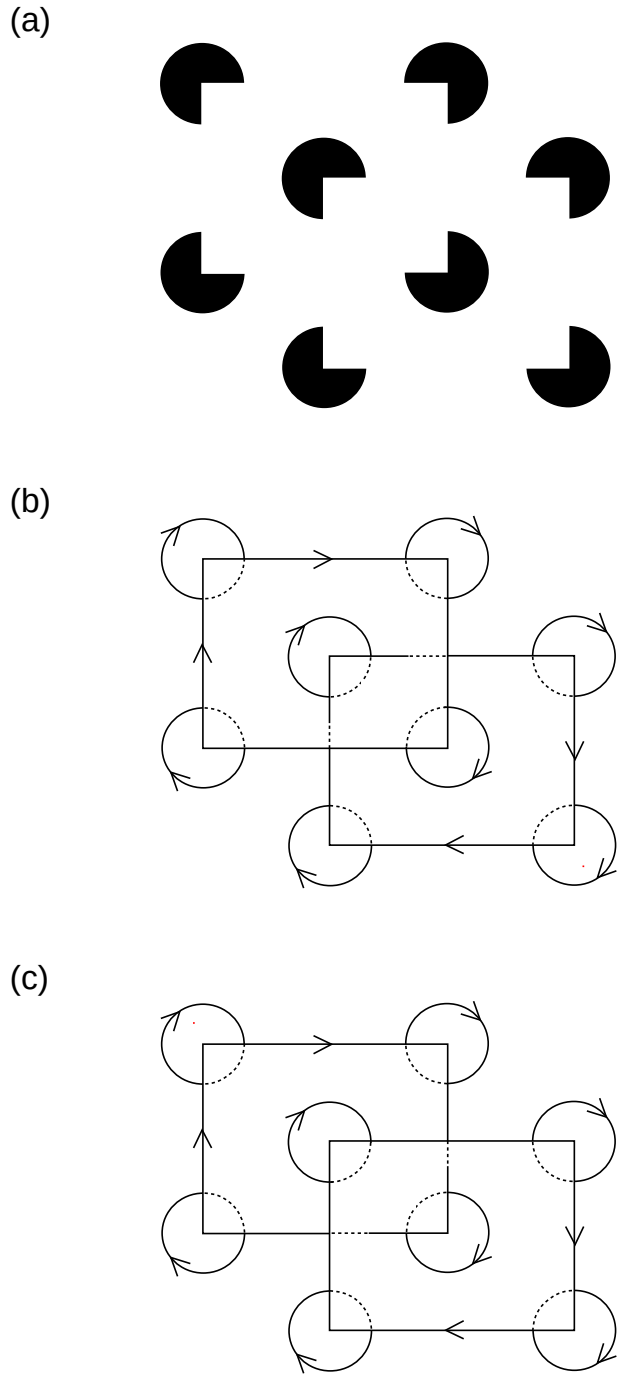

Figure 5.3: Prediction of unit/depth organization model. (a) Same stimulus as the previous figure. (b) and (c) Bistable interpretation predicted by unit/depth organization model.

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 Figure 5.2(b) and (c) will be perceived with equal likelihood. Each of these is a distinct surface 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 Figure 5.2(b) and (c) is perceived. In fact, most subjects experience something more closely resembling Figure 5.3(b) and (c). 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.

5.2 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, $G_{non-planar}$. The overall structure of the unit/depth organization model is compared to the surface organization model in Figure 5.4. If x_j equals one when completion $j \in completions(p)$ is instantiated, and x_j equals zero when completion j is not

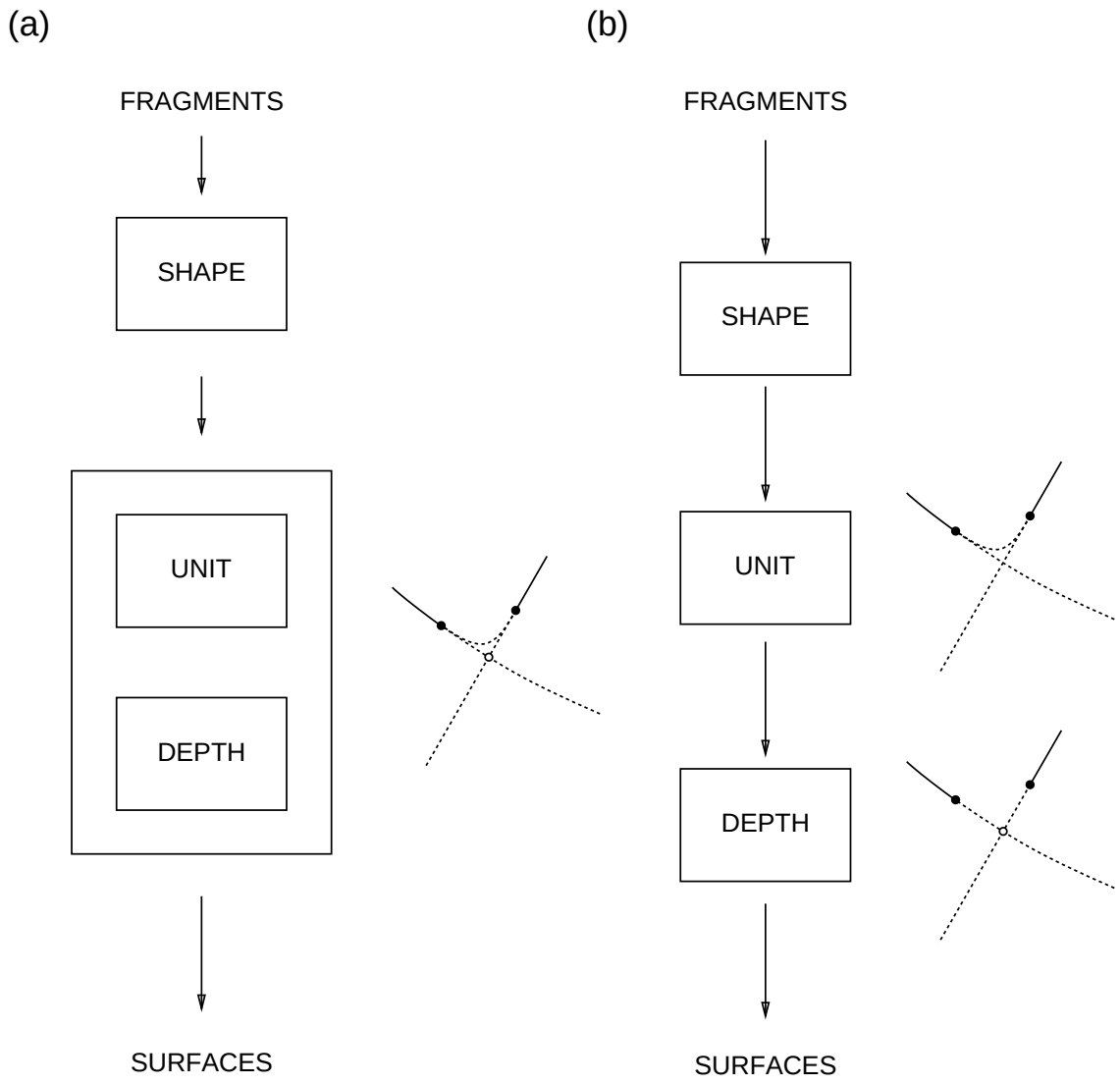

Figure 5.4: A comparison of the labeling problems. (a) The integer linear program for the surface organization model (i.e. $ILP_{surface}$) is a labeling problem on G_{planar}. Unit and depth organization constrain one another, resulting in true surface organization. (b) In the unit/depth organization 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, $G_{non-planar}$. Completions not instantiated in the unit organization are deleted. A much simpler G_{planar} is then labeled by ILP_{depth}.

instantiated, then ILP_{unit} is formed by generating one constraint per endpoint p of the form:

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

The objective function for ILP_{unit} is Q_{unit} or Q'_{unit} depending on whether contrast or outline stimuli are involved. The optimal feasible solution of ILP_{unit} 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 $G_{non-planar}$. The structure of $G_{non-planar}$ after pruning is very simple. Since the in-degree and out-degree of every endpoint vertex is equal to one, the connected components of $G_{non-planar}$ 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 Figure 5.5. 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 objective function, as defined, does not incorporate all of the preference factors which operate in human vision.

Figure 5.6 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, 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 $G_{non-planar}$ 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}.

