Perceptual Organization based on Temporal Dynamics

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

This paper presents a computational model for perceptual organization. A figure-ground segregation network is proposed based on a novel boundary pair representation. Nodes in the network are boundary segments obtained through local grouping. Each node is excitatorily coupled with the neighboring nodes that belong to the same region, and inhibitorily coupled with the corresponding paired node. The status of a node represents the probability of the node being figural and is updated according to a differential equation. The system solves the figure-ground segregation problem through temporal evolution. Gestalt-like grouping rules are incorporated by modulating connections, which determines the temporal behavior and thus the perception of the system. The results are then fed to a surface completion module based on local diffusion. Different perceptual phenomena, such as modal and amodal completion, virtual contours, grouping and shape decomposition are explained by the model with a fixed set of parameters. Computationally, the system eliminates combinatorial optimization, which is common to many existing computational approaches. It also accounts for more examples that are consistent with psychological experiments. In addition, the boundary-pair representation is consistent with well-known on- and off-center cell responses and thus biologically more plausible.

1 Introduction

Perceptual organization refers to the ability of grouping similar features in sensory data. This, at a minimum, includes the operations of grouping and figure-ground segregation. Here grouping includes both local grouping, generally known as segmentation, and long-range grouping, referred to as perceptual grouping in this paper. Figure-ground segregation refers to the process of determining the relative depth of adjacent regions in input data. This problem setting has several computational implications. The central problem in perceptual organization is figure-ground segregation. When the relative depth between regions is determined, different types of surface completion phenomena, such as modal and amodal completion, shape composition and perceptual grouping, can be solved and explained using a single framework. Perceptual grouping can be inferred from surface completion. Grouping rules, such as those summarized by Gestaltists, can be incorporated for figure-ground segregation.

Many computational models have been proposed for perceptual organization. Many of the existing approaches [8] [4] [12] [13] [3] start from detecting discontinuities, i.e. edges in the input; one or several configurations are then selected according to certain criteria, for example, non-accidentalness [8]. Those approaches to a larger extend are influenced by Marr's paradigm [10], which is supported by that on- and off-center cells response to luminance differences, or edges [6], and that the three-dimensional shapes of the parts can be inferred from a two-dimensional line drawing [1]. While those approaches work well to derive meaningful two-dimensional regions and their boundaries, there are several disadvantages for perceptual organization. Theoretically speaking, edges should be localized between regions and do not belong to any region. By detecting and using edges from the input, an additional ambiguity, the ownership of a boundary segment, is introduced. Ownership problem is equivalent to figure-ground segregation [11]. Due to that, regional attributions cannot be associated with boundary segments. Further more, because each boundary segment can belong to different regions, the potential search space is combinatorial; constraints among different segments such as topological constraints must be incorporated explicitly [13]. Furthermore, obtaining the optimal configuration(s) is computationally expensive.

To overcome some of the problems, we propose a laterally-coupled network based on a boundary-pair representation [7]. An occluding boundary is represented by a pair of boundaries of the two involved regions, and initiates a competition between the regions. Each node in the network represents a boundary segment. A closed region boundary is represented as a ring structure with laterally coupled nodes. A region consists of one or more rings. Regions compete to be figural through boundary-pair competition and the figure-ground segregation is solved through temporal evolution. Gestalt grouping rules are incorporated by modulating the coupling strength between different nodes within a region, which influences the temporal dynamics and determines the perception of the system. Shape decomposition and grouping are implemented through local diffusion using the results from figure-ground segregation. This approach offers several advantages over edge-based approaches:

- Boundary-pair representation makes explicit the ownership of boundary segments and eliminates the combinatorial optimization necessary for many existing approaches [13] [3].
- The model can explain more perceptual phenomena than existing approaches using a fixed set of parameters.
- It can incorporate top-down influence naturally.

