Perceptual Organization of Occluding Contours of Opaque Surfaces

Eric Saund
Xerox Palo Alto Research Center
3333 Coyote Hill Rd.
Palo Alto, CA 94304
saund@parc.xerox.com
(650) 812-4474
(650) 812-4334 (fax)

Abstract

This paper offers computational theory and an algorithmic framework for perceptual organization of contours bounding opaque occluding surfaces of constant lightness. For any given visual scene, a sparse graph is constructed whose nodes are salient visual events such as contrast edges, and L-type and T-type junctions of contrast edges, and whose arcs are coincidence and geometric configurational relations among node elements. An interpretation of the scene consists of choices among a small set of labels for graph elements reflecting physical events such as corners, visible surface occlusion, amodal continuation, and surface occlusion sans visible contrast edge (which perceptually give rise to illusory contours). Any given labeling induces an energy, or cost, associated with physical consistency and figural interpretation biases. Using the technique of deterministic annealing, optimization is performed such that local cues propagate smoothly to give rise to a global solution. We demonstrate that this approach leads to correct interpretations (in the sense of agreeing with human percepts) of popular simple "Colorforms" figures known to induce illusory contours, as well as more difficult figures where interpretations acknowledging accidental alignment are preferred.

1 Introduction

The human visual system is remarkably adept at sorting out the various contrast edges found in images and inferring the surfaces that generated them, making explicit their overlap or depth relations. Often the physical configuration of objects is underconstrained by the limited information available in a single view, and additional constraints or assumptions must be brought into play. Hence the premise underlying all proposed explanations of the Kanizsa Triangle (Figure 1) that the visual system must be "filling in" information about contours not physically present in the image based on rules or mechanisms operating with regard to the surrounding cues. The gestalt psychologists concocted numerous demonstrations involving seemingly simple figures of this sort to show that a wealth of biases and assumptions engage in a complex interplay as the visual system settles on preferred interpretations [16]. The challenge facing the modern computational study of Perceptual Organization is to formalize and extend the gestaltists' intuitive insights in terms of testable theories and implementable algorithms.

This paper assembles a computational theory underlying perceptual organization of occluding surfaces largely from components already existing in prior literature, but argues that the more difficult problem is the design of representations and procedures satisfying constraints at the algorithm level. For this we propose a novel approach incorporating several well-known computational techniques including token grouping, graph labeling, the construction of energy surfaces, and optimization by continuation methods.

While many accounts have been proposed for the shapes adopted phenomenally by illusory contours, e.g. [22, 7, 28], explanations for *why* they are seen at all are of two sorts. One class of explanation treats the image itself as the primary object of interest, which it is the visual system's job to elaborate and refine. Accordingly, mechanisms

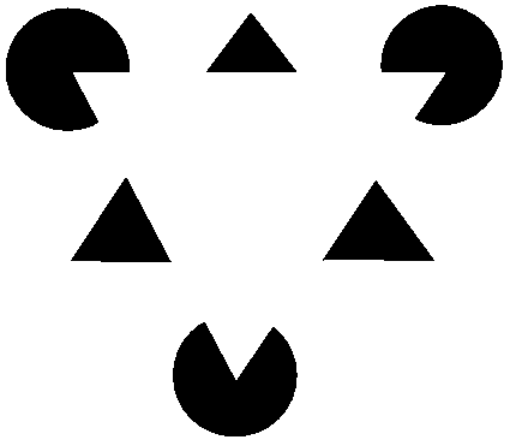
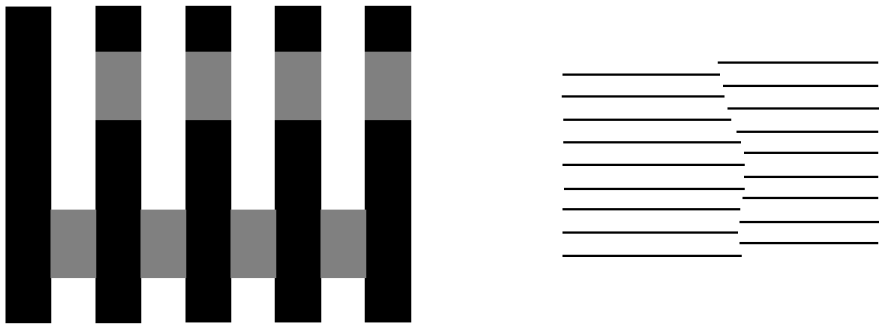


Figure 1: Kanizsa Triangle

are proposed whereby certain types of visible image events such as contrast edges, contour junctions, and thin lines and their endpoints, cause the assertion of additional events in the image representation such as illusory contours and surface brightness variations [7].

The second class of explanation, shared by the present work, maintains that illusory contours are only a byproduct of computations for which the primary goal is to infer $2\frac{1}{2}$-dimensional properties of the physical world such as surfaces, their colors, their boundaries, and the occlusion relations among them [13]. These explanations observe empirically that virtually all illusory contour phenomena are associated with evidence of surface depth discontinuities [9], and they embrace extensions to the theory accounting for transparency effects. See Figure 2. Moreover, this class of theory falls within the modern computational vision paradigm conceiving perception as the application of deeply justified prior knowledge and assumptions about the world to



a. b

Figure 2: a. Image lightness illusion for which Anderson [2] has proposed an explanation based on inferences about surface transparency. b. This figure sometimes held as a counterexample to surface depth inference as an explanation of illusory contours because neither surface is perceived to lie in front of the other. Note however that each line termination along the curving illusory contour is in fact *local* evidence for occlusion, and the lesson to be taken from the global percept may instead be that the representation can be factored: the human visual system is capable of declaring the *presence* of a surface occlusion boundary without committing to the *direction* of occlusion.

underconstrained input data, in order to infer information not directly measurable. For example, it is widely accepted that the visual system prefers interpretations of image events that arise generically instead of interpretations requiring postulation of unlikely "accidents" of arrangement, lighting, motion, viewpoint, etc. [10, 23].

From this vantage point of seeking to infer physical properties from underconstrained image data, Williams [26, 27] pioneered a formulation for "Colorforms"[1] figures adapting the classic line labeling approach of Guzman [8] and Waltz [24], in which the contour interpretation problem becomes one of assigning labels to sparse image events based on two sorts of constraints, physical feasibility and figural biases.

[1] *Colorforms* is a popular toy consisting of vinyl sheets of plastic cut as graphic objects which are laid out to create pictures. *Construction paper* is a similar medium.

More recently, Geiger et al.[6] formulated a dense field relaxation labeling approach whereby interpretation labels are diffused, pinned at the relatively few locations containing contrast data in the input image. Each of these approaches carries drawbacks which are alluded to below; a new synthesis is called for which builds upon this progress.

This paper offers a new formulation for the perceptual organization of occluding contours. Computational theory is developed that incorporates a richer ontology of image junction interpretations than previously has been entertained, holds places for an extensible set of weak constraints or biases associated with figural geometry, and accepts input from additional sources of information such as depth-from-stereo. On the algorithmic level, we employ a token-based representation that is parsimonious and efficient in the declaration of equivalence classes of image events. The formulation permits information from spatially localized cues to be used purely locally as well as to propagate globally. Finally, the solution algorithm contains accessible "hooks" for interaction with top-down or other modules to influence and explore viable perceptual alternatives.

2 Computational Theory for Occluding Opaque Surfaces

A computational theory for the interpretation of images from the domain of constant intensity opaque surfaces under occlusion can be decomposed roughly into two realms, as identified by Malik and Shi[11]. *Ecological Optics* is about what *can* happen in the mapping from the physical world to images, while *Ecological Statistics* is about what *tends* to happen. These are discussed below in turn. Falling outside the scope of this paper, but fully subject to extensions, are theoretical consideration of painted or shaded surfaces; thin-lines; moving surfaces; transparent surfaces; and lighting effects such as shadows.

2.1 Ecological Optics: Junction Label Catalog

With regard to modeling the physical domain of interest, and images resulting from it, Williams focused on the depth ordering of complete surfaces at different places in the image as they overlap one another. Here we are less concerned with tracing the depth trajectories of contours over long distances, and more with articulating and disambiguating among local image cues. In particular, we introduce in Figure 3 a catalog of possible physical interpretations of image events occurring along contrast edges as a result of local surface shape and occlusion. This catalog enumerates interpretation labels for three types of image event, the *boundary contour*, *T-junction*, and *L-junction*.

The most interesting distinctions among interpretation labels arise from *generic* versus *nongeneric*, or "accidental," events. This distinction has been explored at length [18, 4], and its rigorous justification requires careful analysis of a given visual world of interest and an organism's place in it. For the present purposes, we employ an informal conceptualization and define an event as *generic* when the number of parameters required to specify its specific occurrence is fully the number of parameters required to specify *any* event of that class. For example, as shown in Figure 4 in general four parameters are required to specify the relative pose of two line segments in scale-space (Δx, Δy, $\Delta\theta$, $\Delta scale$). Any class of configurations of line segment pairs that could be specified with one fewer parameter, e.g. colinear, same size, parallel, would be regarded as having one degree of *nongenericity*. Parallel *and* same size would be nongeneric of degree two, and so on. The interpretation label ontology of Figure 3 considers events of qualitative nongenericity degree 0 and 1 only, arising from "accidental" alignment of contour edges or "accidental" congruity in the colors of distinct surfaces, as discussed below. Departing from previous expositions, however, we regard it as necessary to go beyond purely qualitative characterizations

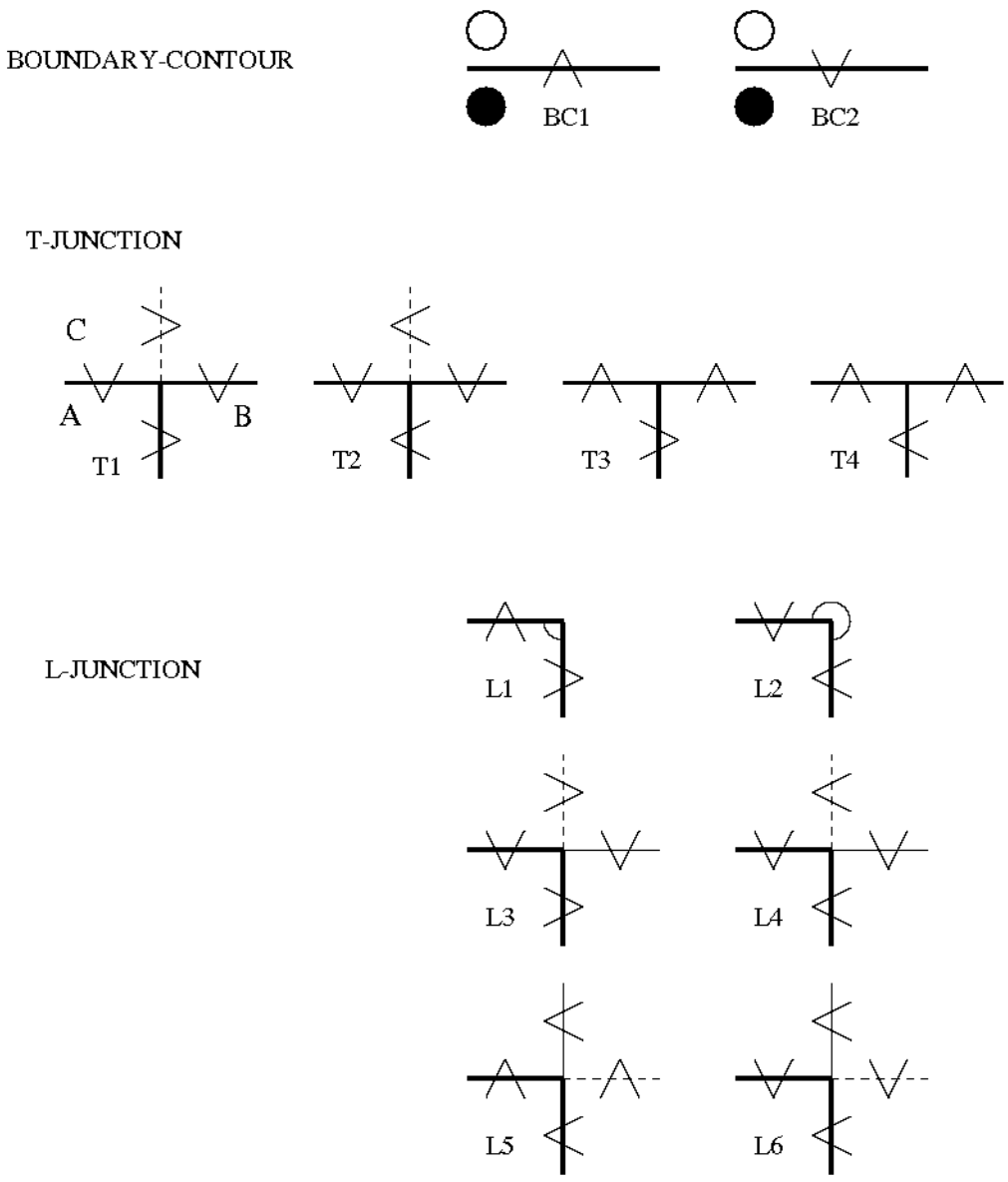


Figure 3: Catalog of interpretation labels for BOUNDARY-CONTOURS, T-JUNCTIONS, and L-JUNCTIONS. Heavy lines denote contrast edges, solid thin lines denote modal completion contours, dashed lines denote occluded contours. Arrows indicate direction of surface overlap: tip of arrow indicates occlud*ed* surface.

of genericity and accidentalness and consider graded penalties for different severities of accidents. The quantitative consequences of accidental alignments and the like are developed in Section 2.2.