The principal simplification is that the sign of occlusion of all edges in the same boundary component can be represented by a single 0-1 valued variable. The constraint that boundary component, b, must have a unique sign of occlusion is then:

$$x_b + x'_b = 1 \tag{5.1}$$

The set of constraints which must be enforced at crossing vertices is also somewhat simplified. Recall that in the surface organization model, the four states associated with crossing c were represented by 0-1 valued variables $x_c^{\top}, x_c^{\perp}, x_c^{\vdash}$ and $x_c^{\dashv}$. If either of the completions intersecting at crossing c is not instantiated then the value of these four variables is zero. However, in the unit/depth organization model (since every completion in G_{planar} must appear in the surface organization) the constraint requiring crossing c to be in one of the four crossing states when two intersecting completions (i.e. u and l) are instantiated is extraneous, since the left side of the inequality always equals one:

$$x_u + x'_u + x_l + x'_l \leq x_c^{\top} + x_c^{\perp} + x_c^{\vdash} + x_c^{\dashv} + 1 \tag{5.2}$$

Furthermore, the inequality constraint in $ILP_{surface}$ which ensures that the four states are mutually exclusive is replaced by a stronger, equality constraint in ILP_{depth}:

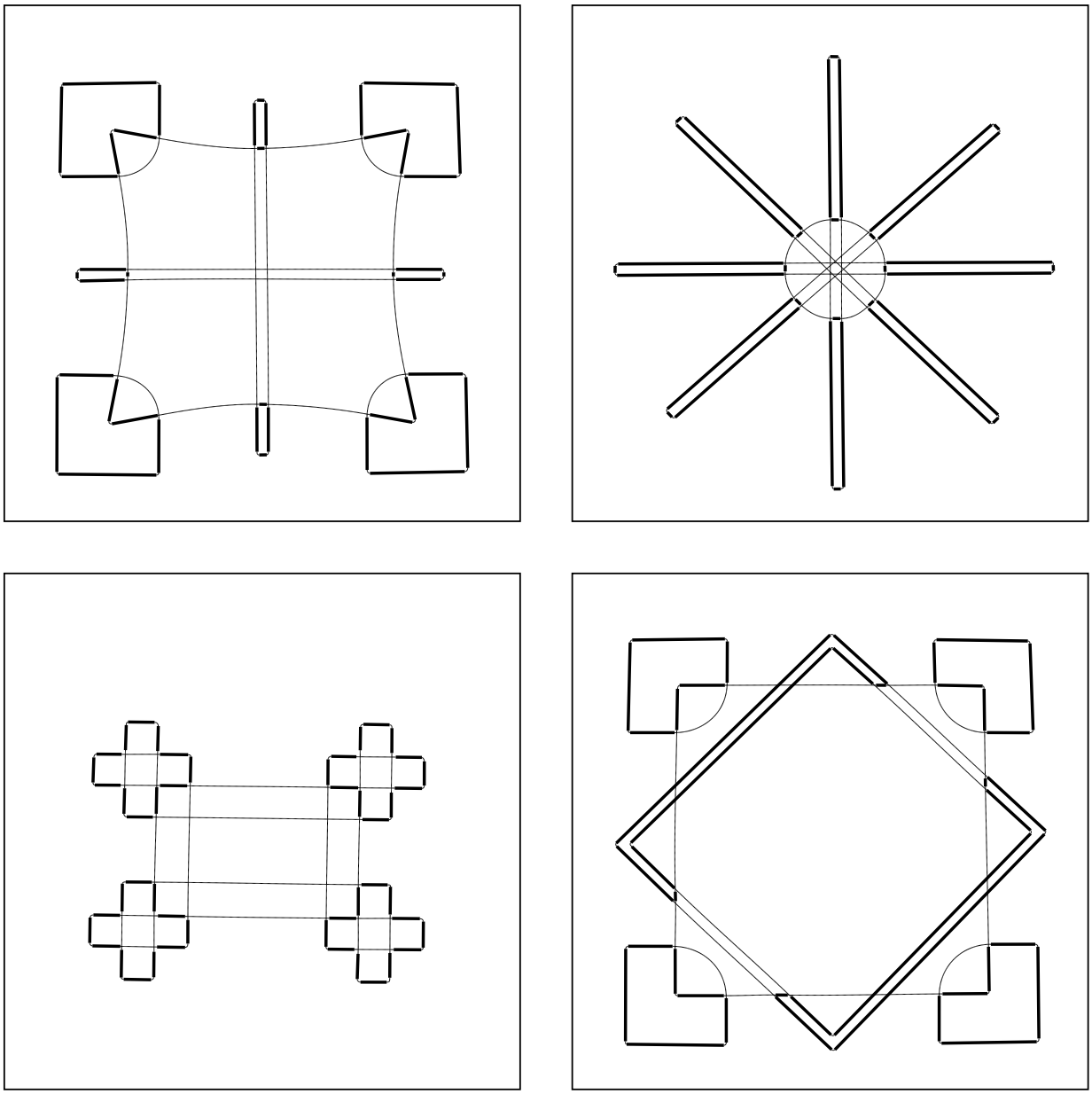

Figure 5.5: Unit organizations for the contrast test figures. In three of the four cases, the unit organization can be consistently labeled; these act as the precursors of topologically valid anterior scenes. However, in the case of the *Kanizsa Plusses* test figure (lower left), no consistent labeling is possible.

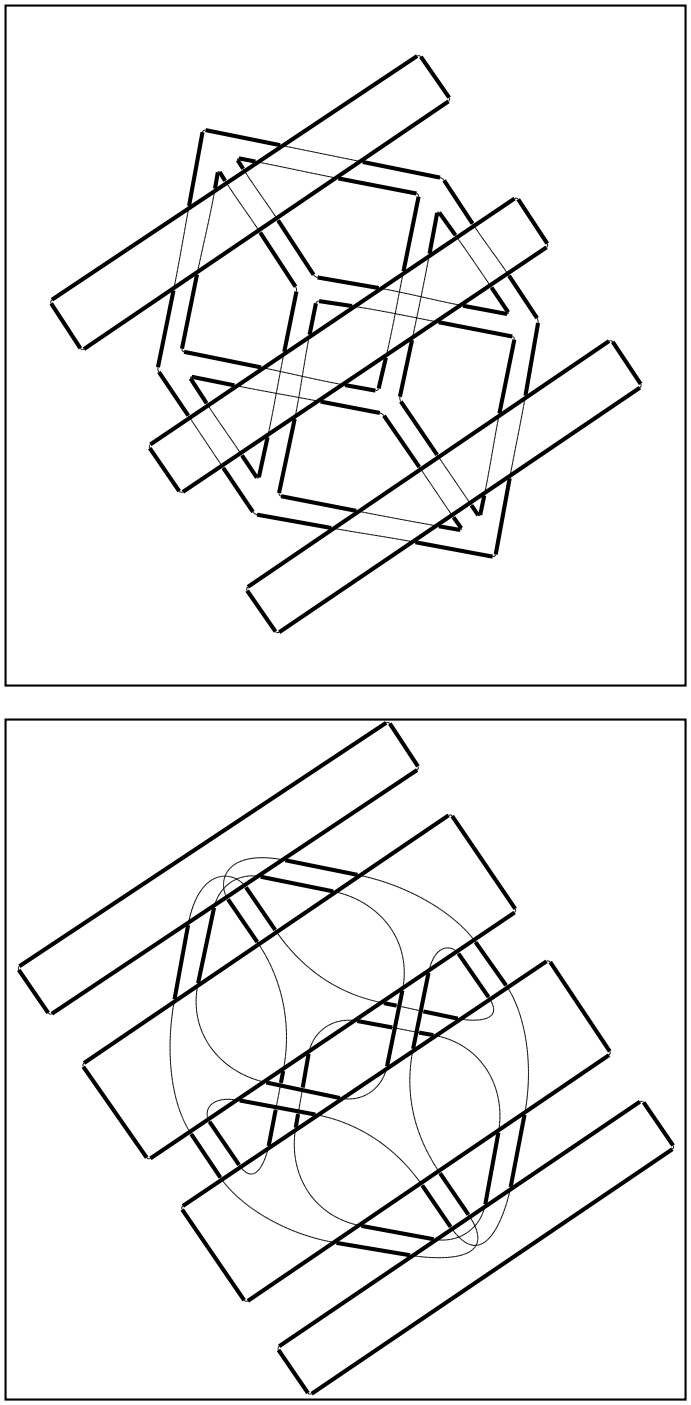

Figure 5.6: Unit organizations for the outline test figures. In the case of the *Y's with Bars* test figure (top), the unit organization is the same as for the surface organization model. However, in the case of the *X's with Bars* test figure (bottom), the unit organization admits no consistent depth labeling.

$$x_c^{\top} + x_c^{\perp} + x_c^{\vdash} + x_c^{\dashv} \;=\; 1 \tag{5.3}$$

The other constraints enforced at crossings are the same as for the surface organization model. These include the constraints which ensure that the contour bounding the uppermost surface possesses the appropriate sign of occlusion, the constraints which relate boundary depth to the four crossing states, and the constraints which enforce the depth order requirement.