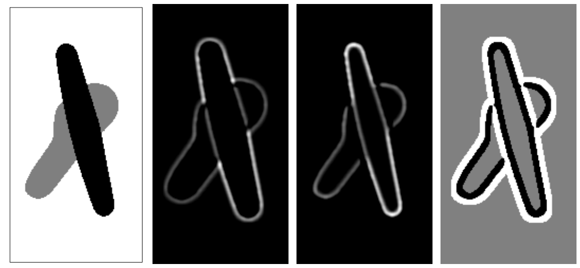
In Section 2 we introduce figure-ground segregation network and demonstrate the temporal properties of the network. Section 3 shows how surface completion and decomposition are achieved. Section 4 provides experimental results. Section 5 concludes the paper with further discussions.

2 Figure-Ground Segregation Network

The central problem in perceptual organization is to determine the relative depth among regions. As figural reversal occurs in certain circumstances, figure-ground segregation cannot be resolved only based on local attributes. By using a boundary-pair representation [7], the solution to figure-ground segregation is given by temporal evolution.

2.1 Boundary-Pair Representation

The boundary-pair representation is motivated by on- and off-center cell responses. Fig. 1(a) shows an input image. Fig. 1(b) and (c) show the on-center and off-center responses. Without zero-crossing, we naturally obtain double responses for each occluding boundary, as shown in Fig. 1(d).



(a) (b) (c) (d)

Figure 1: On- and off-center cell responses. (a) Input image. (b) On-center cell responses. (c) Off-center cell responses (d) Binarized on- and off-center cell responses. White regions represent on-center response regions and black off-center regions.

More precisely, closed region boundaries are obtained from segmentation and then segmented into segments using corners and junctions, which are detected through local corner and junction detectors. A node i in the figure-ground segregation network represents a boundary segment, and its status P_i represents the probability of the corresponding segment being figural, which is set to 0.5 initially. Each node is laterally coupled with neighboring nodes on the closed boundary. The connection weight from node i to j, w_{ij}, is 1 and can be modified by T-junctions and local shape information. Each occluding boundary is represented by a pair of boundary segments of the involved regions. A node in a pair competes with the other to be figural temporally. This competition determines the figure-ground segregation. Here the critical point is that each occluding boundary has to be represented using a pair before we solve the figure-ground segregation problem; otherwise, a combinatorial search would be inherit in order to cover all the possible configurations. Fig. 2 shows an example. In the example, nodes 1 and 5 form a boundary pair, where node 1 belongs to the white region, or the background region and node 5 belongs to the black region, or the figural region.

Node i updates its status by:

$$\begin{aligned} \tau \frac{dP_i}{dt} \quad = \quad & \mu_L \sum_{k \in N(i)} w_{ki}(P_k - P_i) \\ & +\mu_J (1-P_i) \sum_{l \in J(i)} H(Q_{li}) \\ & +\mu_B (1-P_i) exp(-B_i/K_B) \end{aligned} \tag{1}$$

Here $N(i)$ is the set of neighboring nodes of i, and μ_L, μ_J, and μ_B are parameters to determine the influences from lateral connections, junctions, and bias. $J(i)$ is the set of junctions that are associated with i and Q_{li} is the junction strength of node i of junction l. $H(x)$ is given by:

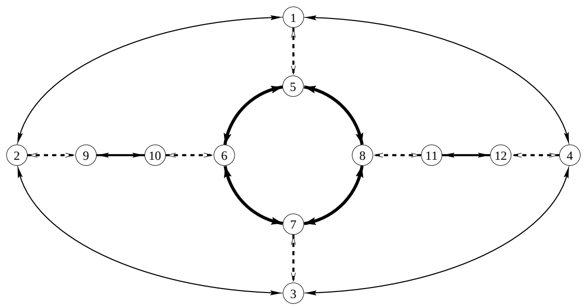


Figure 2: The figure-ground segregation network architecture for Fig. 1(a). Nodes 1, 2, 3 and 4 belong to the white region; Nodes 5, 6, 7, and 8 belong to the black region; Nodes 9 and 10, 11 and 12 belong to the left and right gray regions respectively. Solid lines represent excitatory coupling while dashed lines represent inhibitory connections.

$$H(x) = tanh(\beta(x - \theta_J)$$

Here β controls the steepness and θ_J is a threshold.