The most elemental relation between imaged surfaces is a boundary contour created by the occlusion of one surface by another, the direction of which is depicted by a wedge arrow. A boundary contour thus takes one of two label values indicating which surface is in front.

We assume in the generic condition that different surfaces will have different lightnesses. Generically, therefore, the occlusion of two surfaces A and B by a third, C, creates a visible T-junction in which the stem and both halves of the bar appear as contrast edges. The generic T-junction is of two types, labels T1 and T2, depending upon whether A or B is in front of the other. The depiction of these junctions includes a dotted line indicating the presence of a contour boundary occluded by the nearest surface, Surface C. However, two additional causes for observed T-junctions exist, T3 and T4, for which Surface C is behind. These are nongeneric because they involve the coincidental alignment of the boundary contours of the two distinct surfaces, A and B.

A surface that overlaps another of the same lightness generates an invisible contour boundary known as a *modal completion* edge. It is important to distinguish the common usage of this term as referring to phenomenal appearances generated by certain stimuli, from our definition of a modal completion edge as the formal assertion of surface overlap sans contrast edge—whether it is perceived in some fashion or not. The present work makes no attempt to predict the vividness with which illusory contours will be experienced by human observers, nor which side of a modal completion edge will appear lighter or darker than the other, nor to what degree. A modal completion edge is regarded as a nongeneric, or accidental, event because in general

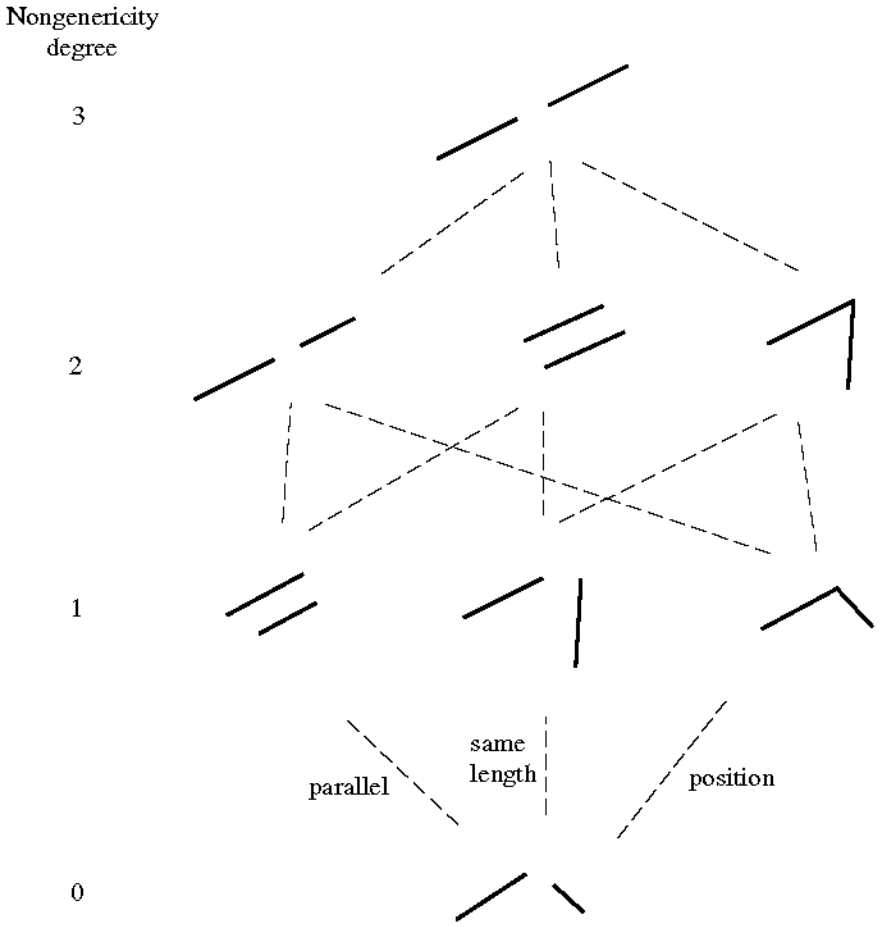


Figure 4: Two line-segment illustration of *nongenericity*. In general two line segments are related by four degrees of freedom. Increasing degrees of nongenericity are introduced in stages by constraining parameters specifying relative location, orientation, and size.

two parameters would be required to specify the colors of different surfaces instead of the single parameter characterizing a shared value.

Referring to Figure 3, an image L-junction arises from one of six causes. L1 and L2 are generic, occurring when a contour boundary undergoes an orientation discontinuity. In its graphic depiction, an arc helps to distinguish a convex corner from a concave partial hole, in addition to the drawn arrows. The remaining four L junction labels are degenerate T-junctions arising from the nongeneric event of the occluding surface matching the lightness of one of the occluded surfaces.

A fully comprehensive accounting of the ecological optics of the Colorforms domain would include additional labels for image events of qualitative nongenericity degree 2 and higher. See Figure 5. Although images do occur requiring these labels, they are relatively rare and are unnecessary to the development of the conceptual and algorithmic machinery sought by this paper. With regard to X-junctions in particular, these can indeed occur through certain configurations of opaque surfaces, but we defer their consideration for future work because X-junctions become really interesting only in the context of a richer physical imaging domain including surface "atmospheric" effects of transparency, smoke, fog, and shadow [1, 2].

2.2 Ecological Statistics: Figural Biases

Once the mapping between an underlying physical domain model and a catalog of image events is spelled out, it remains to further refine prior knowledge of the visual world needed to constrain interpretations of potentially ambiguous image data. We need to articulate in greater detail a topography, over the space of things that *could* occur, of what is more *likely* to occur.

The delineation of labels associated with generic versus nongeneric or accidental conditions does part of the job: generic interpretations will be preferred over non-

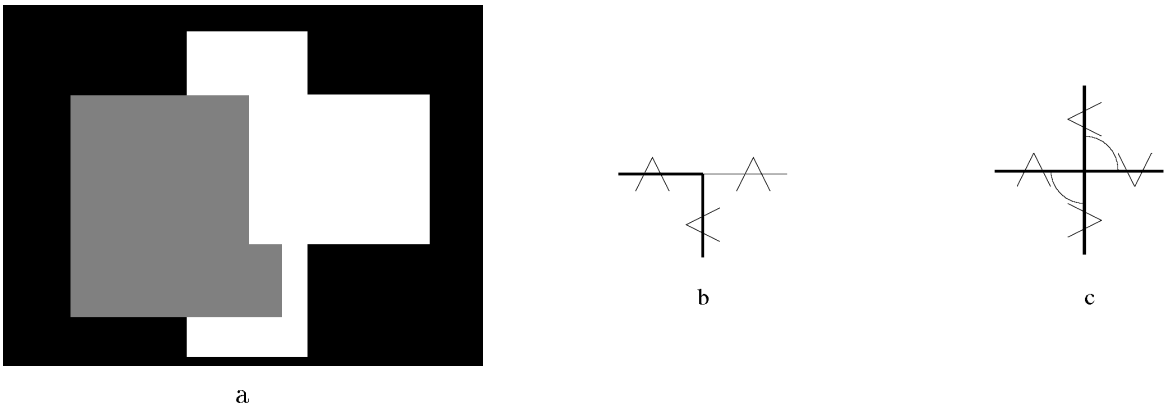


Figure 5: a. The interpretation of a white square overlapping and aligning the grey square requires a junction label of nongenericity degree 2, shown in b., due to accidental edge alignment *and* surface color match. c. An X-junction label, nongenericity degree 2 due to double accidental edge alignment. The algorithm implemented limits itself to junctions of nongenericity degree 0 and 1.

generic ones. But what is the relative genericity of a surface color match on the one hand, and a contour alignment, on the other? Or what is the tradeoff between a very good alignment and two so-so ones? We find it necessary to depart from any purely qualitative accounting, and begin to entertain quantitative measures for relative preferences of image events' interpretations with respect to one another. This strategy is in keeping with Williams' establishment of a quantitative objective function expressing figural biases according to empirically observed perceptual phenomena.

More formally, we express penalties for junctions adopting certain labels, in the form of an *energy cost*. The terms contributing to energy cost are all mathematical expressions engineered to take particular functional forms justified on the basis of commonsense evaluations of sample example configurations and the human visual system's behavior on simple stimuli. Our choices for these expressions are presented in the Appendix, but they are subject to modification, refinement, and testing against either human psychophysical data or arguments from first principles deriving from

statistical analysis of the visual world. What is most important for the present purposes is to get their qualitative behavior right, more or less. To date we include the following figural biases, illustrated in Figure 6:

- **Generic Positioning.** An energy cost E_{aa} (*aa :: accidental alignment*) is imposed whenever two edges align with one another but their associated junction labels interpret them as arising from unrelated contours. E_{aa} reaches a maximum for putatively unrelated edges that abut and align perfectly, and decreases with distance and misalignment.

- **Contour Smoothness.** An energy cost E_{cs} (*cs :: contour smoothness*) is imposed whenever two distinct contours are hypothesized by their associated junction labels to belong to a common contour, blocked from view by occlusion. The energy cost decreases with nearness and smooth continuation of the two contours, and increases as the gap between them increases or their hypothesized invisible join becomes more contorted.

- **Generic Surface Color.** An energy cost E_{mc} (*mc :: modal completion*) is incurred for the assertion of junction labels proposing occlusion by a surface that happens to be the same color as the occluded surface. As with contour smoothness, this cost is at a minimum when the hypothesized modal completion edge is very short and smooth, and increases with its length and contortion.

- **Figural Convexity.** An energy cost E_{fc} (*fc :: figural convexity*) is incurred for hypothesizing locally concave occluding surfaces. Curving boundary contours assigned overlap labels corresponding to concave occlusion boundaries, or partial holes, are assigned cost according to the angular extent. Likewise, concave corners, corresponding to Type L2 L-junctions, incur energy cost according to their internal angle.