Having described the integer linear constraints which define ILP_{depth}, all that remains is to define its objective function. 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 shown in Chapter 3, the relative likelihoods of different figure-ground assignments can not be compared when the unit organizations are of different size. 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.[1]

5.3 Cost Comparison of Two Models

To provide some indication of the relative computational complexity of the surface and unit/depth organization models, it will be useful to compare the problem size and execution statistics of their experimental implementations. The important

[1]For the experimental results, $Q_{surface}$ was actually used. However, because the contribution of Q_{unit} to $Q_{surface}$ remains constant, this was equivalent to using Q_{depth}.

Table 5.1: Complexity statistics for $ILP_{surface}$.

Figure	Obj. Fun.	Variables	Constraints	Pivots	Multiplies	B.B.
Warped Square	41	676	966	1124	7.34×10^8	1
Ehrenstein	51.2	792	1073	1209	1.03×10^9	1
Kanizsa Plusses	97	875	1204	1411	1.49×10^9	1
Woven Square	53.4	782	1222	1263	1.21×10^9	1
Y's with Bars	249.8	2510	4181	13377	1.40×10^{11}	1
X's with Bars	119.9	738	1611	5339	6.35×10^9	9

problem size statistics are the number of variables and constraints comprising each integer linear program. Execution statistics include the total number of nodes in the search tree for the branch and bound enumeration (each node represents a separate application of the simplex algorithm), and the total number of simplex pivot steps. The statistics for $ILP_{surface}$, ILP_{unit} and ILP_{depth} were collected for the four contrast and two outline test figures. Table 5.1 lists these statistics for $ILP_{surface}$ while Tables 5.2 and 5.3 list them for ILP_{unit} and ILP_{depth}. These tables also include the value of the objective function for the optimal feasible solution and an estimate of the total number of floating point multiplies required to solve the problem. Assuming that the complexity of a simplex pivot step is proportional to the size of the constraint matrix (i.e. the product of the number of variables and constraints), then the product of the size of the constraint matrix and the total number of pivot steps provides a good estimate of the total number of multiplies. This, in turn, provides an indication of the time-complexity on a sequential computer.

The computational advantage of decomposing the figural completion problem according to the unit/depth organization model are convincingly demonstrated in Tables 5.4 and 5.5. Table 5.4 compares the total number of simplex pivot steps for $ILP_{surface}$ to the sum of the totals for ILP_{unit} and ILP_{depth}. Table 5.5 similiarly compares the total number of floating point multiplies. The average reduction in the total number

Table 5.2: Complexity statistics for ILP_{unit}.

Figure	Obj. Fun.	Variables	Constraints	Pivots	Multiplies	B.B.
Warped Square	17	112	161	44	7.93×10^5	1
Ehrenstein	35.2	96	129	46	5.70×10^5	1
Kanizsa Plusses	61.2	159	193	39	1.20×10^6	1
Woven Square	25.4	114	161	36	6.61×10^5	1
Y's with Bars	195.8	198	181	167	5.98×10^6	1
X's with Bars	73.3	119	133	68	1.08×10^6	1

Table 5.3: Complexity statistics for ILP_{depth}.

Figure	Obj. Fun.	Variables	Constraints	Pivots	Multiplies	B.B.
Warped Square	41	264	407	118	1.27×10^7	1
Ehrenstein	51.2	352	521	459	8.42×10^7	1
Kanizsa Plusses	inf.	336	521	139	2.43×10^7	3
Woven Square	53.4	264	441	134	1.56×10^7	1
Y's with Bars	249.8	408	829	337	1.14×10^8	1
X's with Bars	inf.	318	741	120	2.83×10^7	1

Table 5.4: 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 5.5: 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

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 percent. The average reduction in the total number of floating point multiplies, for the same five instances is 97 percent.

5.4 Context Dependency of Figural Completion

Although the philosophy of the Gestalt psychologists is supposedly captured by the sentence "The whole is greater than the sum of its parts," Kanizsa[29] points out that this statement is actually somewhat of a misrepresentation. Gestalt philosophy, in small part, was an early 20th century attempt to describe a set of phenomena in human vision using metaphors from physics[34, 35]. For example, the Gestalt "field" is the medium through which visual stimuli exhert influence on one another in the process of perceptual organization. This is analogous to the electromagnetic field, which mediates the attractive force among charged particles. While the failure of the Gestalt program is often attributed to its lack of the metaphor of computation, Gestalt demonstrations can still provide important clues to the nature of the computational processes underlying human vision.

As a computational problem, early vision effects a mapping from an image on the retina (the *stimulus*) to a more abstract intermediate representation (the *percept*). Both the stimulus and the percept can be thought of as vectors in appropriate vector spaces. Assuming that the components of the stimulus vector are indexed spatially, then the *spatial support* of a function can be defined as the subset of the components of the stimulus vector required to compute that function. The spatial support of a function is said to be *local* when it occupies a bounded area of the retina and *global* when it occupies the entire retina. As for the percept vector, although the precise details are unimportant, it is reasonable (at least) to expect that some subset of its components represent the figure-ground sense of different image contours.

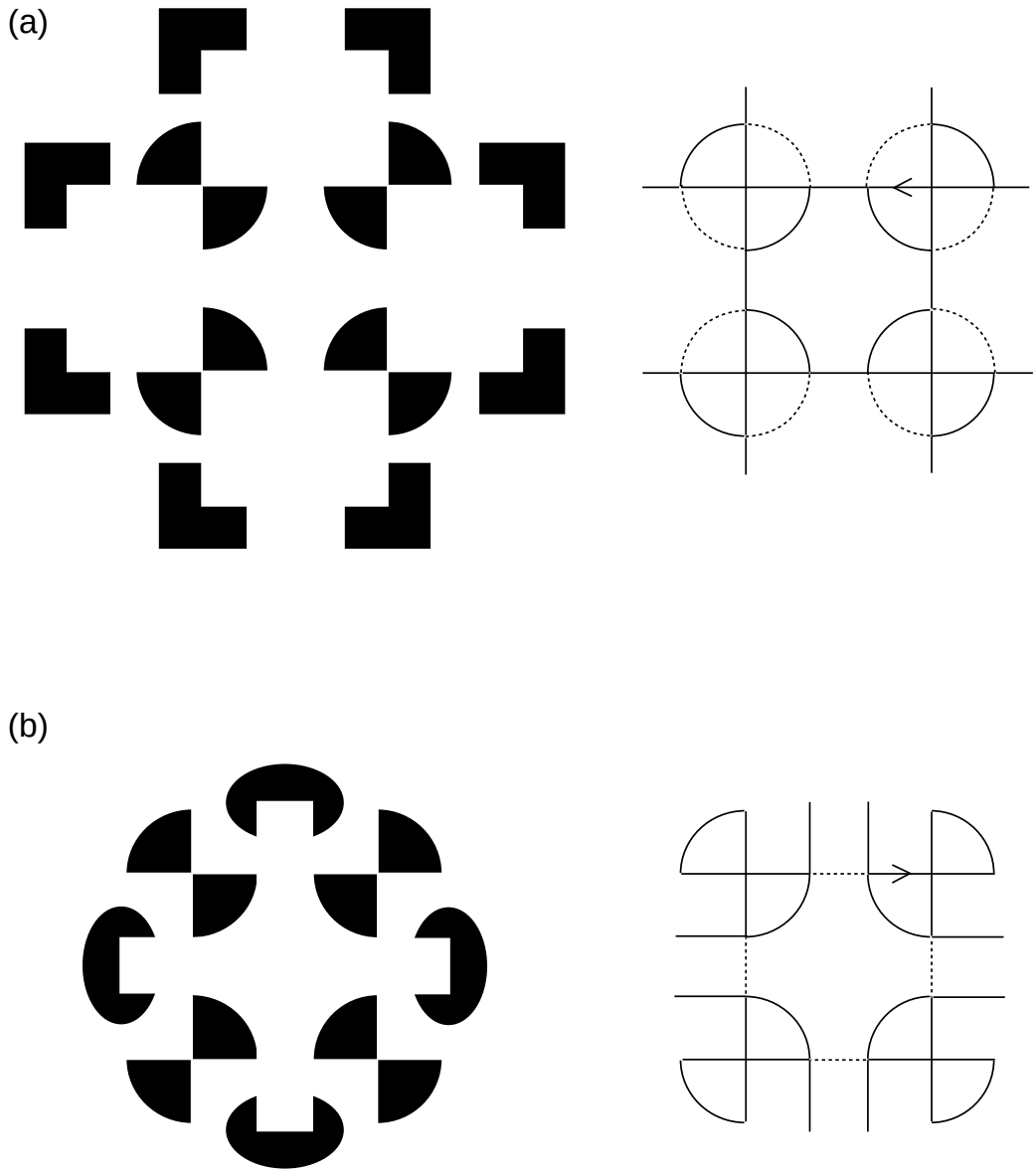

Figure 5.7: Context dependency in Gestalt display. Although, the central portion of the upper and lower displays is the same, identical elements in different contexts play markedly different roles. In particular, the figure-ground sense of the contour marked with the arrow is opposite in the two different contexts. This demonstration was designed by Kanizsa[29].

