Perceptual Organization of Occluding Contours Generated By Opaque Surfaces

Eric Saund
Xerox Palo Alto Research Center
3333 Coyote Hill Rd.
Palo Alto, CA, 94304
http://www.parc.xerox.com/saund

Abstract

This paper offers computational theory and an algorithmic framework for perceptual organization of image contours arising from static occluding surfaces of constant lightness. We articulate constraints and biases underlying the inference of such physical events as visible surface overlap and invisible (modal and amodal) surface boundaries, from ambiguous visual evidence including visible contrast edges and L-type and T-type junctions. For any given scene, an energy or cost function is constructed over interpretation labels for nodes of a sparse graph, or belief net. Annealing-style optimization permits local cues to 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

A longstanding problem in computational vision is to sort out the various contrast edges found in images to infer overlap and depth relations among the surfaces that generated them. Often the physical configuration of objects is underconstrained by the limited information available in a single view, requiring additional constraints or assumptions to be brought into play.

A classic figure illustrating this phenomenon is the "Kanizsa Triangle" [5](Figure 1). Several physically consistent interpretations of the image are possible. But the preferred human percept rejects the perfectly plausible and simple interpretation that all of the black figures are isolated objects (1b), in favor of the apparition of a large white occluding triangle (interpretation 1c, and alternatively, 1d). This is presumably due to the visual system's refusal to accept that certain contour alignments could have arisen by "accident," and that, therefore, an occluding triangle is the most likely remaining explanation. Note however that interpretations 1c and 1d require the postulation of a competing accident of a different kind, namely, that a background plane and the foreground triangle are coincidentally the same lightness, which accounts for the lack of a contrast edge where they overlap.[1] Through

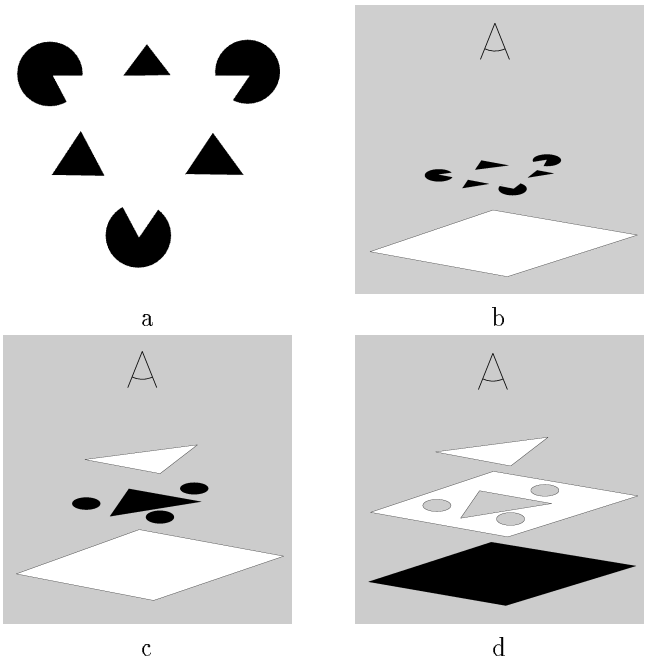

Figure 1: a. Kanizsa Triangle b. Plausible but non-preferred interpretation. c. Preferred interpretation. d. Alternate preferred interpretation.

[1] In fact, the implausibility of *this* accident may underlie the apparent lightness edges induced at the triangle's bound-

this and similar illustrations the Gestalt psychologists demonstrated that a complex and subtle interplay of biases and assumptions must underlie perception of seemingly simple scenes [6]. We seek formal accounts for these processes.