In (1), the first term on the right reflects the lateral influences. When nodes are strongly coupled, they are more likely to be in the same status, either figure or background. Second term incorporates junction information. In other words, at a T-junction, segments that are more smooth are more likely to be figural. The third term is a bias term, where B_i is the bias introduced to simulate human perception.

After all the nodes are updated, the competition between paired nodes is through normalization based on the assumption that only one of the paired nodes should be figural at a given time. Suppose that j is the corresponding paired node of i, we have:

$$P_i^{(t+1)} = P_i^t/(P_i^t + P_j^t) \quad (2a)$$

$$P_j^{(t+1)} = P_j^t/(P_i^t + P_j^t) \quad (2b)$$

As a dynamic system, this shares some similarities with relaxation labeling techniques [5]. Because the status of a node is only influenced by the nodes in a local neighborhood in the network, as shown in Fig. 2, the figure-ground segregation network defines a Markov random field. This shares some similarities with the Markov random fields proposed by Zhu [14] for perceptual organization. As will be demonstrated later, our model can simulate many perceptual phenomena while the model by Zhu [14] is a generic and theoretical model for shape modeling and perceptual organization.

2.2 Incorporation of Gestalt Rules

Without introducing grouping cues such as T-junctions and preferences, the solution of the network is not well defined. To generate behavior that is consistent with human perception, we incorporate grouping cues and some Gestalt grouping principles. As the network provides a generic model, many other rules can be incorporated in a similar manner.

T-junctions T-junctions provide important cues for determining relative depth [11] [13]. In Williams and Hanson's model [13], T-junctions are imposed as topological constraints. Given a T-junction l, the initial strength for node i that is associated with l is:

$$Q_{li} = \frac{exp(-\alpha_{(i,c(i))}/K_T)}{1/2\sum_{k\in N_J(l)} exp(-\alpha_{(k,c(k))}/K_T)}$$

where K_T is a parameter, $N_J(l)$ is a set of all the nodes associated with junction l, $c(i)$ is the other node in $N_J(l)$ that belongs to the same region as node i, and $\alpha_{(ij)}$ is the angle between segments i and j.

Non-accidentalness Non-accidentalness tries to capture the intrinsic relationships among segments [8]. In our system, an additional connection is introduced to node i if it is aligned well with a node j from the same region and $j \notin N(i)$ initially. The connection weight w_{ij} is a function of distance and angle between the involved ending points. This can also be viewed as virtual junctions, resulting in virtual contours and conversion of a corner into a T-junction if involving nodes become figural. This corresponds to the organization criterion proposed by Geiger et al. [3].

Shape information Shape information plays a central role in Gestalt principles. Shape information is incorporated through enhancing lateral connections. In this paper, we consider local symmetry. Let j and k be the two neighboring nodes of i.

$$\begin{aligned} w_{ij} = 1 + C\, exp(-|\alpha_{ij} - \alpha_{ki}|/K_\alpha)* \\ exp(-(L_j/L_k + L_k/L_j - 2)/K_L)) \end{aligned} \quad (3)$$

Essentially (3) strengthens the lateral connections when the two neighboring segments of i are symmetric. Those nodes are then strongly grouped together according to (1).

Preferences Human perceptual systems often prefer some organizations over the others. In this model, we incorporated a well-known figure-ground segregation principle, called closeness. In other words, the system prefers regions over holes. In current implementation, we set $B_i = 1.0$ if node i is part of a hole and otherwise $B_i = 0$.