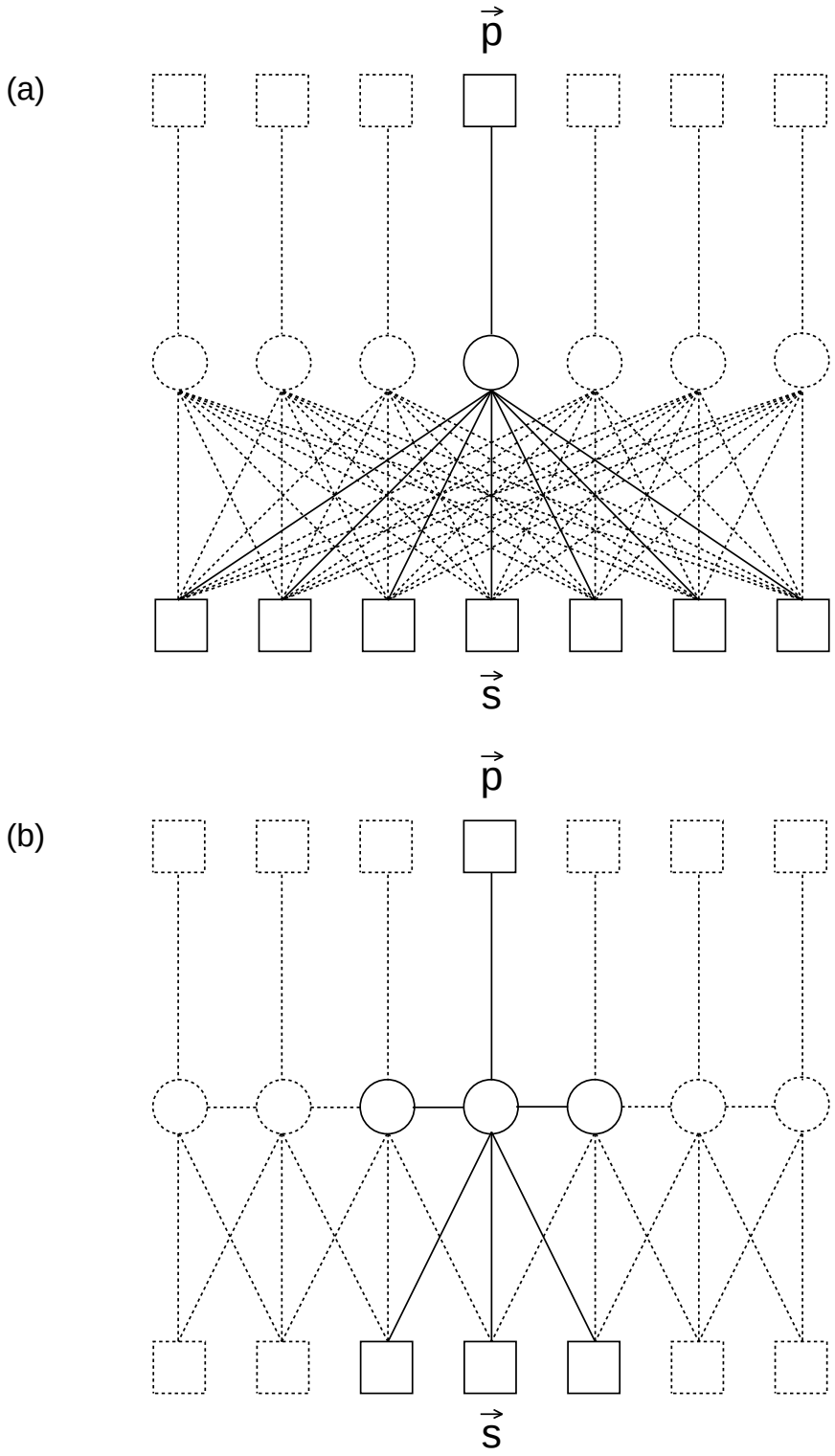

Figure 5.8: Two extremes in network computation. (a) A network computing a global transform. (b) A relaxation network.

One of the connotations of the word "Gestalt" is irreducible whole. Using the language developed in the previous paragraph, we might say that many Gestalt demonstrations strongly suggest that the function mapping the stimulus to the percept can not be decomposed into independent functions computable with local spatial support. Ullman[58] called a device for computing such a non-decomposable function a "Gestaltron." The fact that the computation underlying illusory contours possesses this property is dramatically illustrated in Figure 5.7, a demonstration designed by Kanizsa[29]. Although, the central portion of the upper and lower displays is the same, identical elements in different contexts play markedly different roles. In particular, the figure-ground sense of the contour marked with the arrow is opposite in the two different contexts.

5.5 Network Computation

Ballard[2] considered the structure of networks which might compute the mapping function and identified two fundamental types. The first are networks consisting of independent processing elements with global spatial support. These networks are actually global associative memories, and in principle, are capable of implementing completely arbitrary mappings of input to output. However, because interconnectivity requirements are so severe, their application is limited in practice to mappings between low dimensional parameter spaces. A typical example is a network which detects straight lines by computing a global Hough transform[24], where the components of the percept vector represent individual straight lines parameterized by their orientation and distance from the origin.

In a recent paper, Guy and Medioni[16] describe an abstract transform for computing a tangent field representing global image structure from local tangent measurements. Like the Hough transform, the key to their approach is the local summation

of a set of global voting patterns. Unlike the Hough transform, the accumulator is spatially registered with the image and the voting pattern is a vector, not a scalar field. Elements of the vector field represent orientations which are co-circular to the tangent measurements. Also, the magnitude of the field decreases exponentially with distance. The vectors are combined locally through analysis of moments, and the principal axis of the vectors which accumulate at a location is used as an estimate of the dominant direction.[2]

The second type consists of processing elements with local spatial support coupled in a locally connected network. For example, Parent and Zucker[41] describe a network method for computing a discrete tangent and curvature field consistent with local orientation and curvature measurements produced by operators analogous to simple cells in the visual cortex. Their network solves a *relaxation labeling* problem[27, 50]. A relaxation labeling problem requires that one element of a set of labels be assigned to each element of a set of objects. The probability that object i has label λ is denoted by $p_i(\lambda)$. The compatibility of object i possessing label λ and object j possessing label λ' is denoted by $r_{ij}(\lambda, \lambda')$. The goal is to find the most probable assignment of labels to objects subject to the compatibilities. This is accomplished through an iterative gradient ascent algorithm. The update is computed locally and in parallel for each object. In Parent and Zucker's network, the λ represent discrete orientation and curvature values at image point i. The $p_i(\lambda)$ are assigned values proportional to the initial tangent and curvature responses of the simple cells. The $r_{ij}(\lambda, \lambda')$ reflect the degree of co-circularity of tangent and curvature labels at points i and j. In the course of relaxation, local tangent and curvature measurements are adjusted to reflect global context.

[2]Although Guy and Medioni never explicitly formulate an optimization problem, and although their method of computation is very different from relaxation, their choice of computational goal and their use of the co-circularity constraint is very much in the spirit of the earlier work of Parent and Zucker[41].

Ullman's[58] minimal mapping theory of the correspondence problem in apparent motion is a second example of a global optimization problem solved by a locally connected network of processing elements. Apparent motion in human vision is a natural consequence of a presumed correspondence between elements of the visual field (i.e. correspondence tokens) at two different points in time. Since there are many possible one-to-one mappings between correspondence tokens at two points in time, the correspondence problem in apparent motion is underconstrained. The problem of computing a unique mapping is formulated by Ullman as a linear programming problem.