The pioneering computational study of perceptual organization in the "Colorforms" domain occurred with Williams' [11, 12] delineation of a mapping between sparse, observable, *image* level events such as contrast edges and junctions, and a set of *interpretation labels* for physical events which might or might not give rise to visible evidence. Of particular interest is the inference of unobserved surface boundaries known as modal and amodal completion edges, as discussed below. Williams' formulation reflected his goal of obtaining a globally consistent map of relative surface depths: a scene gives rise to an integer linear programming problem in which continuous-valued figural biases reflect Gestalt preferences for closure and good continuation within a strict physical feasibility space circumscribing the set of possible interpretations for that scene. This model unfortunately foregoes purely local use of local evidence, as evidenced by its prohibition of globally *inconsistent* but perceptually phenomenal interpretations of scenes such as Figure 2.

More recently, Geiger et al. [3] adopted a set of physical interpretations for L-junctions more closely resembling that recognized by psychophysicists [1], but formulated in a relaxation labeling scheme whereby interpretation labels are diffused in a fine-grain two-dimensional field while pinned at the relatively few locations containing contrast data in the input image.

The present work offers a new formulation for the perceptual organization of occluding contours. In Section 2, computational theory is developed that incorporates a richer ontology of image junction interpretations than previously has been entertained, and that elucidates tradeoffs among competing perceptual baises. Section 3 proposes an algorithmic framework based on a token-based representation that is parsimonious and efficient in the declaration of equivalence classes of image events. Token labels reflecting the physical interpretation of a given scene are assigned through annealing-style optimization. The formulation permits information from spatially localized cues to be used purely locally as well as to propagate globally.

ary where none exists in the image data.

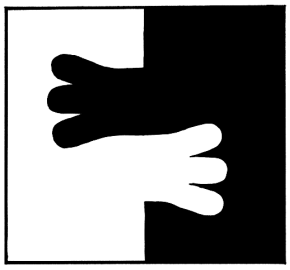

Figure 2: Strong local figure/ground pressures can prohibit globally consistent figure/ground assignments for all contour edges.

2 Computational Theory for Occluding Opaque Surfaces

A computational theory for the perceptual organization of occluding surfaces must address both what *can* happen in the mapping from the physical world to images, and assumptions about what *tends* to happen, the latter providing justification for interpretation biases. 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 Junction Label Catalog

We introduce in Figure 3 a catalog of possible physical interpretations of visible contrast edges and junctions resulting from 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 elemental primitive is the visible contrast edge, or boundary contour, which arises from the overlap of two surfaces of different lightness. This event can take one of two label values indicating which of the surfaces is in front, as indicated by the direction of a wedge arrow in the figure.

The junction catalog articulates the rich ambiguity inherent in local image measurements. For example, a visible L-junction can arise from a convex or concave physical corner (L1 and L2), or, as a "degenerate" T-junction if surfaces coincidentally share the same lightness (L3-L6). Such ambiguities can be resolved only by marshalling more global evidence under the guidance of interpretational biases.

The invisible contour boundary generated when one surface overlaps another of the same lightness is known as a *modal completion* edge.[2] An *amodal completion*,

[2] 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

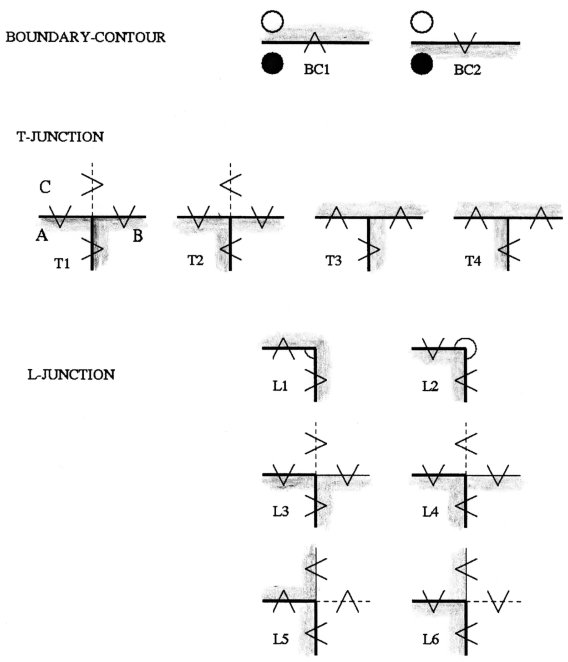

Figure 3: Catalog of interpretation labels for BOUNDARY-CONTOURS, T-JUNCTIONS, and L-JUNCTIONS. Heavy lines denote contrast edges, solid thin lines denote modal completion contours, dashed lines denote occluded contours. Arrows and shading indicate direction of surface overlap: tip of arrow indicates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

