A Boundary-pair Representation for Perception Modeling

Xiuwen Liu and DeLiang L. Wang
Department of Computer and Information Science
The Ohio State University
Columbus, OH 43210-1277
{liux, dwang}@cis.ohio-state.edu

Abstract

It is widely accepted that responses from on- and off-center cells give rise to edges and are equivalent to edge detectors. In this paper, we point out that on- and off-center cell responses provide more information than edges. We show that an edge-based representation makes the ownership of boundaries ambiguous and requires a combinatorial search to model perceptual grouping. By analyzing the differences between edges and responses from on- and off-center cells, we propose a boundary-pair representation, which makes the ownership of boundaries explicit and eliminates the need of a combinatorial search computationally. Each boundary in the boundary-pair representation is associated with regional attributes. We show that this representation is equivalent to a surface representation through a local diffusion. This provides a unified representation for perception modeling. Based on this representation, a figure-ground segregation network is constructed to demonstrate the capabilities of the model in explaining many perceptual phenomena.

1 Introduction

The ultimate goal of a generic machine vision system is to extract meaningful description from input images. Motivated by physiological and psychological experiments [4], Marr and his colleagues proposed the most influential paradigm for computational vision [7]. Based on that simple cells in the visual cortex response to luminance differences, Marr and Hildreth proposed a widely used edge detector, namely zero-crossing of response of Laplacian of Gaussian (LoG) operator [8]. Marr further proposed a computational paradigm for a generic vision system based on the output from edge detector, grouping edge elements and making cues such as depth explicitly. While Marr's paradigm for computational vision is most influential, his ideas have not been fully implemented. One of the main reasons is that each computational module in the paradigm is underconstrained [9].

Nakayama and his colleagues argued based on extensive psychological experiments that visual surface representation is a critical link between low-level and high-level vision [10]. For perceptual organization and grouping, they propose that the ownership of boundaries determines the relative depth between competing boundaries. However, they did not give a computational model for determining the ownership of boundaries and generating virtual surface representation.

To illustrate the problem, Figure 1(a) and (b) show the on- and off-center cells modeled using Laplacian of Gaussian operator. Figure 1(c) shows ideal step edges, and Figure 1(d) and (e) show the filter responses of on-center and off-center filters, respectively. Figure 1(f) shows the edge points detected by zero-crossing. While the edge detector generates good engineering results, Using the output as the input to high-level modules for grouping introduces further ambiguities. As shown in Figure 1(f), edge points do not belong any region and so no regional attributes can be associated. The ownership of the boundaries becomes inherently ambiguous. This ambiguity results in the need for a combinatorial search to model perceptual organization [12] [1].

Figure 1(g) and (h) show the responses from on- and off-center cells, obtained by applying a half-way rectified function on the corresponding filter responses. If we examine more carefully the output from on- and off-center cells, we can see that there are important differences from the edge points generated by the edge detector shown in Figure 1(e). In the on- and off-center cell responses, each boundary can be associated with regional attributes such as intensity values and colors. The ownership is not ambiguous and can be inferred from the associated attributes. Grossberg and his group have studied perceptual grouping extensively based on on- and off-center cell responses [3] [2]. In their model, on- and off-center cells are modeled as relatively independent channels. However, as shown in Figure 1, a region, namely, the central gray region, is associated with both on- and off-center cell responses. For grouping purposes, on- and off-center cells should not be treated as independent channels.