In Ullman's formulation, q_{ij} is the logarithm of the likelihood that token i at time t matches token j at time $t+1$. Within the linear program, 0-1 valued variables x_{ij} form a permutation matrix which defines the mapping between tokens at the two times. A priori knowledge of the probability density function defining the distribution of image velocities provides the values of the different q_{ij}. Assuming independence, the maximum likelihood correspondence maximizes the sum of the product of x_{ij} and q_{ij} subject to linear inequality constraints which ensure that every token has at least one match.

The constraint matrix of this particular linear program is of a special type possessing a property known as *total unimodularity.* The basic feasible solutions of such a linear programming problem always have integer components (See Figure 5.9). In addition to bipartite matching, other linear programs with this property include maximum network flow. This has very important practical consequences. Since the solution vectors always have integer components, the constraint that the x_{ij} be 0-1 valued need not be explicitly enforced, as it must be in more general integer linear programs. Ullman suggests that the problem of computing the optimal feasible solution can be solved by a gradient search method due to Arrow and Hurwicz[1]. The

gradient search locates a saddle point in a Lagrangian constructed according to the theory of Lagrange multipliers. Significantly, the different components of the Lagrangian gradient can be computed locally and in parallel, which leads to a network implementation.

5.6 Total Unimodularity

In this section we briefly discuss the notion of total unimodularity and its importance to integer linear programming. Understanding total unimodularity will allow us to gauge the feasibility of reformulating the figural completion problem as a set of more tractable linear programming problems (i.e. LP's), which can be solved by the method of Arrow and Hurwicz[1] (following the suggestion of Ullman[58]). The key ideas are from Papadimitriou and Steiglitz[40].[3]

Although for reasons of clarity, both equality and inequality constraints have been used when defining integer linear programs, any ILP can be reduced to the following canonical form:

$$\begin{aligned} \max \; & c^t \; \mathbf{x} \\ \text{subject to } & A \; \mathbf{x} \; \leq \; b \\ & \mathbf{x} \; \geq \; 0 \\ & \mathbf{x} \;\; \text{integer} \end{aligned}$$

Here, A, b and c have only integer components.

A square, integer matrix is *unimodular* if its determinant is ± 1. An integer matrix is *totally unimodular* (TUM) if and only if the determinant of every square, non-

[3]Specifically, the use of Theorems 5.1 and 5.2 to establish a sufficient condition for an integer linear program having only integer solutions is due to Papadimitriou and Steiglitz[40].

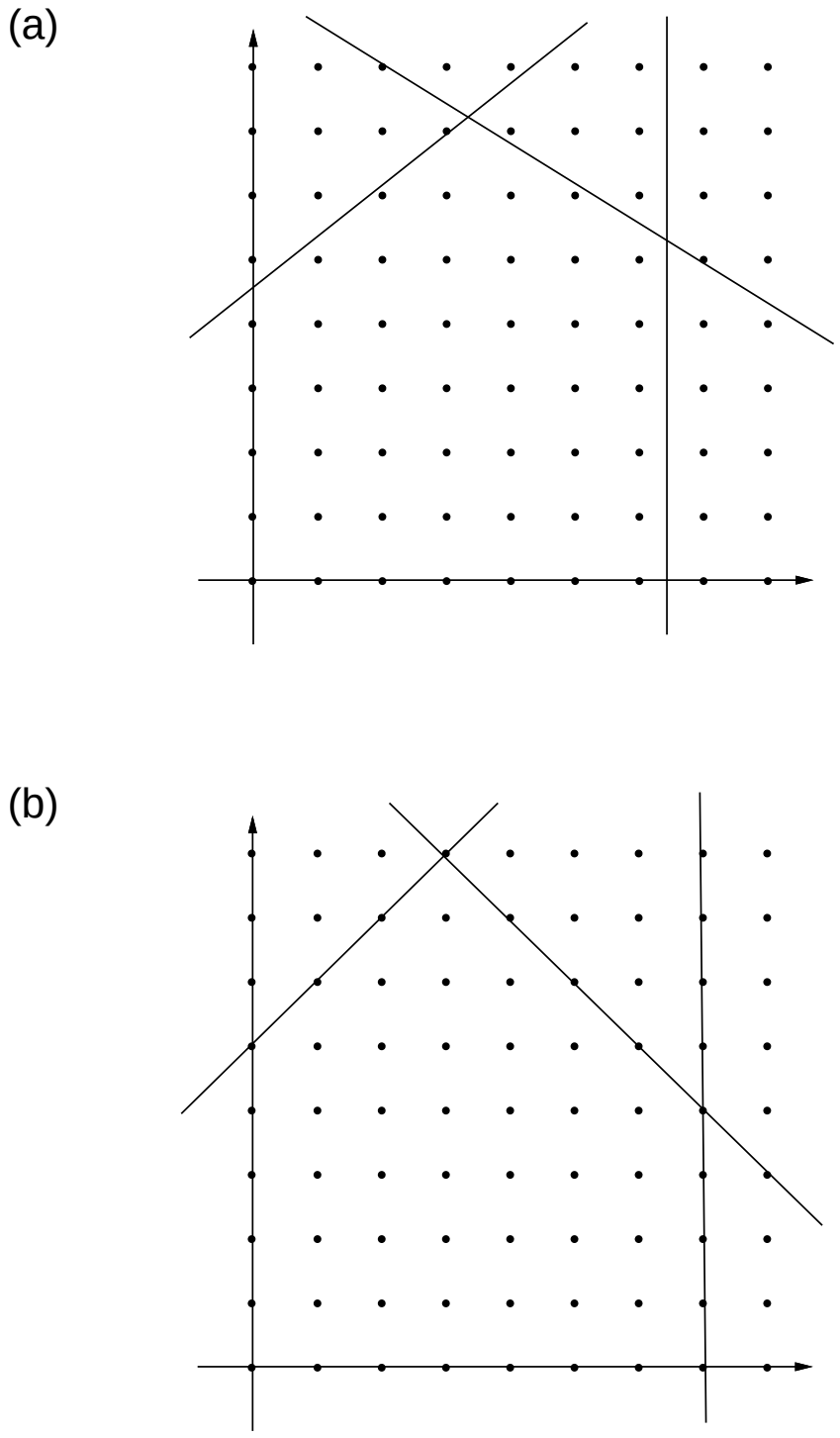

Figure 5.9: Total unimodularity. (a) An integer linear programming problem with a non-unimodular constraint matrix. (b) An integer linear programming problem with a totally unimodular constraint matrix. The basic feasible solutions all lie on the integer lattice.

singular submatrix is ± 1. The following theorem, due to Heller and Tomkins[17] establishes a sufficient condition for total unimodularity.

Theorem 5.1 (Heller and Tompkins) *An integer matrix A with $a_{ij} = 0, \pm 1$ is TUM if no more than two nonzero entries appear in any column, and if the rows of A can be partitioned into two sets such that:*

1. *If a column has two entries of the same sign, their rows are in different sets;*

2. *If a column has two entries of different signs, their rows are in the same set.*

The following theorem, due to Hoffman and Kruskal[20], establishes the connection between total unimodularity and integer basic feasible solutions:

Theorem 5.2 (Hoffman and Kruskal) *If A is TUM and b is integer, then every basic feasible solution is integer.*

This theorem can be appreciated by noting that the simplices of the constraint polytope of a linear program are defined by square submatrices within its tableau. The basic feasible solutions are equal to the product of the inverses of square submatrices and the vector b. By Cramer's rule, the inverse of an integer submatrix will be integer if its determinant is equal to ± 1. Since b is also integer, this ensures that the basic feasible solution is integer.

5.7 A Complexity Result

We will show that ILP_{unit} is TUM under certain additional assumptions which will now be described. Recall that $G_{non-planar} = (V_{endpoints}, E_{fragments} \cup E_{completions})$. Since the fragments in $E_{fragments}$ are also luminance boundaries, it is possible to assign an orientation (i.e. sign of contrast) to each fragment so that the darker of the two regions it borders lies to the right when a fragment is traversed in the direction

of its orientation. Now, divide the set of endpoints, $V_{endpoints}$, into two disjoint sets, P and Q, such that each endpoint $p \in P$ is an initial endpoint of some fragment and each endpoint $q \in Q$ is a terminal endpoint of some fragment.