or *occluded* contour occurs when a surface boundary is occluded by another surface, as indicated by dashed lines.

2.2 Figural Biases

The Kanizsa Triangle demonstrates that modal completion edges are readily inferred in the context of conspicuously aligning visible contrast edges. Although the assertion of a modal completion interpretation requires postulation of a nongeneric or "accidental" coincidence in lightness of distinct and overlapping surfaces, this explanation allows the alignment to be explained by a single boundary contour instead of by the "accidental" alignment of contours from independent objects.[3] Thus, we require means to express tradeoffs among interpretational biases or preferences drawing from underlying assumptions about what types of physical events are more or less likely to occur.

This may be accomplished though the mechanism of an energy or cost function that assigns penalties for junctions adopting certain labels, depending on the severity of any nongeneric inferences these labels induce in the context of the particular image geometry. In implementation, the terms contributing to energy cost are all simple mathematical expressions engineered to take particular functional forms justified on the basis of commonsense evaluations of sample configurations and the human visual system's behavior on simple stimuli. See [8] for details. What is most important for the present purposes is to get their qualitative behavior right, to reflect the following figural biases, illustrated in Figure 4:

- **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.

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 their perceived shapes.

[3] A more refined treatment of the nature of generic versus nongeneric events is presented in [8], from which the present paper is condensed.

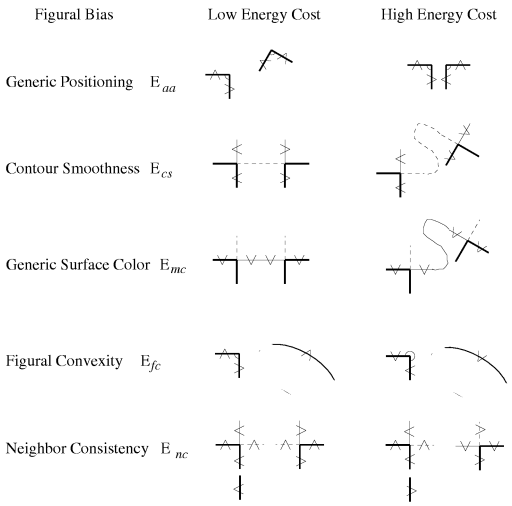

Figure 4: Schematic illustration of quantitative figural biases.

- **Neighbor Consistency.** An energy cost E_{nc} (*nc :: neighbor consistency*) expresses a penalty incurred for every instance that a junction interpretation label conflicts with that of the constituent boundary contours that comprise it. This amounts to a firm but not unyielding bias for figural consistency, that is, that foreground objects do not arbitrarily flip occlusion direction to become background.

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-contour and junction labels. Figure 5 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 5a 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 5b 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 5c shows that the strongly nonpreferred interpretation—isolated objects with no occlusion—pays a very high penalty for accidental edge alignments. In all cases presented in the paper the optimal labeling is attained by the algorithm of the next section.

3 Search Over Labelings of a Junction Graph

This machinery may be instantiated for real and synthetic Colorforms scenes through an algorithm proceeding in two phases, a problem setup phase and a solution phase.

Input data consists of chain-coded contours such as found by edge detection and curve tracing processes, and annotated with the colors of the surfaces on each side. Straightforward techniques are used to detect corners and junctions. Spatial relations among junctions are analyzed to construct a *junction graph* whose nodes are boundary contours, L-junctions, and T-junctions, and whose links are of two kinds. *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. *Alignment links* found by simple grouping techniques declare pairs of contour ends that are preferably near to and align with one another across pairs of L- or T-junctions. See Figure 6.