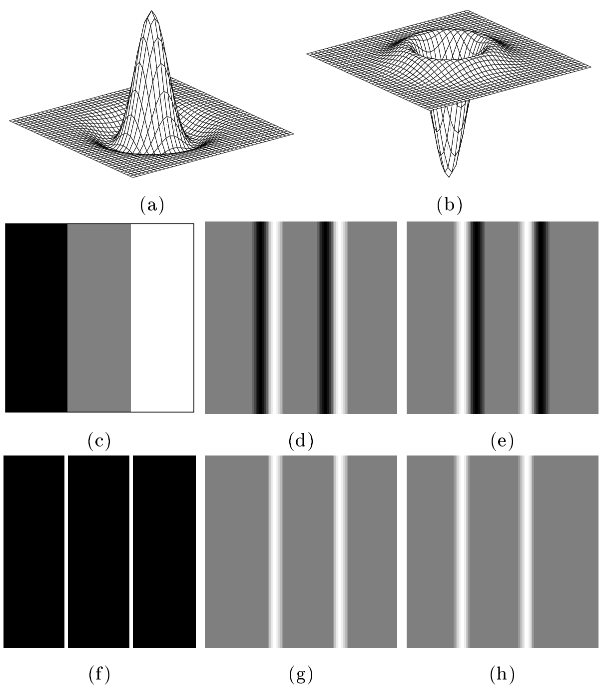


Figure 1: On- and off-center cell responses. (a) On-center filter modeled using LoG. (b) Off-center filter using LoG. (c) Input image. (d) On-center filter response. (e) Off-center filter response. (f) Edges using zero-crossing. (g) On-center cell responses. (h) Off-center cell responses.

To overcome some of the problems and unify different representations, we propose a boundary-pair representation. A boundary between two regions is represented by a pair of boundaries with different region attributes associated. The paired boundaries compete with other of being figural. In case of an occluding boundary, one boundary is more likely to be figural based on the contextual information. Under this representation, the problem for grouping and perceptual organization is to determine the status of the boundary, not the ownership. The dynamic nature of competition can also explain unstable percept for images like the vase-face example. This also eliminates the need of a combinatorial search. Based on this representation, we have proposed and implemented a network for figure-ground segregation [5]. In this paper, we mainly focus on the properties of the boundary-pair representation. Please refer to [5] for details regarding the figure-ground segregation network.

In Section 2 we introduce boundary-pair representation and show that it is equivalent to a surface representation. Section 3 gives a brief summary about our figure-ground segregation network. Section 4 shows experimental results. Section 5 concludes the paper with further discussions.

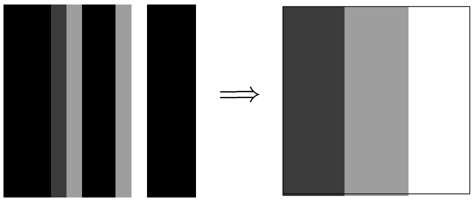
2 Boundary-pair Representation

Given an input image, we apply both on- and off-center cells to extract boundaries. The responses from the on- and off-center cells are combined to obtain a boundary-pair representation.

2.1 Necessity for a boundary-pair representation

If we ignore the signs of cell responses, we obtain an edge representation as proposed by Marr [7]. Because an edge corresponds to luminance discontinuity, it does not belong to any region. As shown in Fig. 1(f), the detected edges are between regions with different intensity values. Because of this, no regional attributes can be associated with an edge unless we can solve the ownership problem. Because of this inherent ambiguity under an edge representation, to model perceptual organization, one has to try different combinations of edge elements in order to derive the most likely percept. For example, Williams and Hanson [12] posed the perceptual organization as a line drawing explanation problem. The grouping problem is then formulated as an integer linear program, which is an NP problem. Geiger et al [1] posed the grouping problem as hypothesis selection. The number of possible hypotheses is combinatorial. For their system, in order to generate desirable outputs, the size and shape of hypotheses have to be given manually.

This clearly demonstrates that a boundary-pair representation is necessary in order to model the perceptual organization and grouping more efficiently and effectively. In cases of occluding boundaries, the boundary-pair encodes the ownership through a boundary's status being figural. However, this boundary-pair representation is more generic than ownership determination. By formulating the problem as an ownership determination, one implicitly assumes that every boundary is an



(a) (b)

Figure 2: A boundary-pair representation and the resulting surface representation through local diffusion. (a) A boundary-pair representation. (b) The result from a local diffusion.

occluding one, which may not be not true in general. In the boundary-pair representation, both the boundaries in a pair can be figural, resulting in capabilities for modeling boundaries between two non-overlapping boundaries, as shown in Fig. 1(c).