Assuming that the sign of contrast remains constant around every boundary, then every completion in $E_{completions}$ is adjacent to exactly one element of P and one element of Q. $G_{non-planar}$ is therefore a bipartite graph.[4]

We now show that if $G_{non-planar}$ is bipartite then the rows of the constraint matrix defining ILP_{unit} can be divided into two disjoint sets satisfying the conditions of Theorem 5.2.

Theorem 5.3 *If $G_{non-planar}$ is bipartite, then ILP_{unit} is TUM.*

Proof Divide the rows of the constraint matrix into two sets, such that a row is in set I_p if it is of the form:

$$x_p - \sum_{j \in completions(p)} x_j = 0 \tag{5.4}$$

(where $p \in P$) and a row is in set I_q if it is of the form:

$$x_q - \sum_{j \in completions(q)} x_j = 0 \tag{5.5}$$

(where $q \in Q$).

The variables of ILP_{unit} can themselves be divided into two sets, namely those representing boundary fragments (i.e. $\{x_p \mid p \in P \cup Q\}$) and those representing completions (i.e. $\{x_j \mid j \in E_{completions}\}$). Observe that in both Constraint 5.4 and Constraint 5.5 the column entries for variables representing boundary fragments are always +1 (i.e. x_p and x_q) while those representing completions (i.e. x_j) are always

[4]If this were not true, it would imply that there is a completion which joins two fragments with opposite signs of contrast, which contradicts our assumption.

-1. Since the column entries for a given variable are always of the same sign, to prove that ILP_{unit} is TUM, it is sufficient to show that every column has exactly two non-zero entries, one in a row in I_p and one in a row in I_q. This is shown first for variables representing boundary fragments.

By convention, if p and q are opposite endpoints of a single boundary fragment then $x_p = x_q$ so that these variables form a single column of the constraint matrix. Clearly, one of these entries (i.e. x_p) is in I_p and the other (i.e. x_q) is in I_q.

The situation for a variable representing instantiation of a completion is only slightly more involved. We observe that since $G_{non-planar}$ is bipartite, for any two endpoints $p, p' \in P$, it is the case that $completions(p) \cap completions(p') = \emptyset$. It follows that the column of the constraint matrix representing variable x_j, where $j \in completions(p)$ has at most one non-zero entry among the rows comprising set I_p. Since for any two endpoints $q, q' \in Q$, it is the case that $completions(q) \cap completions(q') = \emptyset$ it also follows that the column of the constraint matrix representing variable x_j has at most one non-zero entry among the rows comprising set I_q. Since the column of the constraint matrix representing variable x_j is non-zero at most twice (i.e. once within a row of I_p and once within a row of I_q), this establishes that ILP_{unit} is TUM when $G_{non-planar}$ is bipartite.□

5.8 Other Considerations

Until this chapter, there had been no attempt to study the problem of figural completion at any other level than that of computational theory. However, in this chapter, the computational theory was revised. This revision was motivated in part by evidence from human vision and in part by speculation about biologically plausible forms of computation. This was the first and only time that issues of algorithms and representations were entertained.

The result of that analysis was fairly modest: since the integer linear program defining the unit organization problem is TUM, it can, in principle, be solved by numerical relaxation in a locally connected network.[5] Still, this falls far short of proposing a biologically plausible algorithm and representation. Probably the best precedent for this is the work of Parent and Zucker[41], which was briefly described earlier in this chapter.

Their work is notable for a number of reasons. First, the representation passes the first litmus test for biological plausibility: in contrast with many neural network models,[6] contour organizations in Parent and Zucker's scheme are represented as linear combinations of a *fixed* basis set of tangent/curvature "units."[7] Unless the brain has evolved dynamic memory allocation and garbage collection, the use of a fixed set of units is absolutely essential.

Second, when compared to representing contours by "grouping items" or "tokens," the discrete tangent/curvature field possesses important representational advantages. Recall that in Chapter 4, a procedure for identifying and eliminating "redundant" completions was described. A completion is considered redundant when its likelihood is lower than the product of the likelihoods along an alternative path joining the same fragments. Because sets of redundant completions overlap in the image, they are likely to be products of a single physical process. However, because completions are represented as distinct tokens in the experimental system, overlapping contours significantly increase the complexity of the grouping problem. The irony is that, in an appropriately designed parallel distributed representation, such as the discrete

[5] An open problem is whether or not some variation of the depth labeling problem can be shown to be TUM as well. As defined, it very clearly is not. Again, the awkward depth order requirement is to blame: it can not be written as a set of integer linear inequalities with ± 1 coefficients.

[6] Including Ullman's network formulation of the correspondence problem in apparent motion, which motivated the use of linear programming in this thesis.

[7] Here the word *unit* is synonymous with "neural unit," and has nothing to do with the Gestalt concept of unit.

tangent/curvature field, overlapping stimuli could combine additively, and therefore reinforce a single percept.

This "redundancy" phenomenon, more than any other, illustrates that token representations are ill-suited to the task of representing and grouping image contours. The problem is not peculiar to the experimental system described in this thesis, but shows up whenever token representations are used in contour grouping. When used to represent image contours, tokens make all the wrong things explicit. For example, an artist drawing a rough charcoal sketch draws contours with sets of crude overlapping strokes. Where more emphasis is required, the artist retraces old strokes with bolder, darker strokes. The precise image trace is unimportant. Even less important is where individual strokes begin and end. If these observations are correct, then why are the locations of a contour's endpoints, together with its precise trace, the basis of most contour representations? In the case of a charcoal sketch, tokens faithfully model the idiosyncrasies of the method of manufacture, but fail to make plain what the eye actually sees.

Finally, and of special significance, is the fact that Parent and Zucker's representation permits multiple tangent/curvature labels to exist at a single image location. Because this allows contours to cross without penalty, it suggests that the discrete tangent and curvature field can be adapted to the problem of representing the different graphs (i.e. G_{planar}, etc.) required by our formulation of the figural completion problem. With some cleverness, it might even be possible to cast the different subproblems required to construct a labeled knot diagram as relaxation labeling problems using this same set of tangent/curvature units. This is a promising direction for future work.

CHAPTER 6

CONCLUSION

In their choice of goals, representations and methods, conventional theories of visual reconstruction have failed to address the problem of reconstructing environmental structure which is not in plain sight. Because conventional theories assume that parallel projection maps visible surface neighborhoods to image neighborhoods in one-to-one fashion, the problem of deducing the topology of environmental structure has largely gone unrecognized. This thesis, in contrast, has suggested that deducing the topology of environmental structure, whether visible or occluded, is the fundamental problem of perceptual organization, and that figural completion phenomena must be understood in this context.

6.1 Contributions

The major contribution of this thesis is a computational theory of figural completion. By computational theory, it is meant first of all that the goal of the computation has been clearly elucidated (this was the topic of Chapter 2). It is my opinion that, in this regard, labeled knot diagrams are as fundamental to the theory of figural completion as correspondence is to the theory of apparent motion. Second, the natural constraints which assist the human visual system in achieving this goal, together with the inherent ambiguities, have been identified. Here, the topological constraints implicit in the labeling scheme play much the same role as rigidity does in theories

of motion understanding. Together, these elements allowed the figural completion problem to be formulated abstractly as a combinatorial optimization problem. This combinatorial optimization problem, in the form of an integer linear program, was described in detail in Chapter 3.

One of the things which distinguishes a good theory from a bad theory is that a good theory can be wrong, that is, it makes predictions which can be tested. While the jury is still out on whether or not the computational theory proposed here is correct, it is specific enough that it allows some testable predictions to be made:

- In Chapter 2, it was conjectured that the goal of figural completion is to compute a labeled knot diagram representing the boundaries of the anterior surfaces. This leads to the following prediction: In human vision, the image traces of perceptual completions (whether modal or amodal), together with their signs of occlusion and relative depths at points of intersection, are *explicitly* represented.

- In Chapter 3, we observed that the problem of computing a labeled knot diagram from visible boundary fragments requires overcoming three different forms of ambiguity. These were termed *shape*, *unit* and *depth*. It was conjectured that shape ambiguity is resolved independently, and in advance of unit and depth ambiguity. In this way, it was suggested, figural completion is reduced to a combinatorial optimization problem. This leads to the prediction that the shape of a perceptual completion is independent of the subset of completions which is instantiated, their signs of occlusion and relative depths.

- Also in Chapter 3, it was conjectured that the shape of a perceptual completion is solely a function of the tangent, curvature and relative positions of the endpoints of the occluded boundary. This leads to the prediction that completions joining pairs of boundary fragments sharing these attributes will always be of the same shape.