An interpretation of contour overlap relations consists of a labeling of the nodes of the junction graph, where the links provide and propagate constraint on nodes' labels by means of the figural biases presented above.

In the solution phase, search over the space of junction label assignments is conducted with the goal of minizing energy cost. We employ a continuation method which permits decisions to be made gradually as local preferences propagate around the graph.

While the final solution goal is a discrete choice over junction labels, we employ a "soft" representation ad-

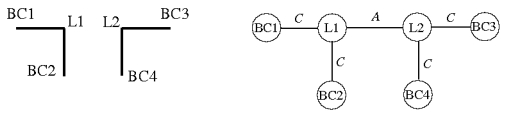

Figure 6: Four BOUNDARY-CONTOURS forming two L-JUNCTIONS, and their corresponding junction-graph containing coincidence links (C) and an alignment link (A).

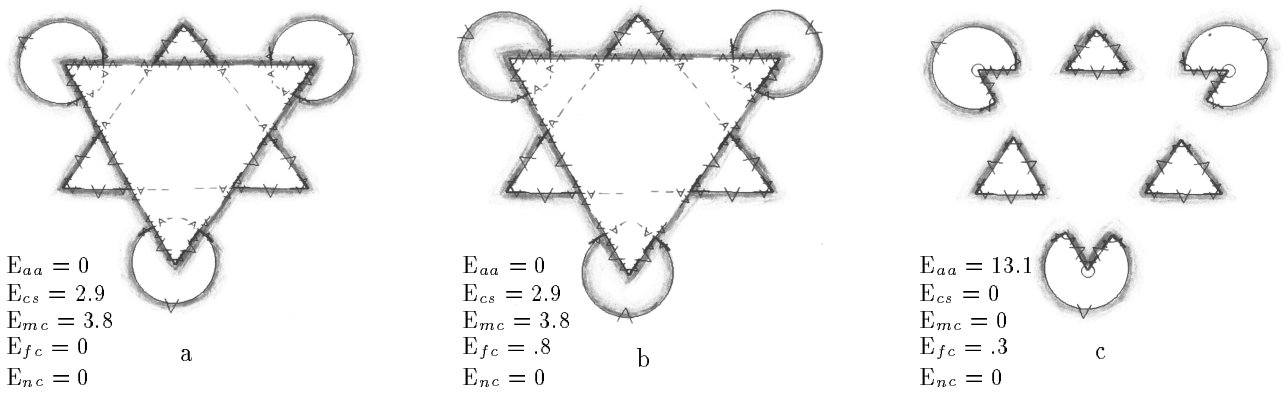

Figure 5: Three among many possible interpretation labelings of the Kanizsa Triangle, and their associated energy costs. All labelings are stable at high inverse-temperature, but a. (the global optimum) is converged to by the algorithm annealing from low inverse-temperature. Shading was done by hand to accentuate program output (arrowheads).

mitting continuous-valued beliefs $0 \leq b_{j,l} \leq 1$ in the validity of interpretation label l for node j. Because each visible junction is assumed to arise from one and only one physical cause, we impose the further constraint, $\{\forall j : \sum_l b_{j,l} = 1\}$, giving the belief vector resemblance to a probability distribution over interpretation states. Calculation of an energy cost for a given belief state of the junction graph proceeds using the figural biases above, now weighted according to the belief vectors of nodes participating across coincidence and alignment links. Subtleties of this process are discussed further in [8].

The global energy cost is simply the sum of energy costs for individual nodes in the junction graph. A node's local energy cost is determined by its own belief vector in relation to those of its link neighbors according to the figural biases 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*[2, 7]. 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, $\beta = 0.5, 1, 2, 3, 10$. Scenes of this complexity are interpreted in less than a minute on a Symbolics XL1200 Lisp Machine.