2.2 Equivalence to a surface representation

In the boundary-pair representation, each boundary element can be associated with regional attributes, for example, intensity value. By doing so, this boundary-pair representation is equivalent to a surface representation but more parsimonious.

To generate a surface representation, a local diffusion is used. In the diffusion, each boundary acts as a fixed heat source. Here the temperature represents the intensity value. Also each boundary does not allow diffusion from neighboring heat sources. This defines a well-studied mathematical problem for heat diffusion with boundary conditions. By solving the diffusion equation or through numerical simulation, one can generate a surface from the sparse data. Figure 2(a) shows a boundary pair representation for the image shown in Fig. 1(c). In Fig. 2(a), the values of black pixels are not available and need to be determined. In order to show the background, values of boundary values are shifted. Figure 2(c) shows the resulting surface representation.

As discussed earlier, the equivalence does not hold for an edge-representation because regional attributes of edges are not known. In addition, grouping has to be performed based on the regional attributes associated with boundaries because the responses of on- and off-center cells depend on surrounding regions and do not give an intrinsic representation.

3 Figure-Ground Segregation Network

In the boundary-pair representation, the problem of perceptual organization and grouping is to determine the status of boundaries, and the ownership and regional attributes are explicit. This leads to a network for figure-ground segregation [5] . In the network, each boundary is represented by a node with associated attributes. Neighboring nodes that have similar attributes are connected exicatorily. Paired boundaries are connected inhibitorily and compete with each other to be figural. Long-range connections are introduced based on Gestalt-like grouping rules and corners and junctions. In this network, the figure-ground segregation is solved through temporal evolution.

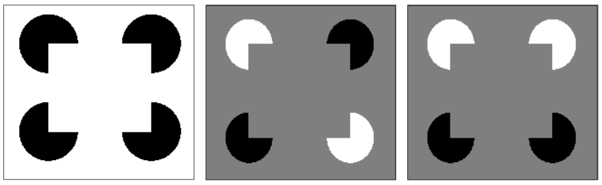
The network provides a way to incorporate contextual information and explains many perceptual phenomina in a unified way. Through temporal dynamics, some perceptual phenomina that are difficult to model computationally, such as unstable percept, can be explained using the same network for virtual contours and amodal completion. This will be demonstrated through examples in the next section. For more details regarding the network, refer to [5].

4 Experimental Results

The experiments shown in this paper are generated using a figure-ground segregation network [5]. For an input image, the nodes and the connections in the network are constructed automatically. The parameters for the figure-ground segregation network are fixed.

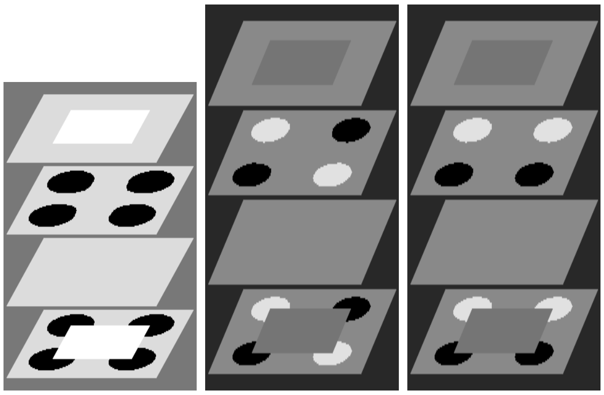
4.1 Virtual contours

We first demonstrate that the system can simulate virtual contours and modal completion. Figure 3 shows three variations of pacman images. The results from our system are shown in Fig. 4. Each layer represents regions belong to the same depth order and layers are arranged according to the relative depth. Even though the pacman patterns vary in colors and thus the associated edges vary in constrast signs, the virtual contour of the central square is evident in all the three cases. In the boundary-pair representation, because the color of the central square is same regardless of the variations of pacman patterns, our system correctly handles all of them in a similar way and generates correct results. On the other hand, edge-based approaches tend to have problems, as pointed out by Williams and Hanson [12]. This is because the edges have different contrast signs due to the variations of the pacman patterns. This



(a) (b) (c)

Figure 3: Images with virtual contours. (a) Original pacman image. (b) Mixed pacman image. (c) Alternate pacman image.