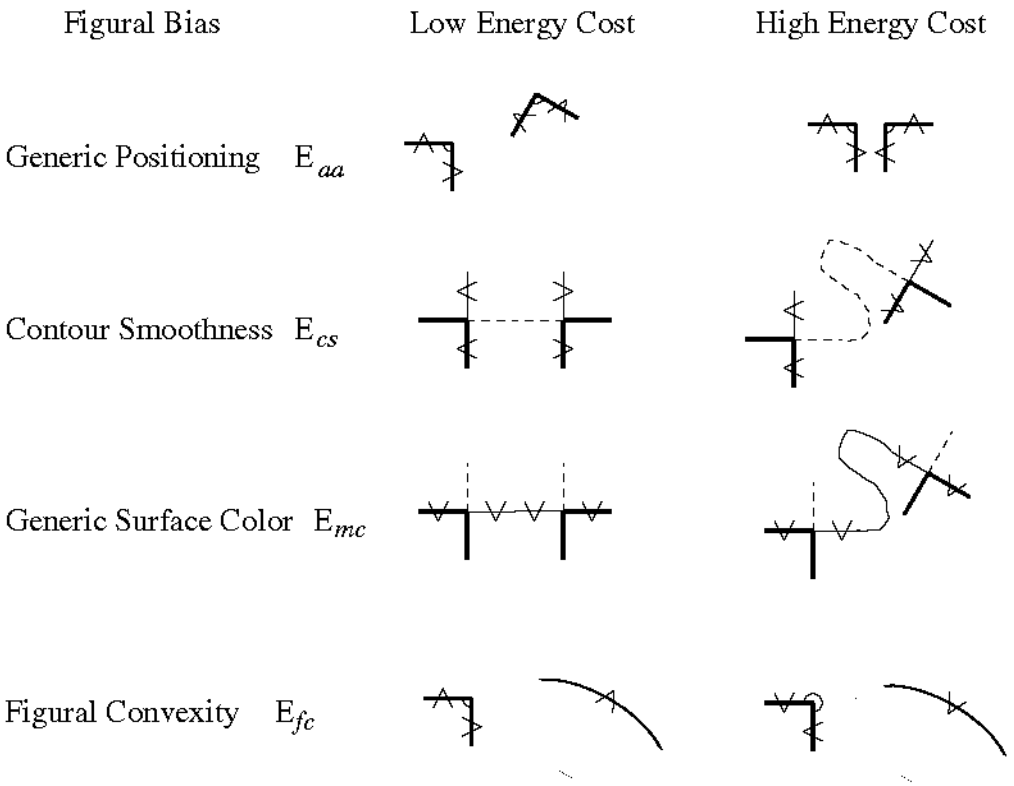


Figure 6: Schematic illustration of quantitative figural biases. See text for explanation and the Appendix for mathematical expressions developed to reflect these biases.

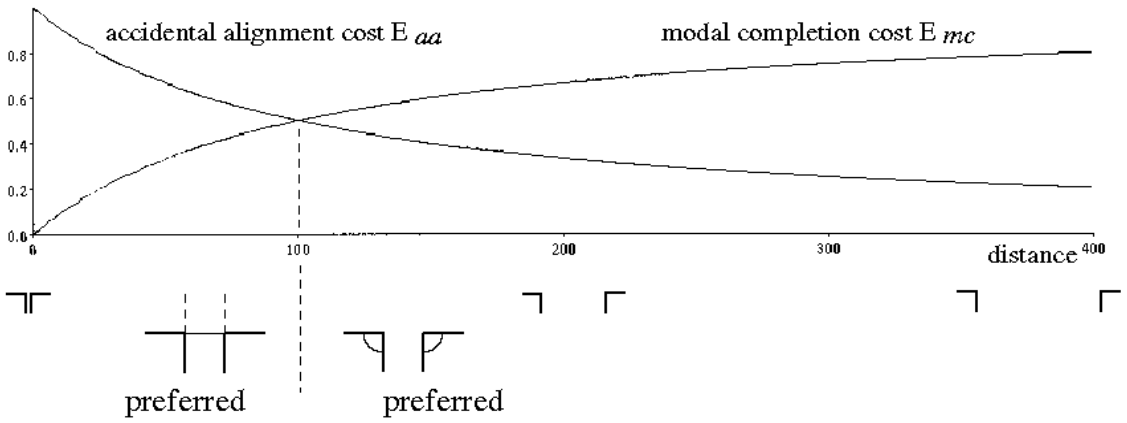


Figure 7: A tradeoff in E_{aa} (*accidental alignment*) versus E_{mc} (*modal completion*) energy costs occurs as a pair of aligning L-JUNCTIONS increase in distance from one another, reflecting a transition in preference from a modal completion interpretation to an independent object interpretation.

Embedded in these figural biases are the determinants of the tradeoffs between accidental alignment, smooth continuation, and modal completion interpretations under ambiguous image evidence. For example Figure 7 shows, as two L junctions containing aligning edges are brought nearer, where we have effectively chosen to place the tradeoff between an accidental alignment interpretation and a modal completion interpretation.

Of course, it would be easy to augment these figural biases with additional ones motivated by psychophysical or other sources of evidence. For example, Williams [26] included a bias for perceiving nearby parallel lines as figure. We are fully open to augmenting the figural biases presented above with others, and welcome the notion that these need not be fixed in form, but are subject to adjustment dynamically in the course of perception by top-down mechanisms, global contextual cues, or other visual modules.

3 Representational and Algorithmic Proposal

In this framework the core problem of perceptual organization of occluding contours becomes one of assigning interpretation labels to perceptually significant events in image data. Our proposal is to form a *junction graph* whose nodes are symbolic tokens denoting BOUNDARY-CONTOUR, T-JUNCTION, and L-JUNCTION events, and whose links represent coincidence and geometric configurational relations among these events, where these links provide and propagate constraint on nodes' labels. Search over the space of junction label assignments is conducted by allowing local preferences to propagate around the graph as labeling decisions are made gradually, using a continuation method. The overall approach of searching a hypothesis space governed by weakly interacting constraints through parallel iterative local propagation is descended from relaxation labeling [20, 15, 3, 14].

3.1 Energy Cost Objective Function

In general any global interpretation must obey constraints of local overlap consistency, that is, physically feasible interpretations obey the condition that the overlap directions occurring at a junction match those of the boundary contours forming the junction. Note however by the example in Figure 8 that this need not be a strict requirement; in fact humans perceptually can entertain interpretations that because of strong local figure/ground pressures must be globally inconsistent with regard to occlusion direction along visible contours. Conveniently for trading off overlap consistency with other evidence, the local overlap consistency constraint can be expressed in the same energy cost terms as figural constraints:

- **Neighbor Consistency.** An energy cost E_{nc} (*nc :: neighbor consistency*) expresses a penalty incurred for every instance that a junction interpretation

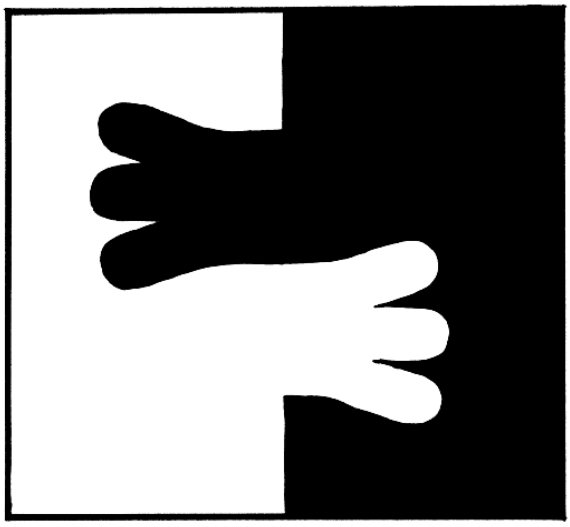


Figure 8: Strong local figure/ground pressures can prohibit globally consistent figure/ground assignments for all contour edges.

label conflicts with that of its constituent boundary contours.

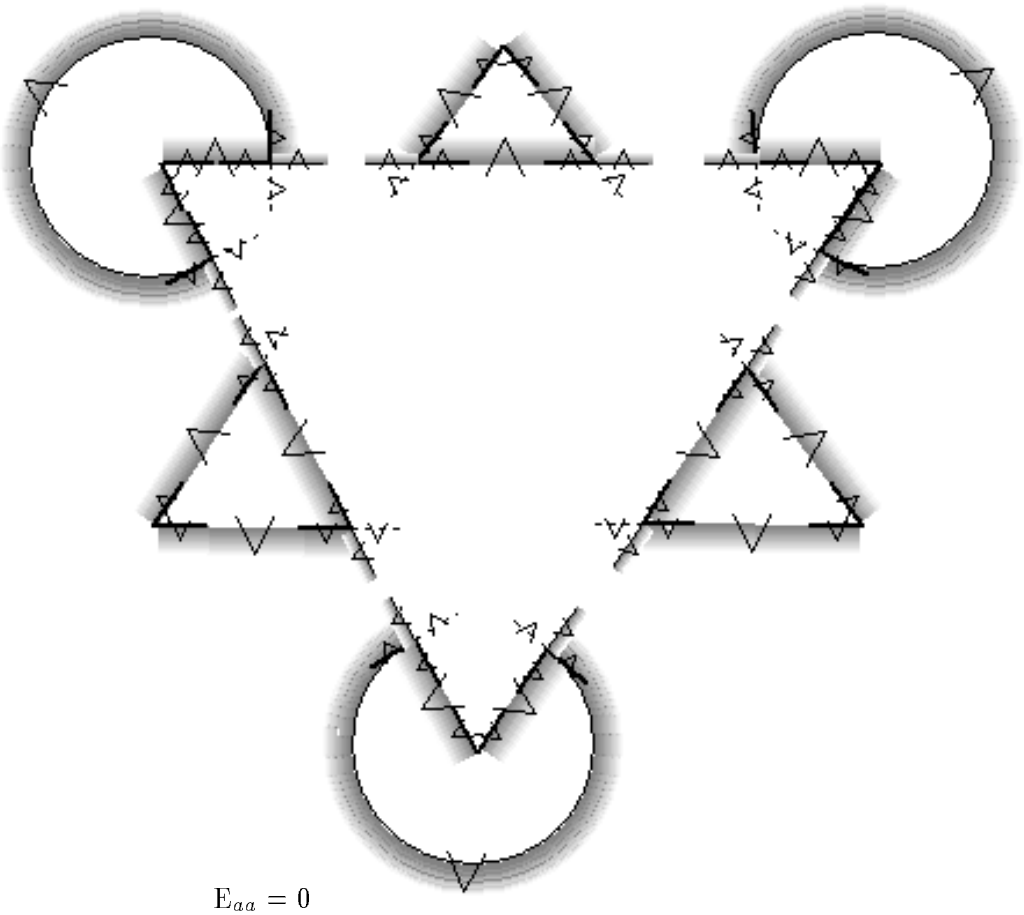
Any assignment of interpretation labels to a figure gives rise to a global interpretation energy cost simply by summing the energy costs of all boundary-contours and junctions. Figure 9 illustrates optimal and suboptimal labelings of the Kanizsa triangle. Our figural biases are chosen so that optimal energy costs will correspond with perceptual interpretations preferred by humans. Note for example in Figure 9a that modal completion contours are asserted to enclose the central white triangle, while amodal continuation contours complete the occluded black triangle and the occluded black circles. But Figure 9b shows that another interpretation that is fairly easy for humans to see—the circles as holes revealing a black background—pays only a small energy penalty for figural nonconvexity, while Figure 9c shows a strongly nonpreferred interpretation—isolated objects with no occlusion—which pays a very high penalty for accidental edge alignments. In all cases presented in the paper, unless otherwise

noted the optimal labeling is attained by the algorithm of Section 3.4.

3.2 Link Formation by Token Grouping

Input data consists of chain-coded contours such as found by edge detection and curve tracing processes, and annotated with the colors of surfaces on each side. Contours are broken at corners and merged at points of smooth alignment, giving rise to BOUNDARY-CONTOUR tokens as shown in Figure 11a. Locations, orientations, and curvatures of contour ends are estimated by fitting circular arcs at each end. Cliques of two and three nearby ends forming L- and T- junctions are found by clustering and performing simple geometric tests. L-JUNCTION and T-JUNCTION tokens are created accordingly. Simple techniques work for computer-generated graphic data and video frames of physical construction paper scenes, but obviously would need much refinement for photographic imagery.