4 Results

In addition to the Kanizsa triangle result of Figure 5a, Figure 7 presents input images along with interpretations found by the algorithm for two representative situations. Figure 7a/b shows the algorithm working with somewhat noisy input obtained from a video frame of a construction paper scene under standard office lighting. Figure 7c/d demonstrates a scene whose preferred interpretation includes a nongeneric edge alignment at a T-junction.

Simulated annealing methods are known to be susceptible to local optima, and Colorforms scenes can be constructed for which the described diffusion-style algorithm fails. For these, more sophisticated methods resembling those for solving Markov nets are under investigation [10].

Figure 8 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.

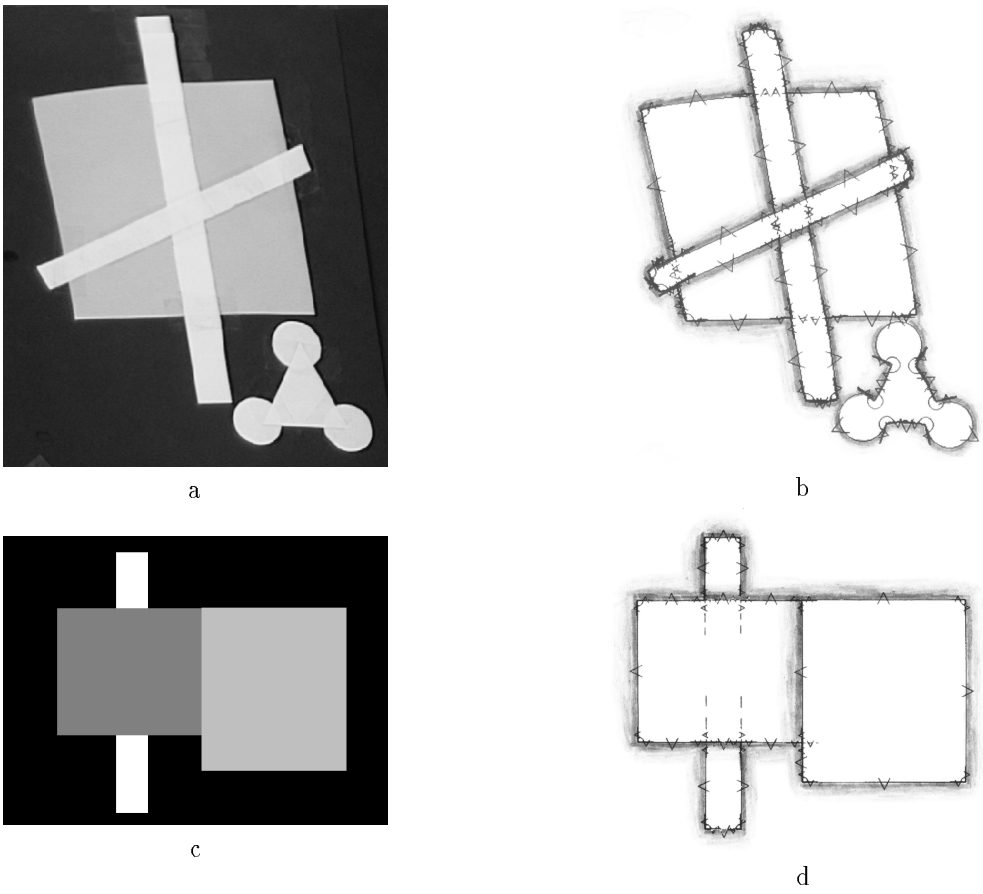

Figure 7: a/b Frame from video input of a construction paper scene. c/d The preferred percept requires a nongeneric interpretation (interpretation label T4) for the upper right T-junction.