(a) (b) (c)

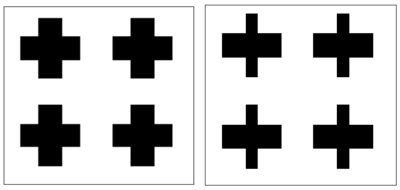
Figure 4: Layered representation of surface completion for images shown in Fig. 3.

example is strong evidence for boundary-bair representation.

Figure 5 shows two images, where the optimal percept is difficult to be simulated by a single existing model. Our system, even with fixed parameters, generates the outputs shown in Fig. 6 due to that the system allows interactions between shape information and non-accidental alignment. In Fig. 5(a), the symmetric crosses are more stable and the lateral connections are much stronger, and the perception of four crosses generated from the system is consistent with that in the psychological literature [6]. In Fig. 5(b), the crosses are not symmetric any more and are perceived as overlapping rectangular bars, which is shown in Fig. 6(b). Both models by Williams and Hanson [12] and Geiger et al. [1] do not correctly handle the case shown in Fig. 5(a).

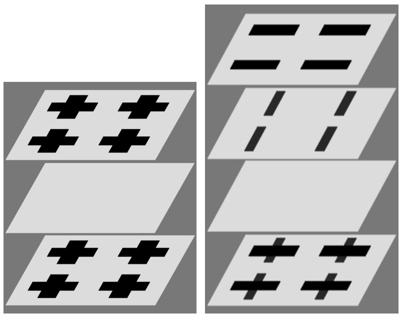
4.2 Multiple solutions

While in many cases the percept is stable and unique, figure-reversal occurs under circumstances. For example, Figure 7(a) shows an image, where either two faces or vase can be perceived, but not both at the same time. In order to account for this phenomenon, clearly multiple stable solutions are needed. Our models can give an explanation by incorporating some form of attention-based top-down influence, which was believed to play an important role in this phenomenon. Through temporal dynamics, we can naturally obtain multiple solutions by utilizing the additional *time* dimension, which is unique to dynamic systems.



(b) (c)

Figure 5: Images with virtual contours. (a) Four crosses. (b) Overlapping rectangular bars.



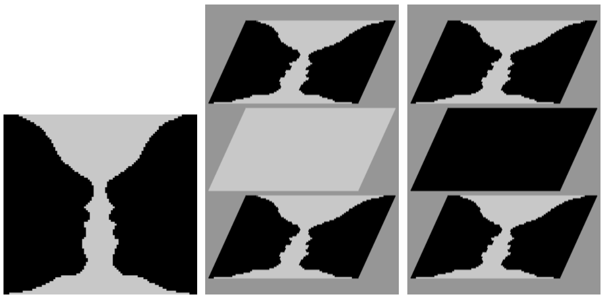
(a) (b)

Figure 6: Surface completion results for the corresponding image in Fig. 5.

In this paper, we demonstrate this through self-inhibition, motivated by habituation [11]. It is well known that the strength of responses decreases when a stimulus is presented repeatedly. Figure 7(b) and (c) show the two possible results using layered representation. In Fig. 7(b), two faces are perceived and and the vase is suppressed into the background; Fig. 7(c) shows the other case. Here the differences can be seen from the middle layer. By introducing habituation, our system offers a computational explanation. As shown in Fig. 8, two faces and vase alternate to be figural, resulting in bistable percept. This example demonstrates that top-down influence from memory and recognition can be naturally incorporated in the network.