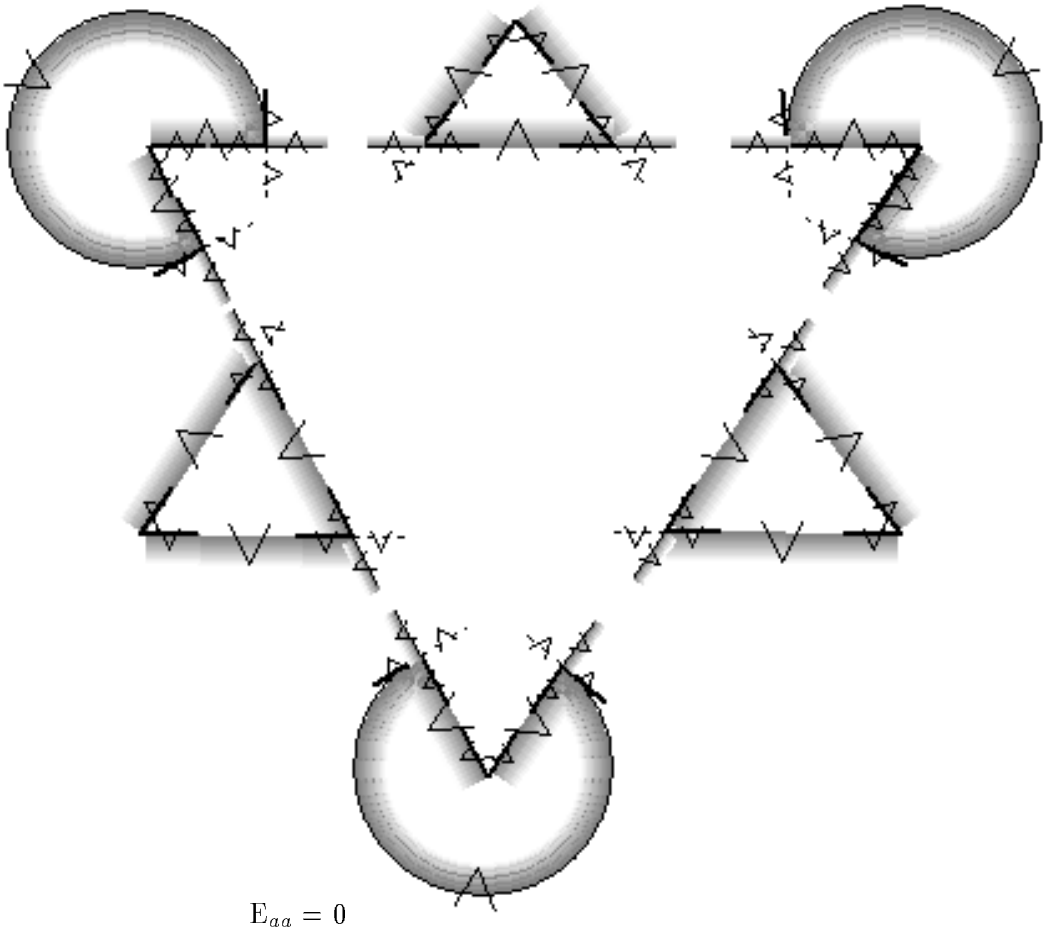
The junction graph contains two kinds of links. First, *coincidence links* denote associations between L- and T-JUNCTION tokens and the BOUNDARY-CONTOUR tokens contributing to their formation. These represent the visible structure of the contrast edges in the scene. Second, *alignment links* declare pairs of contour ends that are preferably near to and align with one another across pairs of L- or T- junctions. See Figures 10 and 11b. Search for aligning contour pairs is conducted by directing an expanding beam from each leg of every L-JUNCTION token and the stems of T-JUNCTION tokens, testing junction tokens it encounters for alignment and compatibility of surface colors on each side. We offer in the Appendix a heuristic mathematical expression assessing degree of alignment which reflects an estimate of the likelihood that the two contour ends belong to the same underlying contour which may have been rendered invisible either by occlusion or color match. Note that although our measure roughly



$E_{aa} = 0$
$E_{cs} = 2.9$
$E_{mc} = 3.8$
$E_{fc} = 0$
$E_{nc} = 0$

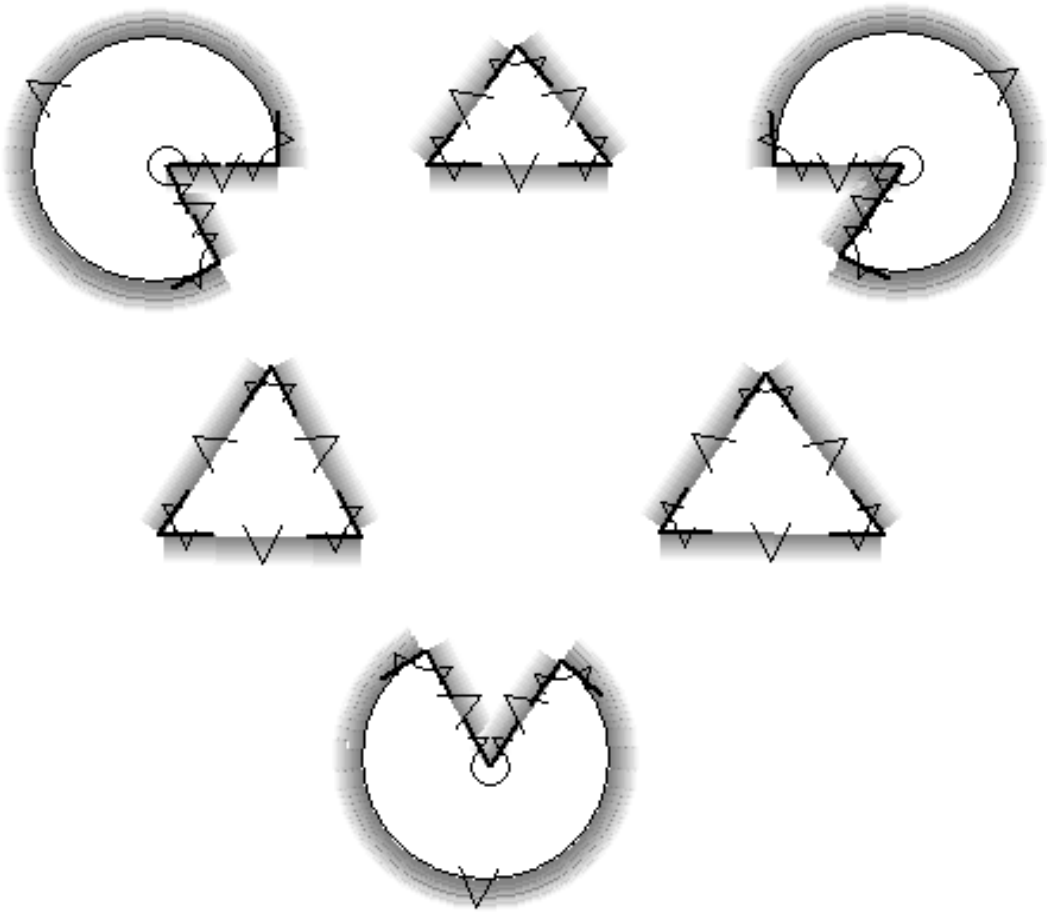
a

Figure 9: Three among many possible interpretation labelings of the Kanizsa Triangle, and their associated energy costs. All labelings are stable at high inverse-temperature, but a. (the global optimum) is converged to by the algorithm annealing from low inverse-temperature. Shading resembling shadowing is added to enhance visualization of program output.



$E_{aa} = 0$
$E_{cs} = 2.9$
$E_{mc} = 3.8$
$E_{fc} = .8$
$E_{nc} = 0$

b



$E_{aa} = 13.1$
$E_{cs} = 0$
$E_{mc} = 0$
$E_{fc} = .3$
$E_{nc} = 0$

c

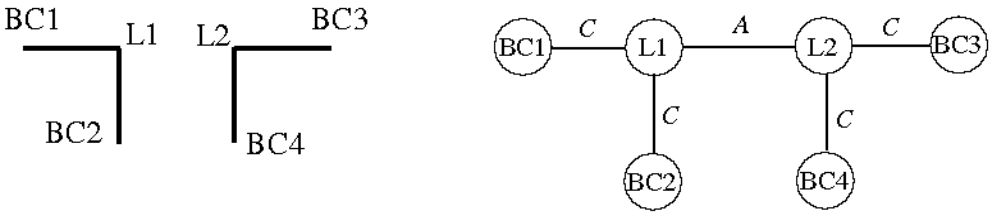
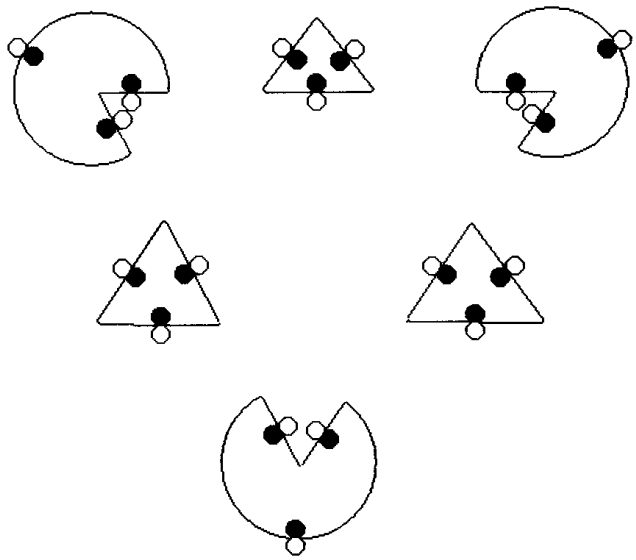


Figure 10: Four BOUNDARY-CONTOURS forming two L-JUNCTIONS, and their corresponding junction-graph containing coincidence links (C) and an alignment link (A).

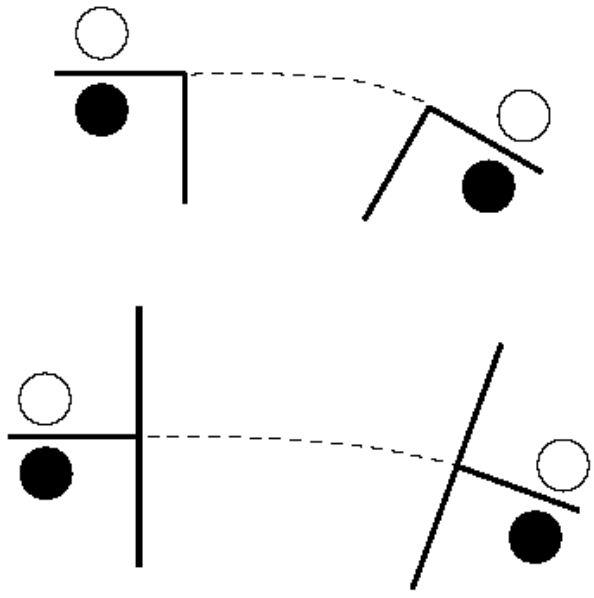
includes terms for both distance and bending energy, no attempt is made to infer the detailed shape of the missing contour. In cases where more than one good alignment match is found, a heuristic algorithm is used to prune the set of alignment links to one per L-JUNCTION leg and T-JUNCTION stem. Alignment links found for the Kanizsa triangle are shown in Figure 11c. Although this works for simple illusory contour figures this step begs for a more sophisticated approach as mentioned further in the Discussion section.