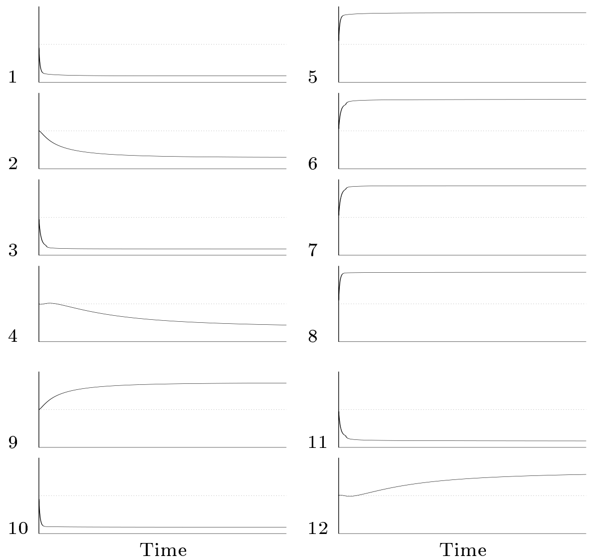


Figure 3: Temporal behavior of each node in the network shown in Fig. 2. Each plot shows the status of the corresponding node with respect to the time. The dashed line is 0.5.

2.3 Temporal Properties of the Network

After we construct the figure-ground segregation network, there are two fundamental questions to be addressed. First we need to demonstrate that the equilibrium state of the system gives a desired solution. Second, we need to show that the system converges to the desired state. Here we demonstrate those using the example shown in Fig. 2. Fig. 3 shows the temporal behavior of the network. First, the system approaches to a stable solution. For figure-ground segregation, we can binarize the status of each node using threshold 0.5. In this case, the system convergences very quickly. In other words, the system outputs the solution in a few iterations. Second, the system generates the correct perception. The black region is occluding other regions while gray regions are occluding the white region. For example, P_5 is close to 1 and thus segment 5 is figural, and P_1 is close to 0 and thus segment 1 is at background.

3 Surface Completion

After the figure-ground segregation is solved, surface completion and shape decomposition can be implemented in a straightforward manner. Currently this stage is implemented through diffusion. Because the ownership of each boundary segment is known, fixed heat sources are generated along occluding boundaries, and the occluding boundaries naturally block diffusion. This method is similar to the one used by Geiger et al. [3] for generating salient surfaces. However, in their approach, because the hypotheses are defined only at junction points, fixed heat sources for diffusion have to be given. On the other hand, in our model, fixed heat sources are generated automatically along the occluding boundaries. In other words, the hypotheses in our system are defined along boundaries, not at junction points.

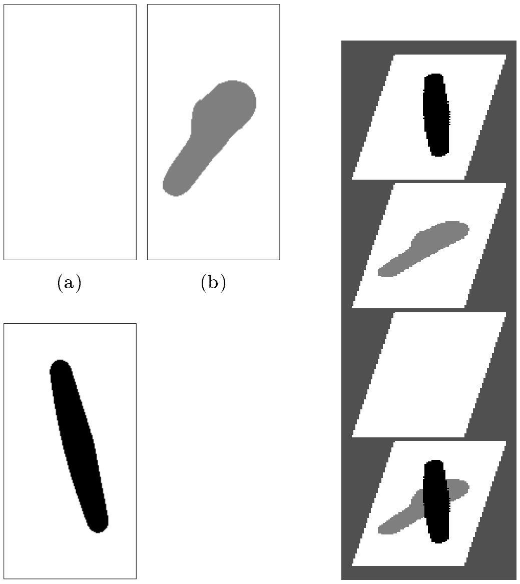
To be more precise, regions from local segmentation are now grouped into diffusion groups based on the average gray value and that if they are occluded by common regions. Segments that belong to one diffusion group are diffused simultaneously. For a figural segment, a buffer with a given radius is generated. Within the buffer, the values are set to 1 for pixels belonging to the region and 0 otherwise. If there is no figural segment in the diffusion group, it is the background, which is always the entire image. Because the figure-ground segregation has been solved, with respect to the diffusion group, only the parts that are being occluded need to be completed. Now the problem becomes a well-defined mathematical problem. We need to solve the heat equation with given boundary conditions. Currently, diffusion is implemented through local diffusion. The results from diffusion are then binarized using threshold 0.5.