5 Conclusions

In this paper, we propose a boundary-pair representation for perception modeling based on the fact that there exist both on- and off- center cells in the visual cortex. By using a pair to represent boundaries between regions, the ownership of a boundary is made explicit and thus regional attributes can be associated with a boundary. This representation is equivalent to a surface representation through local diffusion and thus provides a unified representation for edge- and surface-based representation. This also eliminates the neeed of a combinatorial search for perceptual organization and grouping, which is widely used in the literature. While the boundary-pair representation makes the ownership explicit, it does not necessarily assume that only one boundary from a pair can be figural, resulting a more generic representation than ownership determination.



(a) (b) (c)

Figure 7: Bistable perception. (a) Face-vase input image. (b) Faces as figures. (c) Vase as figure.

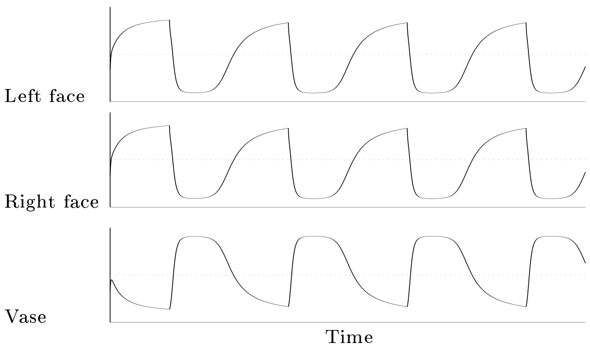


Figure 8: Temporal behavior of the system for Fig. 7(a). Dotted lines are 0.5.

Acknowledgments

Authors would like to thank valuable discussions with S. C. Zhu and M. Wu. This research is partially supported by an NSF grant (IRI-9423312) and an ONR Young Investigator Award (N00014-96-1-0676) to DLW and a Presidential Fellowship to XL from the Ohio State University.

References

[1] D. Geiger, H. Pao, and N. Rubin, "Salient and multiple illusory surfaces," In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 118-124, 1998.

[2] A. Gove and S. Grossberg and E. Mingolla, "Brightness perception, illusory contours, and corticogeniculate feedback," *Visual Neuroscience*, vol. 12, pp. 1027-1052, 1995.

[3] S. Grossberg and E. Mingolla, "Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations," *Perception & Psychophysics*, vol. 38, pp. 141-170, 1985.

[4] D. H. Hubel, *Eye, Brain, and Vision*, W. H. Freeman and Company, New York, 1988.

[5] X. Liu and D. L. Wang, "Perceptual Organization based on Temporal Dynamics," In *Proceedings of International Joint Conference on Neural Networks*, 1999.

[6] G. Kanizsa, "Quasi-perceptual margins in homogeneously stimulated fields," In *The Perception of Illusory Contours*, S. Petry and G. E. Meyer, eds., Springer-Verlag, New York, pp. 40-49, 1987.

[7] D. Marr, *Vision*, W. H. Freeman and Company, New York, 1982.

[8] D. Marr and E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London, Series B*, vol. 207, pp. 187-217, 1980.

[9] J. D. McCafferty, *Human and Machine Vision*, Ellis Horwood, London, 1990.

[10] K. Nakayama, Z. J. He, and S. Shimojo, "Visual surface representation: A critical link between lower-level and higher-level vision," In *Visual Cognition*, S. M. Kosslyn and D. N. Osherson, eds., The MIT Press, Cambridge, Massachusetts, vol. 2, pp. 1-70, 1995.

[11] D. L. Wang, "Habituation," In *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, ed., The MIT Press, Cambridge, Massachusetts, pp. 441-444, 1995.

[12] L. R. Williams and A. R. Hanson, "Perceptual Completion of Occluded Surfaces," *Computer Vision and Image Understanding*, vol. 64, pp. 1-20, 1996.