3.3 Subtleties of Evidence Propagation

At the outset of the computation before any evidence has been considered, no junction has any basis for asserting any one interpretation label over others. The catalog of interpretations available to a junction (or boundary-contour) j forms an interpretation belief vector b_j of elements $b_{j,l}$, where l refers to the lth interpretation label, $0 \leq l \leq L$. L is the number of interpretations available to that junction's type (i.e. $L = 2$ for a

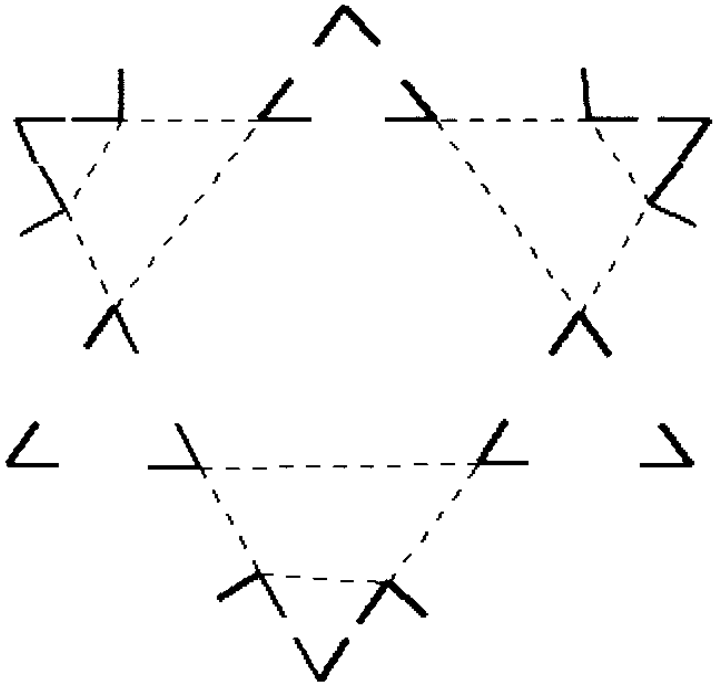


a



b

Figure 11: a. BOUNDARY-CONTOUR tokens for the Kanizsa Triangle figure. b. Alignment links are created between pairs of L-JUNCTION legs and pairs of T-JUNCTION stems that are sufficiently near and aligned with one another, and whose bounding surface colors match. c. L-JUNCTIONS(solid lines) and alignment links (dashed lines) found for the Kanizsa Triangle figure.



c

BOUNDARY-CONTOUR, $L = 4$ for a T-JUNCTION, $L = 6$ for an L-JUNCTION).

One way to imbue the representation with the expressive power to declare that multiple interpretations remain possible at a junction would be to allow multiple beliefs $b_{j,l}$ to take the value 1, holding open the possibility that the lth interpretation is true for multiple choices of l. A final global interpretation of a scene would occur after junction interpretations have been eliminated by switching their label entries to 0, leaving only a single 1 entry in each junction's belief vector. In this scheme, propagation of evidence around the junction graph would consist of propagating vetos of junction interpretations based on consistency across coincidence and alignment links, in the manner of discrete junction labeling. Such a strategy would rely heavily on combinatoric search and backtracking. Using a very different labeling ontology, Williams instead turned to a global integer linear constrained optimization formulation. We believe that this approach violates the Principle of Graceful Degradation [12] because it places severe restrictions on the consistency of the input data and because it foregoes purely local use of local evidence, as evidenced by its prohibition of globally inconsistent but perceptually phenomenal interpretations such as in Figure 8.

A "softer" representation we propose admits continuous valued beliefs, $0 \leq b_{j,l} \leq 1$. Because each visible junction is assumed to arise from one and only one physical cause, we impose the further constraint, $\sum_l b_{j,l} = 1$, giving the belief vector resemblance to a probability distribution over interpretation states. The vector, $b_j = \{1/L, 1/L, ...1/L\}$ represents abstinence of preference for any interpretation.

Propagation of evidence via the Neighbor Consistency constraint consists of accumulating energy cost for every available interpretation label of each junction, due to the interpretation vectors of its link neighbors in the junction graph. Figure 12 includes illustration of the consistency relation between the left leg of an L-JUNCTION ($j = 2$) and the North end of a BOUNDARY-CONTOUR ($j = 1$). The contribution of

energy cost E_{nc} by the BOUNDARY-CONTOUR due to the L-JUNCTION, and vice versa, are generated according to the following *propagation consistency matrices*, respectively (refer to Figure 3):

$$\begin{bmatrix} E_{1,1} \\ E_{1,2} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} b'_{2,1} \\ b'_{2,2} \\ b'_{2,3} \\ b'_{2,4} \\ b'_{2,5} \\ b'_{2,6} \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} E_{2,1} \\ E_{2,2} \\ E_{2,3} \\ E_{2,4} \\ E_{2,5} \\ E_{2,6} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} b'_{1,1} \\ b'_{1,2} \end{bmatrix} \tag{2}$$

A 1 entry reflects label incompatibility. For example, belief that an L-JUNCTION should be interpreted as a convex orientation discontinuity, $b'_{2,1} \neq 0$, leads to an energy cost penalty for interpretation label 2 of the BOUNDARY-CONTOUR, $E_{1,2}$, but none for interpretation 1. Similar overlap compatibility matrices can be generated for overlap consistency links between BOUNDARY-CONTOURS and T-JUNCTIONS, and for alignment links between pairs of L-JUNCTIONS and the stems of pairs of T-JUNCTIONS.

There is however a subtle consideration that must be addressed concerning the elimination of bias under ignorance. Suppose that for the label belief terms b' in expression (1) we use the pure belief vector values $b_{j,l}$. Then in the case that junction 2 is neutral in its interpretation belief, i.e. $b_{2,l} = \frac{1}{6}$, unequal amounts of energy cost will be propagated through the matrix leading to a biased contribution to BOUNDARY-CONTOUR 1's energy cost vector,

$$\begin{bmatrix} \frac{4}{6} \\ \\ \frac{2}{6} \end{bmatrix}.$$

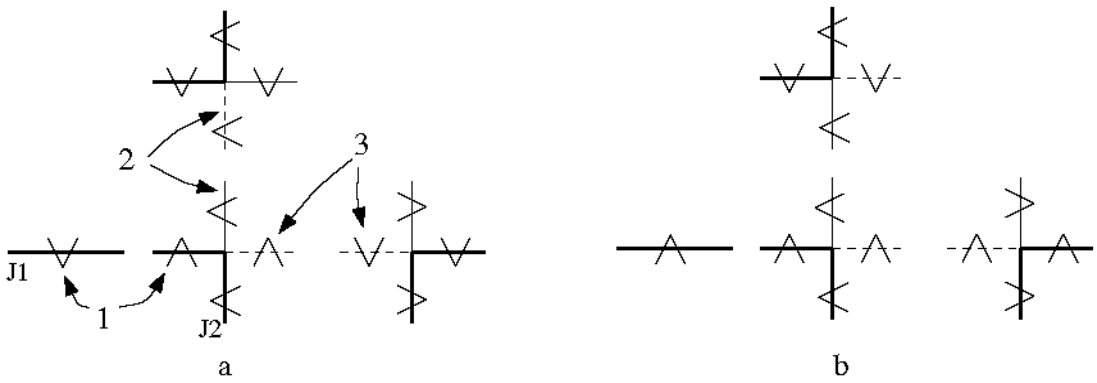


Figure 12: a. An incompatible assignment of junction labels due to: (1) incompatibility between an L-JUNCTION and one of its associated BOUNDARY-CONTOURS; (2) incompatibility between modal completion and amodal contour continuation interpretations; (3) incompatibility between overlap directions across an amodal continuation link. b. A compatible assignment of junction labels.

This bias can be eliminated by interposing a transformation in each belief vector before multiplying it by the propagation consistency matrices. The transformation is a shift to a *zero-based* representation for belief [21]. A zero-based representation maps the interval [0, 1] to the interval [-1, 1] such that a uniform probability distribution maps to zero. For this, we use the formula[2],

$$b' = 2b^{\log_L 2} - 1. \tag{3}$$

3.4 Optimization by Deterministic Annealing

For any given states of belief vectors held by its link neighbors, a node in the junction graph can compute an energy cost distribution over its catalog of interpretation labels by summing energy costs associated with figural biases and with overlap consistency

[2]I thank Josh Tenenbaum for suggesting this functional form.

constraints as determined by alignment and coincidence links. In order to give local evidence the opportunity to propagate around the junction graph, we desire that belief vectors not immediately choose lowest energy cost labels in winner-take-all fashion, but instead iteratively gravitate from neutrality toward a single interpretation. A mechanism for accomplishing this is provided by the technique of *deterministic annealing*[5, 19]. An *inverse-temperature* parameter β is used to govern the mapping between energy cost and belief distribution using the Softmax operator:

$$b_{t+1,j,l} = \frac{e^{-\beta E_{t,j,l}}}{\sum_l e^{-\beta E_{t,j,l}}},$$

where t is an index of time or iteration number. Low inverse temperature spreads belief more evenly over all available states, while raising inverse temperature corresponds to "cooling" toward a winner-take-all state. All experiments reported in this paper were performed using a simple predetermined annealing schedule consisting of ten iterations at each of five temperatures, $\beta = 0.5, 1, 2, 3, 10$.

4 Results and Discussion

4.1 Representative Results

In addition to the Kanizsa triangle result of Figure 9a, Figures 13 and 14 present input images along with interpretations found by the algorithm for several representative situations. Figure 13a/b are another figure from Kanizsa's book showing interweaving due to the greater cost of modal completion contours versus amodal continuation contours, per unit contour length. Figure 13c/d demonstrates a scene whose preferred interpretation includes a nongeneric edge alignment at a T-junction. Figure 14 illustrates how the algorithm is affected by degradation in its feature input. The algorithm scales linearly in computational cost with amount of feature input, but not surprisingly requires correct junction features in order to infer contour occlusions

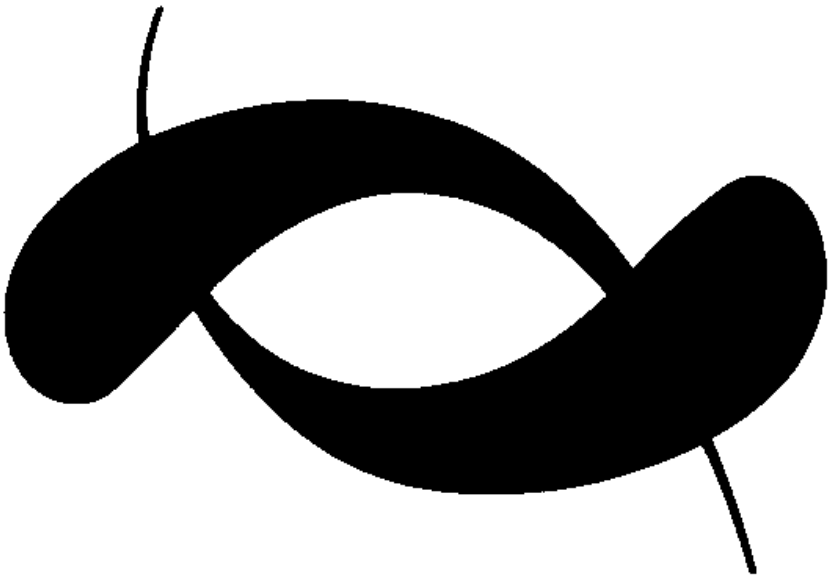
correctly. Local interpretations are however robust with respect to disruptive feature input elsewhere in the scene.

Figure 15 shows that the algorithm is amenable to accepting augmented evidence such as stereo cues. Stereo evidence of relative surface depth at some L-junctions of the Kanizsa triangle was simulated simply by injecting energy cost for junction interpretations violating the stereo depth cues. Note that resulting weaving interpretation matches human perception of the stereo scene.