- Finally, in Chapter 5, we considered the hypothesis of Kellman and Loukides[32] that unit ambiguity is resolved independently and in advance of depth ambiguity. This leads to the prediction (partially confirmed) that topologically invalid unit organizations exist.

Although the primary reason for building the experimental system (described in Chapter 4) was to test the computational theory which is the major contribution of this thesis, its algorithms and representations should interest researchers in computer vision also. It is widely acknowledged that perceptual organization is among the most difficult problems facing researchers in computer vision today. In my opinion, research in bottom-up visual reconstruction has virtually come to a standstill because of lack of progress in this area. This thesis has explored a previously unidentified source of constraints in image contour grouping—that the grouped contours must form the boundaries of topologically valid surfaces. The usefulness of the labeling scheme in limiting search in computer vision application areas such as visual recognition is independent of whether or not the computational theory outlined here correctly models human vision.

6.2 Future Directions

Several topics for future research were introduced in previous chapters. Some of these topics involved extending or modifying the computational theory while others involved formulating a complementary theory of figural completion phenomena at the level of algorithm and representation:

- Demonstrating that labeled knot diagrams, as defined in Chapter 2, can be used to represent the anterior surfaces of scenes composed of smooth manifold-solids.

- In the surface organization model, integrating figure-ground and other depth preference criteria with unit preference criteria in a single objective function without introducing a bias for (or against) organizations containing larger numbers of completions.

- Determining whether or not topological validity can be ensured without committing to a specific depth ordering of the completed boundaries. This addresses the issue raised by the example of the *Ehrenstein* figure in Chapter 4.

- Identifying additional unit preference criteria which might account for the inability of the experimental implementation of the unit-depth organization model to correctly organize the *Kanizsa Plusses* figure.

- Determining whether or not there exists some approximation of the depth labeling subproblem of the unit/depth organization model (i.e. ILP_{depth}) which is totally unimodular.

- Investigating how the labeled knot diagram and other graphs central to our formulation of the figural completion problem might be computed using biologically plausible representations (e.g. the discrete tangent/curvature field of Parent and Zucker[41]) and algorithms (e.g. relaxation labeling).

Some directions for future work have not been previously discussed. For example, probably the most important "unspoken" assumption of this thesis is the issue of scale. Implicit in the computational theory, as formulated, is the assumption that tangent and curvature have unique values at boundary fragment endpoints. The theory also assumes that the boundary fragments which the figural completion process organizes are the result of segmenting image luminance boundaries at points of high curvature. Unfortunately, unlike mathematically ideal curves, the tangent and

curvature of image contours are functions of the scale at which they are measured. The problem of segmenting a set of image contours into piecewise smooth segments, or even identifying criterion by which alternate segmentations can be compared, is difficult and largely unsolved.[1]

In all of the examples considered in this thesis, the tangents and curvatures have been stable in "scale-space," that is, they remain constant across an order-of-magnitude range of scales which happens to overlap the size of gaps due to occlusion. Yet, in the real world, the backs of cats with fur standing on end are perceptually completed behind scratching posts covered with shag carpet. As Marr[38] and others have pointed out, image contours originate in physical processes operating at multiple natural scales. Therefore, it is essential that any comprehensive theory of figural completion address the issue of scale.

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 (e.g. Figures 2.1– 2.3) suggests that the representation computed by the figural completion process is more general than (and probably subsumes) the labeled knot-diagrams developed here.

Towards this end, it is worth re-examining Huffman's[26] 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 developed in this thesis 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.

[1]See the recent paper by Saund[52] for a review of the literature on this topic.

APPENDIX

GLOSSARY

This appendix is a glossary of general and specific terms used in Chapter 2.

General Terms

- **boundary** The set of points of a surface with boundary with neighborhoods homeomorphic to half-discs.
- **boundary edge** In a paneling representing a surface with boundary, an edge which is not identified, and consequently forms part of the boundary.
- **boundary vertex** A point where two boundary edges are incident, and with a neighborhood homeomorphic to a half-disc.
- **contour generator** The locus of points on the surface of a smooth manifold-solid which are tangent to the viewing direction.
- **crossing** In a knot diagram, a point where one contour passes over another.
- **cusp** The image of a point on the surface of a smooth manifold-solid where the direction of the contour generator coincides with the viewing direction.
- **embedding** A homeomorphism between one topological space and a subspace of another. For example, between a set of orientable surfaces with boundary

and three-space. Because an embedding is a homeomorphism, there are no singularities, and embedded surfaces can not intersect.

- **genus** For orientable surfaces without boundary, the number of handles which must be added to a sphere to create a topologically equivalent surface.
- **handle** A tube whose two open ends are added to a surface to increase its genus by one.
- **homeomorphism** A continuous, one-to-one mapping with continuous inverse.
- **identification** In a paneling, an explicit indication that two edges of equal length are to be glued together and in which way.
- **immersion** A locally homeomorphic mapping between one topological space and a subspace of another.
- **interior edge** In a paneling representing a surface, an edge which is identified with exactly one other edge, and consequently forms part of the surface's interior.
- **interior vertex** A point in a paneling where two or more interior edges are incident, and with a neighborhood homeomorphic to a disc.
- **Jordan curve** A closed non-self-intersecting path on any surface.
- **knot** An embedding of a circle in three-space.
- **knot diagram** The generic projection of a knot onto a plane.
- **manifold-solid** an object whose boundary is an orientable surface which divides three-space into an interior and an exterior.

- **multiplicity** The number of points in the domain of a function which map to a single point in the range.
- **neighborhood** A subset of a topological space is a neighborhood of a point p if and only if it contains every other point of that space which lies within a sufficiently small ball centered on p.
- **network** A graph representing a system of difference equations.
- **occluding contour** The projection onto the image plane of the points on a surface which are tangent to the lines of sight.
- **oriented knot diagram** A knot diagram where each contour is assigned one of two orientations.
- **panel** A surface cutout from a flat sheet of paper. The boundary of a panel is formed by an ordered cycle of edges separated by vertices.
- **paneling** A set of panels together with a set of edge identifications which determine which panel edges are glued together and in which ways. A paneling where every neighborhood is homeomorphic to either a disc or half-disc represents a surface with boundary.
- **punctured torus** A surface with boundary of genus one formed by subtracting a disc from a torus.
- **singularity** A point with multiplicity greater than one under some mapping.
- **writhe** For a crossing in an oriented knot diagram, the sign of the cross product of the upper and lower strands.

Specific Terms

- **anterior scene** A set of orientable surfaces with boundary embedded in three-space such that the surface normals everywhere are defined and have a positive component in the viewing direction.

- **anterior surfaces** The locus of environmental surface points where the surface normal is defined and has a positive component in the viewing direction.

- **boundary depth** The number of surfaces which lie between a surface boundary and the eye.

- **depth order requirement** The requirement that the depth of the occluding surface boundary be less than or equal to the depth of the occluded surface boundary.

- **identification scheme** An implicitly defined set of edge identifications. In this case, the identification scheme specifies which panel edges should be glued together to construct a paneling from a labeled knot diagram.

- **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 viewing direction.

- **labeled knot diagram** A representation of an anterior scene consisting of an oriented knot diagram annotated with boundary depths satisfying the labeling scheme depicted in Figure 2.5.

- **planar region** One of a set of regions into which a knot diagram divides the plane.

- **visible surfaces** The locus of environmental surface points first incident along the lines of sight.

BIBLIOGRAPHY

[1] Arrow, K.J., Huricz, L., and Uzawa, H., *Studies in Linear and Nonlinear Programming*, Stanford University Press, 1958.

[2] Ballard, D.H., Parameter Networks, *Readings in Computer Vision*, M. Fischler and O. Firschein (eds.), Morgan Kauffman, Los Altos, pp. 534-550, 1987.

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

[4] Barrow, H.G and Tenenbaum, J.M., Retrospective on Interpreting Line Drawings as Three Dimensional Surfaces, *Artificial Intelligence 59*, pp. 71-80, 1993.

[5] Blake, A. and Zisserman, A. *Visual Reconstruction*, MIT Press, Cambridge, Mass., 1987.

[6] Boldt, M., Weiss, R. and Riseman E.M., Token Based Extraction of Straight Lines, *IEEE Transactions on Systems, Man and Cybernetics, Vol. 19, No. 6*, pp. 1581-1594, 1989.