5 Conclusion

This paper has presented computational analysis of junction labels and figural biases underlying the perceptual organization of occluding contours of static opaque surfaces, and has outlined an algorithmic model that allows locally-derived constraints to influence one another by propagating around a junction graph, leading to a global interpretation. The model's predictions accord with human perception for a variety of scenes, and account for preferred and coherent interpretations under subtle interactions of locally ambiguous cues.

By focusing on labeling of boundary contours alone, the present approach cleanly factors away and postpones decisions about surface segmentation–which local surfaces patches are associated with one another. In addition to setting the stage for a modular surface segmentation process, 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, resolving ambiguous alignment links in the annealing stage, and extensions to motion and transparency.

Acknowledgments

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, and the members of the Xerox PARC Image Understanding Area.

References

[1] B. Anderson and B. Julesz, "A Theoretical Analysis of Illusory Contour Formation in Stereopsis," *Psychological Review*, Vol. 102, 705-743, 1995.

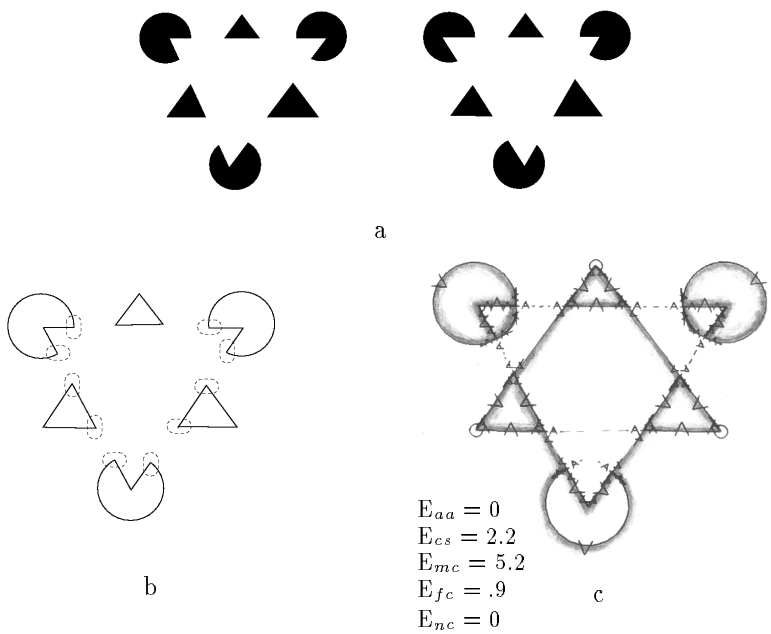

Figure 8: Stereo evidence influences surface overlap perception. a. Stereo disparities at pacmen leads to perception (under cross fusing) of the white triangle weaving behind a flat planar surface. b. Input from a stereo module was simulated by effectively constraining certain L-junction interpretation 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.

[2] D. Geiger and F. Girosi, "Parallel and Deterministic Algorithms from MRF's: Surface Reconstruction," *IEEE TPAMI* Vol. 13 No. 5, pp. 401-412, 1991.

[3] D. Geiger, K. Kumaran, and L. Parida, "Visual Organization for Figure/Ground Separation," *IEEE CVPR,* San Francisco, pp. 155-160, 1996.

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

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

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

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

[8] E. Saund, "Perceptual Organization of Occluding Contours of Opaque Surfaces," CVPR '98 Workshop
on Perceptual Organization in Computer Vision, *http://www.parc.xerox.com/saund/papers.html*, 1998.

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

[10] Y. Weiss, "Belief propagation and revision in networks with loops," MIT AI Memo 1616 (CBCL Paper 155). Presented in NIPS*97 workshop on graphical models, 1997.

[11] L. R. Williams, "Perceptual Organization of Occluding Contours," *Proc. Third International Conference on Computer Vision*, Osaka, pp. 639-649, 1990.

[12] L. R. Williams and A. R. Hason, "Perceptual Organization of Occluded Surfaces," *CVIU* Vol. 64, No. 1, pp. 1-20, 1996.