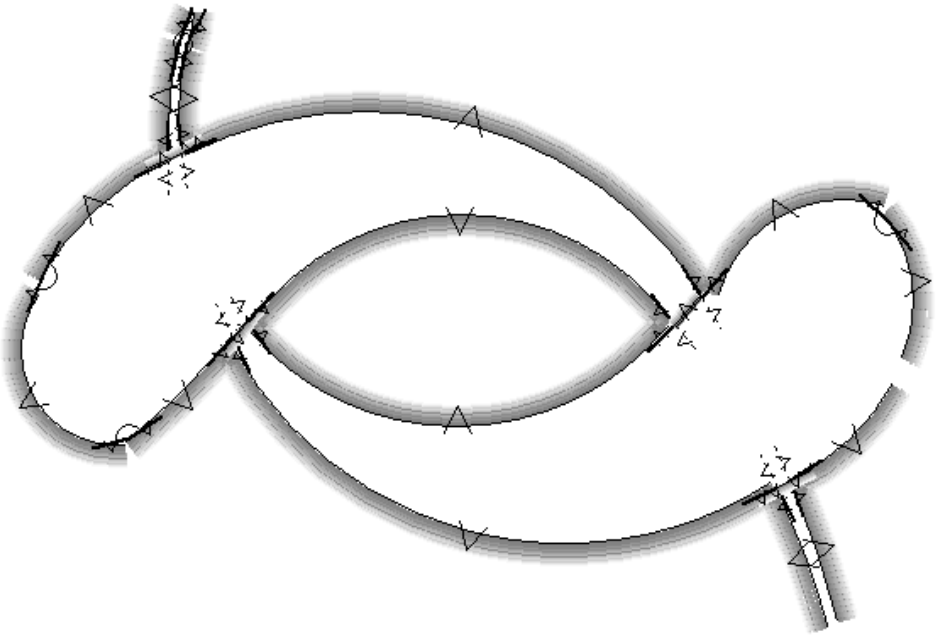
4.2 Algorithm Level Considerations

The junction catalog and figural bias formulation we have presented as computational theory for perceptual organization of occluding surfaces could in principle be deployed under a variety of algorithmic approaches. The choice of representations and computing strategies is in many ways more difficult than the computational theory itself because it involves subtle engineering judgements regarding underspecified requirements of the larger visual system within which this module will ultimately be embedded. Let us review some of the motivations behind the choices made by this work:

I. *Appropriate Granularity.* Following the Principle of Explicit Naming [12], we employ computational elements that treat as unit entities any sections of contour or surface region that behave equivalently. Accordingly, boundary contours, T- and L- junctions, and local surface patches are made explicit as tokens amenable to symbolic grouping, linking, labeling and propagation efficiently over large distances. This is in contrast to a dense field representation more amenable to inherently slow diffusion style processes [6]. It remains an open question how an extremely coarse symbolic token grouping and labeling

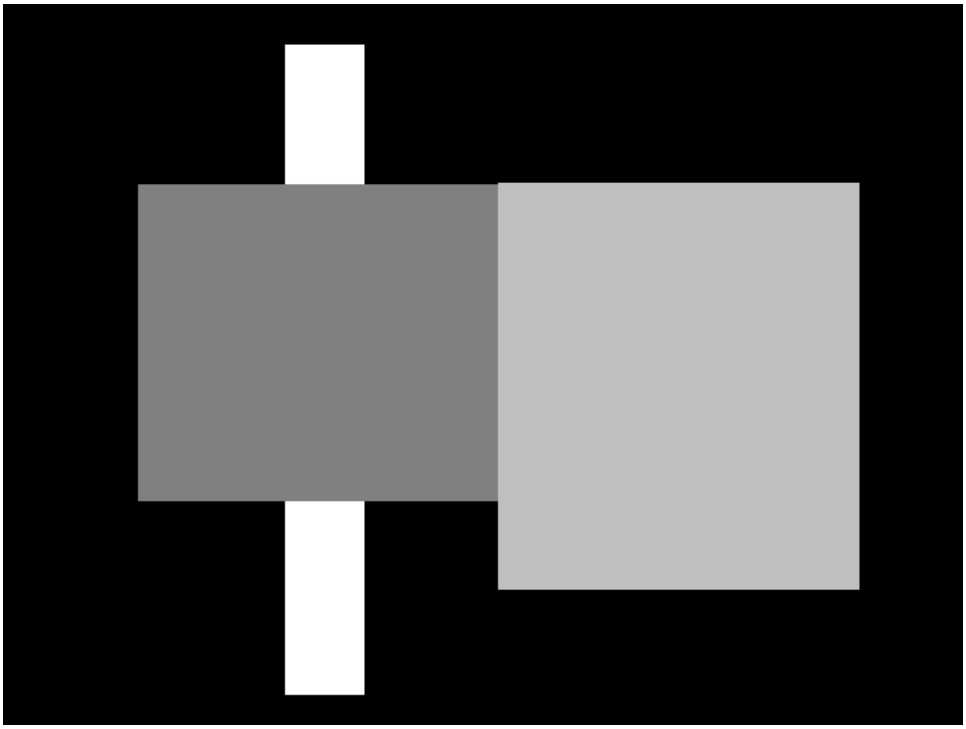


a

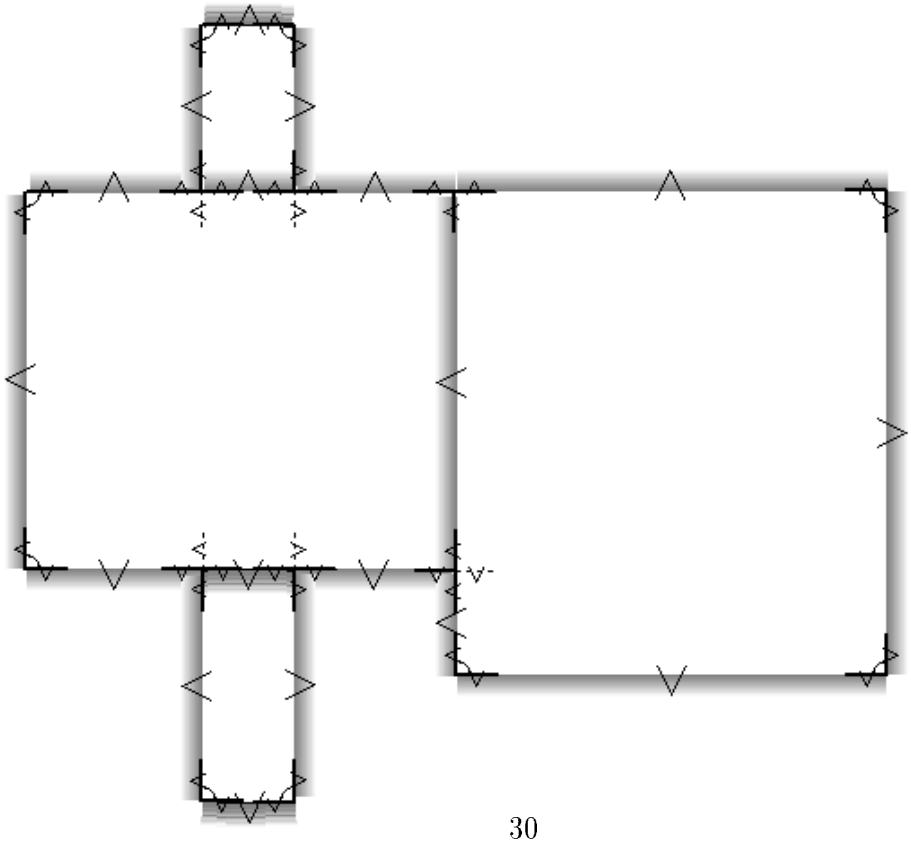


b

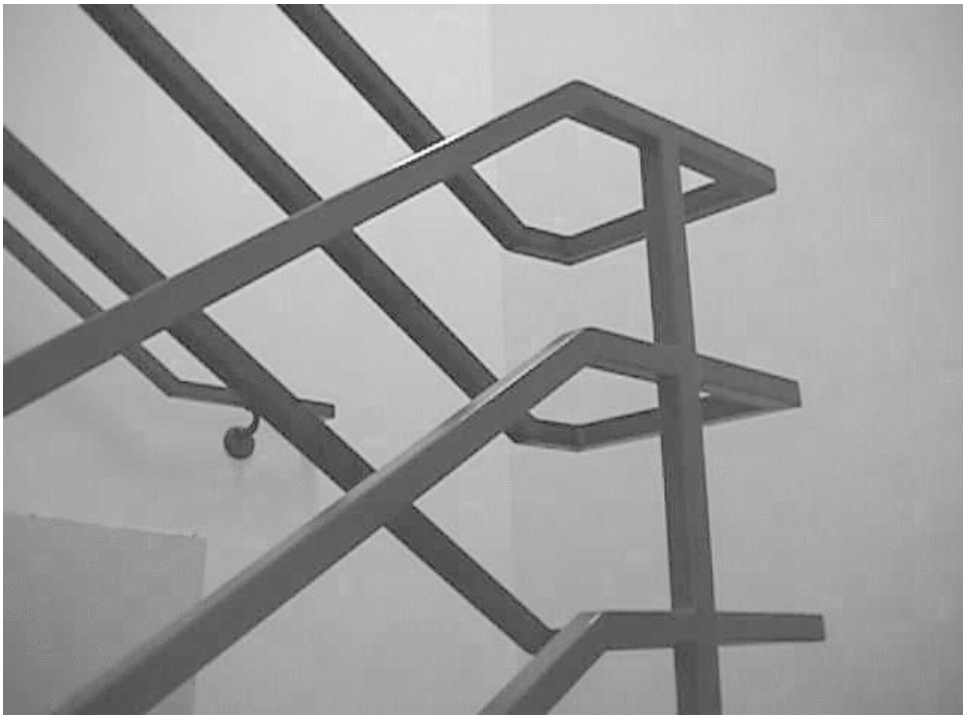
Figure 13: Results. a/b Note interweaving interpretation of Kanizsa Figure 2.13 [9]. c/d The preferred percept requires a nongeneric interpretation (interpretation label T4) for the upper right T-junction.