[7] Burns, J.B., Matching 2D Images to Multiple 3D Objects Using View Description Networks, Ph.D. Dissertation, Dept. of Computer Science, University of Massachusetts, 1992.

[8] Callahan, J., Singularities and plane maps, *Am. Math. Monthly 81*, pp. 211-240, 1974.

[9] Dolan, J. and Riseman E.M., Computing Curvilinear Structure by Token-based Grouping, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 264-270, 1992.

[10] Duda, R.O. and Hart, P.E., *Pattern Classification and Scene Analysis*, John Wiley and Sons, New York, 1973.

[11] Elder, J. and Zucker, S.W., The Integration of Figure Fragments into Representations of Planar Shape, Technical Report, McGill Research Centre for Intelligent Machines, McGill University, Montreal, Quebec, 1993.

[12] Fahle, M. and Palm, G., Perceptual Rivalry between Illusory and Real Contours, *Biological Cybernetics Vol. 66, No. 1*, pp. 1-8, 1991.

[13] Gibson, J.J., *The Perception of the Visual World*, Greenwood Press, Westport, Conn., 1950.

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

[15] Grossberg, S., and Mingolla, E., The Role of Illusory Contours in Visual Segmentation, *The Perception of Illusory Contours*, Petry, S. and Meyer, G. (eds.), Springer-Verlag, New York, pp. 116-125, 1987.

[16] Guy, G. and Medioni, G., Inferring Global Perceptual Contours from Local Features, *Proc. of the DARPA Image Understanding Workshop*, Washington, D.C., pp. 881-892, 1993.

[17] Heller, I. and Tompkins, C.B., An Extension of a Theorem of Dantzig's, *Linear Inequalities and Related Systems*, Kuhn and Tucker (eds.), Princeton University Press, pp. 247-252, 1956.

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

[19] Hilbert, D. and Cohn-Vossen, S., *Geometry and the Imagination*, Translated from the German by P. Nemenyi, Chelsea Publishing Co., New York, 1952.

[20] Hoffman, A.J. and Kruskal, J.B., Integral Boundary Points of Complex Polyhedra, *Linear Inequalities and Related Systems*, Kuhn and Tucker (eds.), Princeton University Press, pp. 223-246, 1956.

[21] Horn, B.K.P., Obtaining Shape from Shading Information, *The Psychology of Computer Vision*, P.H. Winston (ed.), McGraw-Hill, New York, pp. 115-155, 1975.

[22] Horn, B.K.P., The Curve of Least Energy, MIT AI Lab Memo No. 612, Artificial Intelligence Laboratory, MIT, Cambridge, Mass., 1981.

[23] Horn, B.K.P., *Robot Vision*, MIT Press, Cambridge, Mass., 1986.

[24] Hough, P., *A Method and Means for Recognizing Complex Patterns*, U.S. Patent No. 3,069,654, 1962.

[25] Hubel, D.H., and Wiesel, T.N., Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex, *Journal of Physiology 160*, pp 106-154, 1962.

[26] Huffman, D.A., Impossible Objects as Nonsense Sentences, *Machine Intelligence 6*, pp. 295-323, 1971.

[27] Hummel, R.A., and Zucker, S.W., On the Foundations of Relaxation Labeling Processes, *IEEE Transactions on Pattern Analysis and Machine Intelligence 5*, pp. 267-287, 1983.

[28] Jacobs, D., The Use of Grouping in Visual Object Recognition, M.S. Thesis, Dept. of E.E. and C.S., MIT, Cambridge, Mass., 1988.

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

[30] Kass, M., Witkin, A. and Terzopolous, D., Snakes: Active Minimum Energy Seeking Contours, *Proc. of the First Intl. Conf. on Computer Vision*, London, England, pp. 259-268, 1987.

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

[32] Kellman, P.J. and Loukides, M.G., An Object Perception Approach to Static and Kinetic Subjective Contours, *The Perception of Illusory Contours*, Petry, S., and Meyer, G., (eds.), Springer-Verlag, New York, pp. 151-164, 1987.

[33] Koenderink, J.J., The Shape of Smooth Objects and the Way Contours End, *Natural Computation*, W. Richards (ed.), MIT Press, Cambridge, Mass., pp. 115-124, 1988.

[34] Koffka, K., *Principles of Gestalt Psychology*, Harcourt, Brace and Company, New York, 1935.

[35] Kohler, W., *The Task of Gestalt Psychology*, Princeton University Press, Princeton, New Jersey, 1969.

[36] Kumar, R., Model Dependent Inference of 3D Information from a Sequence of 2D Images, Ph.D. Dissertation, Dept. of Computer Science, University of Massachusetts, 1992.

[37] Lowe, D., *Perceptual Organization and Visual Recognition*, Kluwer Academic Press, The Netherlands, 1985.

[38] Marr, D., *Vision*, Freeman Press, San Francisco, Cal., 1982.

[39] Nitzberg, M. and D. Mumford, The 2.1-D Sketch, *Proc. of the 3rd Intl. Conf. on Computer Vision*, Osaka, Japan, pp. 138-144, 1990.

[40] Papadimitriou, C.H., and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Prentice-Hall, Englewood Cliffs, N.J., 1982.

[41] Parent, P. and Zucker, S.W., Co-Circularity and the Role of Curvature in Curve Detection, Technical Report, Computer Vision and Robotics Lab, McGill University, Montreal, Quebec, 1985.

[42] Petry, S. and Meyer, G.E. (eds.), *The Perception of Illusory Contours*, Springer-Verlag, New York, 1987.

[43] Petter, G., Nuove ricerche sperimentali sulla totalizzazione percettiva, *Rivista di Psicologia 50*, pp. 213-227, 1956.

[44] Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T., *Numerical Recipes in C*, Cambridge University Press, 1988.

[45] Richards, W., The Approach, *Natural Computation*, W. Richards (ed.), MIT Press, Cambridge, Mass., pp. 3-13, 1988.

[46] Richards, W., Feldman, J. and Jepson, A., From Features to Perceptual Categories, *Proc. of the Second European Conference on Computer Vision*, pp. 99-108, 1992.

[47] Roberts, L., Machine Perception of Three-Dimensional Solid Objects, *Optical and Electro-Optical Information Processing*, J. Tippett (ed.), MIT Press, Cambridge, Mass., 1966.

[48] Rock, I., *The Logic of Perception*, MIT Press, Cambridge, Mass., 1983.

[49] Rock, I., A Problem Solving Approach to Illusory Contours, *The Perception of Illusory Contours*, Petry and Meyer (eds.), Springer-Verlag, New York, pp. 62-70, 1987.

[50] Rosenfeld, A., Hummel, R.A., and Zucker, S.W., Scene Labeling by Relaxation Operations, *IEEE Transactions on Systems, Man and Cybernetics, Vol. 6, No. 6*, pp. 420-433, 1976.

[51] Salkin, H.M., and Mathur, K., *Foundations of Integer Programming*, North Holland, New York, 1989.

[52] Saund, E., Identifying Salient Circular Arcs on Curves, *Computer Vision, Graphics and Image Processing: Image Understanding Vol. 58, No. 3*, pp. 327-337, 1993.

[53] Sawhney, H., Spatial and Temporal Grouping in the Interpretation of Image Motion, Ph.D. Dissertation, Dept. of Computer Science, University of Massachusetts, 1992.

[54] Schumann, F., Contributions to the Analysis of Visual Perception–First Paper: Some Observations on the Combination of Visual Impressions into Units, *The Perception of Illusory Contours*, Petry and Meyer (eds.), Springer-Verlag, New York, pp. 21-34, 1987.

[55] Sederberg, T.W. and Parry, S., A Comparison of Three Curve Intersection Algorithms, *Computer Aided Geometric Design 18*, pp. 58-64, 1986.

[56] Strang, G., *Linear Algebra and its Applications*, Harcourt Brace Jovanovich, San Diego, Cal., 1988.

[57] Ullman, S., Filling-in the Gaps: The Shape of Subjective Contours and a Model for Their Generation, *Biological Cybernetics 21*, pp. 1-6, 1976.

[58] Ullman, S., *The Interpretation of Visual Motion*, MIT Press, Cambridge, Mass., 1979.

[59] Williams, L.R., Perceptual Organization of Occluding Contours, *Proc. of the 3rd Intl. Conf. on Computer Vision*, Osaka, Japan, pp. 133-137, 1990.

[60] Witkin, A. P. and Tenenbaum, J.M., On the Role of Structure in Vision, *Human and Machine Vision*, Beck, Hope and Rosenfeld (eds.), Academic Press, pp. 481-543, 1983.