c

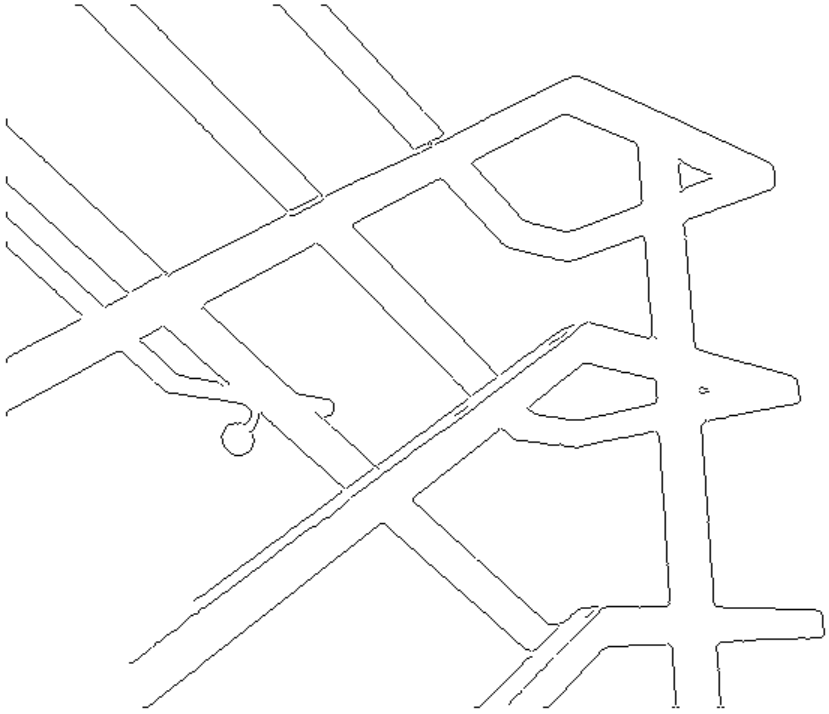


d

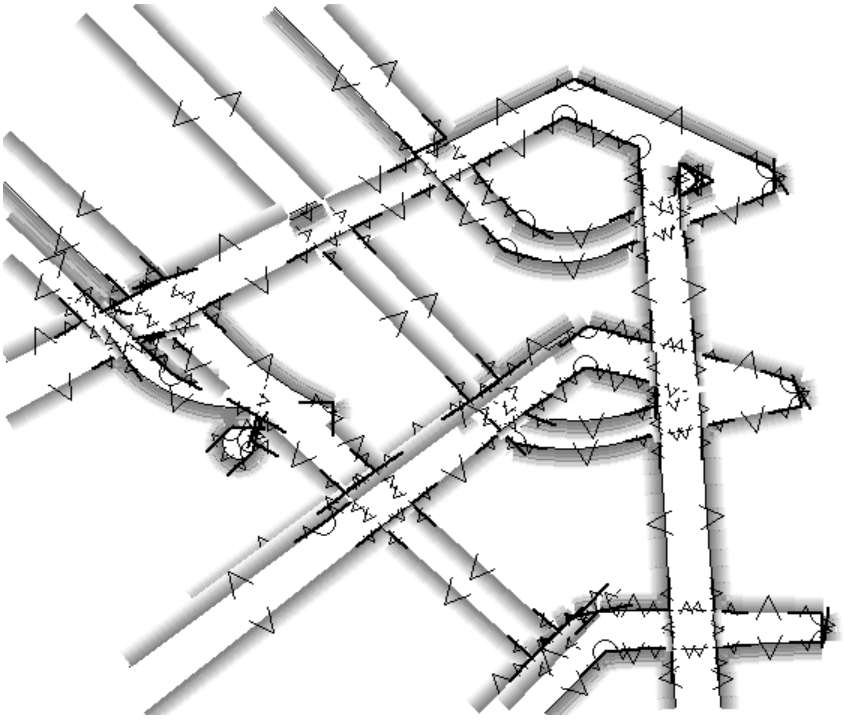


a

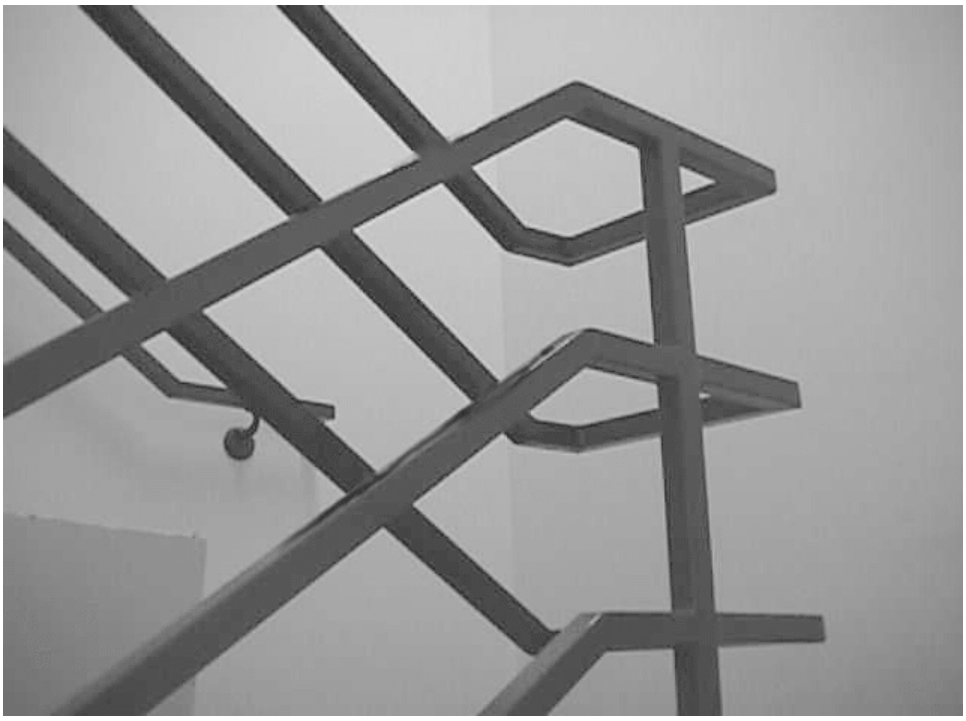
Figure 14: Even in a scene comprised of opaque uniformly colored objects, simple local edge detection and grouping does not produce clean T and L-junctions. Note how highlights on the railings (a) introduce edges not corresponding to object boundary contours (b). These disrupt the junction graph, leading to locally incorrect occlusion inferences (c). Artificially removing the highlight (d) allows these features to be found by simple methods, to the benefit of the resulting interpretation (e).



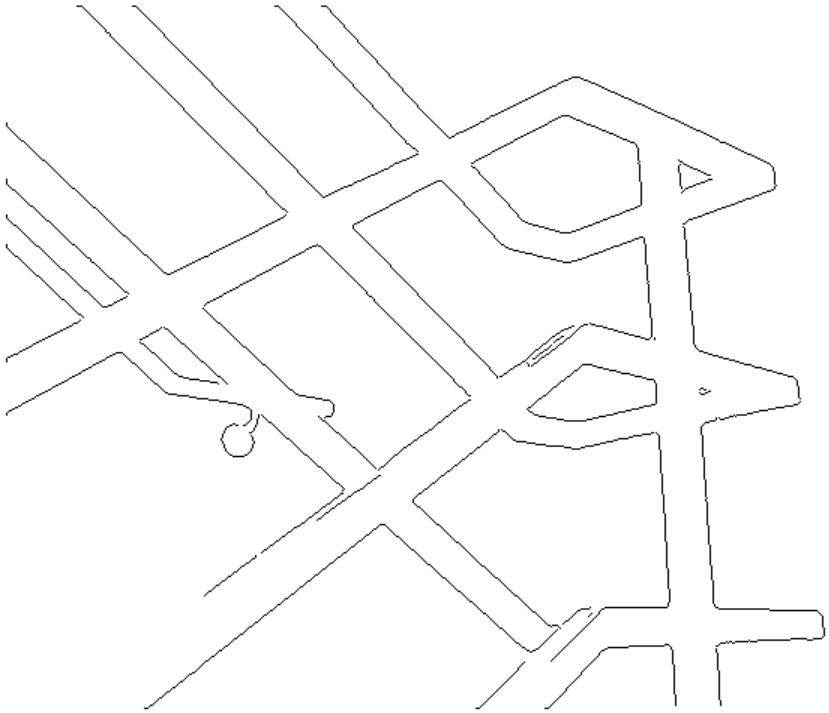
b



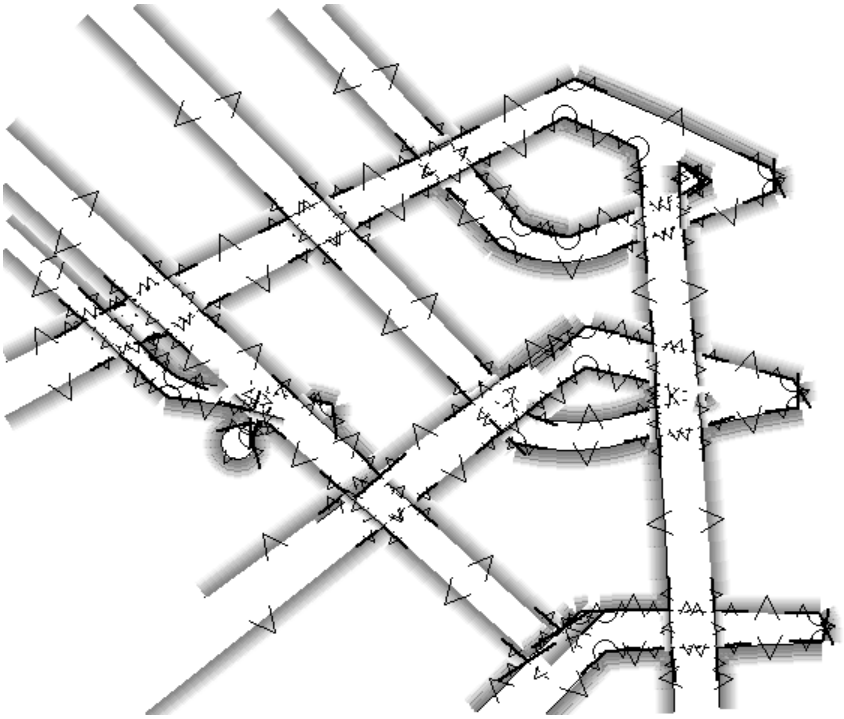
c



d



e

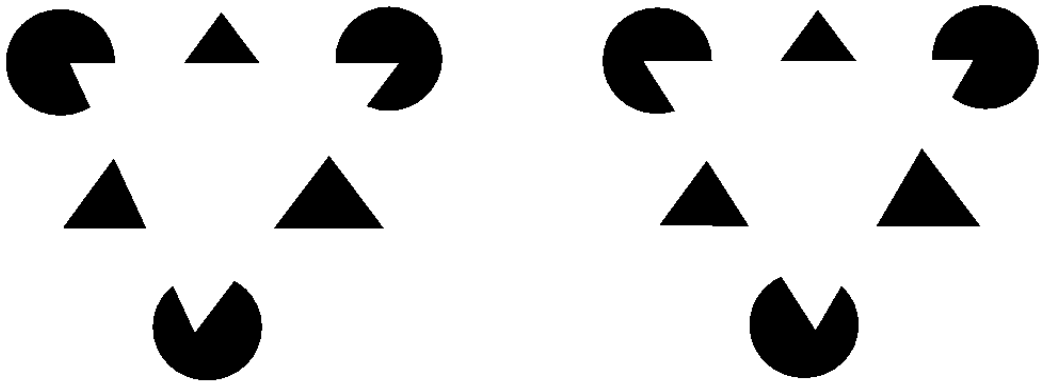


f

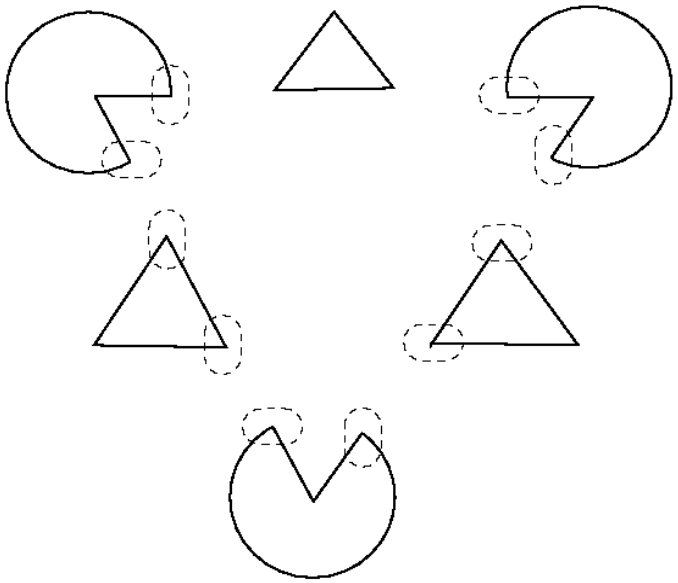
approach will scale to more complex scenes characterized by greater density of significant image events.

II *Preference for Local Evidence.* Demonstrations such as the Devil's Pitchfork, and many human observers' reported experiences of difficult displays such as random dot stereograms, suggest that the human visual system does not combine all available evidence into a single problem statement leading to a global all-at-once solution but instead uses information locally to construct local solutions, which in turn propagate constraints to neighboring regions.

III *Approprite Modularity.* It is well known that human perception of Colorforms displays allows multiple interpretations, and that these are in many cases *cognitively penetrable* [17], that is, influenced by conscious thought. More generally, engineering modularity argues that perceptual organization at this level be rel-

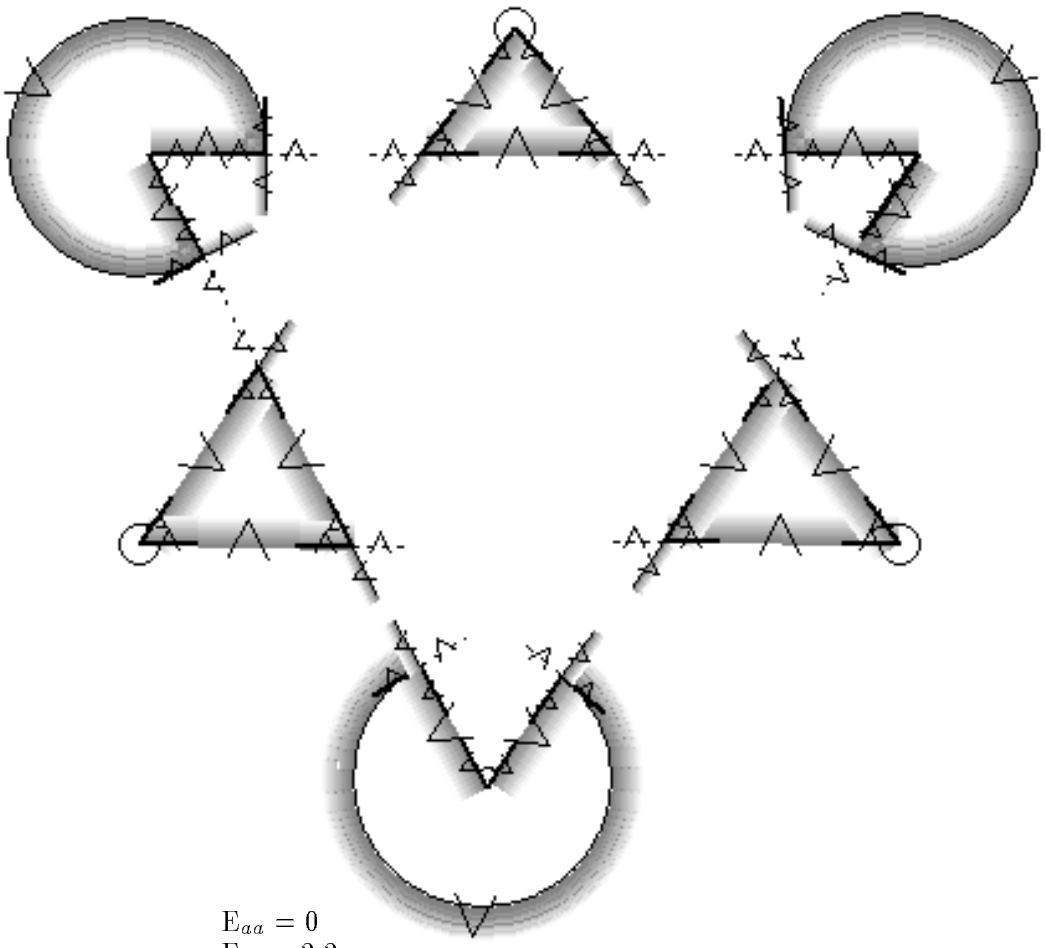


a



b

Figure 15: Stereo evidence influences surface overlap perception. a. Stereo disparities at pacmen leads to perception (under cross fusing) of the white triangle weaving behind a flat planar surface. b. Input from a stereo module was simulated by effectively constraining certain L-junction interpretation labels according to surface overlaps allowed by the local disparity: vertical ovals indicate L-JUNCTIONS constrained to be of type L3 or L4; horizontal ovals indicate L-JUNCTIONS constrained to be of type L5 or L6. c. Resulting interpretation. Note that this interpretation agrees with human perception, including the inference that the black triangle must be a hole.



$E_{aa} = 0$
$E_{cs} = 2.2$
$E_{mc} = 5.2$
$E_{fc} = .9$
$E_{nc} = 0$

c

atively self-contained, yet present an interface to other components of the visual system that admits meaningful influences upon both the parameters and the outcome of the processing, as external evidence demands. This appears in the present work in at least two ways: in the ability of the energy cost objective function to accept manipulation of figural biases and accept externally derived evidence, and in the ability, not explored in this paper, for external processes to locally adjust the malliability of an interpretation by manipulating the annealing processes, for example locally lowering inverse-temperatures in certain regions.

By focusing on labeling of boundary contours alone, our approach cleanly factors away and postpones decisions about surface segmentation –which local surfaces patches are associated with one another. In addition to posing a separate surface segmentation stage, the framework we have presented raises many other possibilities for future work, including enhancement of the collection and massaging of input data, improvements to the evidence propagation machinery [25], shifting resolution of ambiguous alignment links to the annealing stage, and extensions to motion and transparency.

Acknowledgements

Many people contributed to the development of the ideas in this paper. I especially thank Lance Williams, the members of the NEC Vision Group, Allan Jepson, Yair Weiss, David Fleet, and the members of the Xerox PARC Image Understanding Area.

Appendix

The following mathematical expressions are engineered to reflect energy costs obeying figural biases. Energy costs are computed based on the spatial configuration of the BOUNDARY-CONTOUR tokens b_1 and b_2 associated with legs of L-JUNCTIONS or stems of stems T-JUNCTIONS, and these BOUNDARY-CONTOUR tokens' ends e_1 and e_2.

$$E_{aa} = C(e_1, e_2)(1 - G(b_1, b_2))$$

$$E_{cs} = (1 - C(e_1, e_2)) + G(b_1, b_2)$$

$$E_{mc} = 2E_{cs}$$

where $C(e_1, e_2)$ is a measure of the degree to which the two ends are cocircular and pointing toward one another:

$$C(e_1, e_2) = \max\left\{0, 3\max\left[0, \left(1 - \frac{|\phi_1 + \phi_2|}{\pi}\right)\right] - 2\right\} \max\left(0, 1 - \frac{|\phi_1| + |\phi_2|}{4}\right)$$

(see Figure 16a) and $G(b_1, b_2)$ is a scale normalized measure of the distance between the ends:

$$G(b_1, b_2) = \frac{d}{d + l_1 + l_2}$$

(see Figure 16b).

Alignment link quality used to select the best among all candidate alignment links is also based on the spatial relationship of the ends of boundary contours' ends.

$$Q = \frac{1}{1 + C(e_1, e_2)}$$

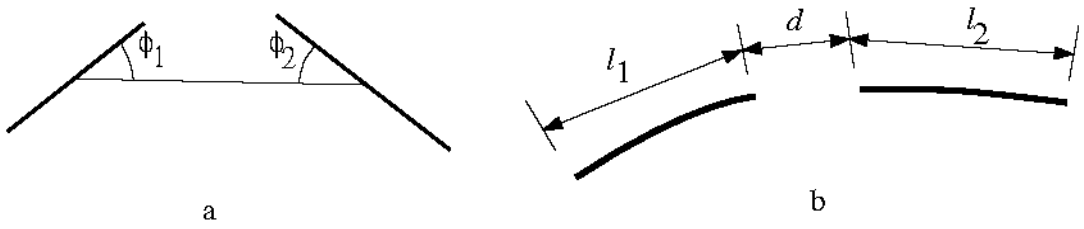


Figure 16: a. The angles ϕ_1 and ϕ_2 used to measure the alignment of two boundary contour ends. b. The lengths l_1, l_2 and d used to measure the scale-normalized distance between the ends of two boundary-contours.

References

[1] Adelson, E., and Anandan, P.; [1990]; "Ordinal Characteristics of Transparency," *AAAI-90 Workshop on Qualitative Vision.*

[2] Anderson, B.; [1997]; "A Theory of Illusory Lightness and Transparency in Monocular and Binocular Image: The Role of Contour Junctions," *Perception*, Vol. 26 No. 4, pp. 419-454.

[3] Breton, P., Iverson, L., Langer, M., and Zucker, S.; [1992], "Shading Flows and Scenel Bundles: A New Approach to Shape from Shading," in Sandini, ed., Lecture Notes in Computer Science 588 (ECCV '92), Springer-Vilag, pp. 135-150.

[4] Feldman, J., [1997]; "Efficient Regularity-Based Grouping," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Puerto Rico, pp. 288-294.

[5] Geiger, D., and Girosi, F.; [1991]; "Parallel and Deterministic Algorithms from MRF's: Surface Reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 13 No. 5, pp. 401-412.

[6] Geiger, D., Kumaran, K., and Parida, L.; [1996]; "Visual Organization for Figure/Ground Separation," *IEEE Conference on Computer Vision and Pattern Recognition,* San Francisco, pp. 155-160.

[7] Grossberg, S., and Mingolla, E.; [1985]; "Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading." *Psychological Review*, Vol. 92, pp. 173-211.

[8] Guzman, A.; [1968]; *Computer Recognition of Three Dimensional Objects in a Visual Scene*, Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.

[9] Kanizsa, G.; [1979]; *Organization in Vision*, Praeger Publishers, New York.

[10] Lowe, D.; [1985]; *Perceptual Organization and Visual Recognition*, Kluwer, Boston.

[11] Malik, J., and Shi, J; [1997]; "Grouping and Perceptual Organization," Research Overview Page,
http://HTTP.CS.BERKELEY.EDU/ jshi/Grouping/overview.html.

[12] Marr, D.; [1976]; "Early Processing of Visual Information," *Phil. Trans. R. Soc. Lond.* B Vol. 275, pp. 483-519.

[13] Marr, D.; [1982]; *Vision*, W.H. Freeman & Co., San Francisco.

[14] Maxwell, B., and Shafer, S.; [1997], "Physics-Based Segmentation of Complex Objects Using Multiple Hypotheses of Image Formation," Computer Vision and Image Understanding, Vol. 65 No. 2, pp. 265-295.

[15] Parent, P. and Zucker, S.; [1989], "Trace Inference, Curvature Consistency, and Curve Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 11 No. 8, pp. 823-839.

[16] Petry, S. and Meyer, G., [1987], eds., *The Perception of Illusory Contours*, Springer Verlag, New York.

[17] Pylyshyn, Z.; [1981]; "The Imagery Debate: Analog Media versus Tacit Knowledge," in Block, N., ed., *Imagery*, MIT Press, Cambridge, MA.

[18] Richards, W., Jepson, A., and Feldman, J.; [1996]; "Priors, Preferences, and Categorical Percepts," in D. Knill and W. Richards, eds., *Perception as Bayesian Inference*, Cambridge University Press,.

[19] Rose, K., Gurewitz, E., and Fox, G.; [1990]; "A Deterministic Annealing Approach to Clustering," *Pattern Recognition Letters,* Vol. 11 No. 9, pp. 589-594.

[20] Rosenfeld, A., Hummel, R., and Zucker, S.; [1976], "Scene Labeling by Relaxation Operations," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 6, pp. 420-433.

[21] Saund, E.; [1995]; "A Multiple Cause Mixture Model for Unsupervised Learning," *Neural Computation,* Vol. 7, pp. 51-71.

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

[23] Ullman, S.; [1979]; *The Interpretation of Visual Motion,* MIT Press, Cambridge, MA.

[24] Waltz, D.; [1975]; "Understanding Line Drawings of a Scene With Shadows," in *The Psychology of Computer Vision*, P. H. Winston, ed., McGraw-Hill., New York.

[25] Weiss, Y.; [1999]; "Correctness of Local Probability Propagation in Graphical Models with Loops," *Neural Computation*, in press.
[26] Williams, L.; [1990]; "Perceptual Organization of Occluding Contours," *Proc. Third International Conference on Computer Vision*, Osaka, pp. 639-649.
[27] Williams, L., and Hanson, A.; [1996]; "Perceptual Organization of Occluded Surfaces," *Computer Vision and Image Understanding* Vol 64 No. 1, pp. 1-20.
[28] Williams, L., and Jacobs, D.; [1995]; "Stochastic Completion Fields: A Neural Model of Illusory Contour Shape and Salience," *Proc. Fifth International Conference on Computer Vision*, pp. 408-415.

References

[1] Adelson-and-Anandan

[2] Anderson

[3] Breton-etal

[4] Feldman

[5] Geiger-and-Girosi

[6] Geiger-etal

[7] Grossberg-and-Mingolla

[8] Guzman

[9] Kanizsa

[10] Lowe

[11] Malik-and-Shi

[12] Marr-primal-sketch

[13] Marr-vision

[14] Maxwell-and-Shafer

[15] Parent-and-Zucker

[16] Petry-and-Meyer

[17] Pylyshyn

[18] Richards-etal

[19] Rose-etal

[20] Rosenfeld-etal

[21] Saund-mcmm

[22] Ullman

[23] Ullman-sfm

[24] Waltz

[25] Weiss

[26] Williams

[27] Williams-and-Hanson

[28] Williams-and-Jacobs