Figure 4 shows the results of Fig. 1 after surface completion. Here the two gray regions are grouped together through surface completion because occluded boundaries allow diffusion. Figure 1(d) shows the result using a layered representation to show the relative depth between the surfaces. While the order in this example is well defined, in general the system can handle surfaces that are overlapped with each other, making the order ill-defined.

4 Experimental Results

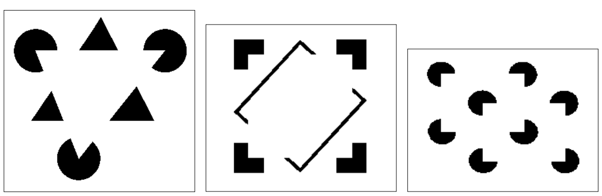
For all the experiments shown in this paper, we use a fixed set of parameters for the figure-ground segregation network. Given an input image, the system automatically constructs the network and establishes the connections based on the rules discussed in Section 2.2.

We first demonstrate that the system can simulate virtual contours and modal completion. Figure 5 shows the input images and Fig. 6 shows the results. The system correctly solves the figure-ground segregation problem and generates the most probable percept. In Fig. 5(b), the rectangular-like frame is tilted, mak-



(a) (b) (c) (d)

Figure 4: Surface completion results for Fig. 1(a). (a) White region. (b) Gray region. (c) Black region. (d) Layered representation by stacking the results according to the relative depth order.

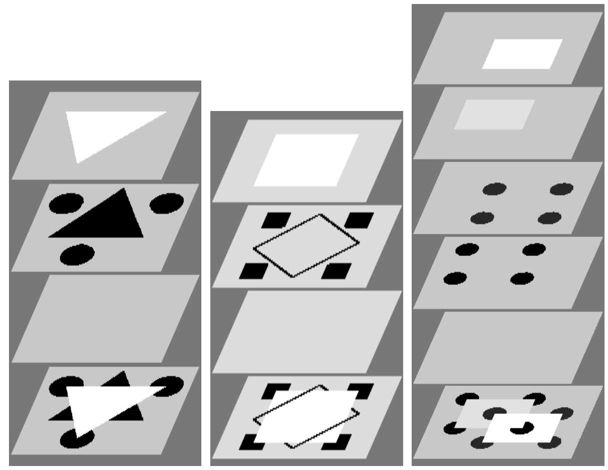


(a) (b) (c)

Figure 5: Images with virtual contours. (a) Kanizsa triangle. (b) Woven square. (c) Double kanizsa.

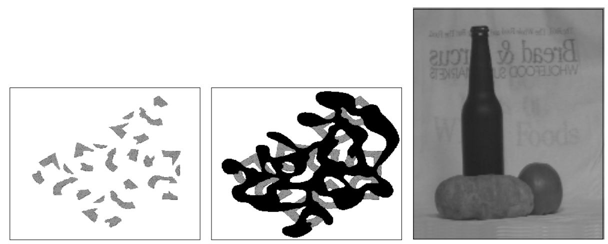
ing the order between the frame and virtual square not well-defined. Our system handles that in the temporal domain. At any given time, the system outputs one of the completed surfaces. Due to this, the system can also handle the case in Fig. 5(c), where the perception is bistable, as the order between the two virtual squares is not defined.

Figure 7 (a) and (b) show well-known examples by Bregman [2]. While the edge elements in both cases are identical, the perception is quite different. In Fig. 7(a), there is no perceptual grouping and parts of B's remain fragmented. However, when occlusion is introduced as in Fig. 7(b), perceptual grouping is evident and fragments of B's are grouped together. These perceptions are consistent with our results shown in Fig. 8



(a) (b) (c)

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



(a) (b) (c)

Figure 7: Bregman and real images. (a) and (b) Examples by Bregman [2]. (c) A grocery store image.

(a) and (b). This is also strong evidence for boundary-pair representation and against edge-based approaches. It shows clearly that grouping plays a very important for recognition. Figure 7(c) shows an image of a grocery store used in [12]. Even though the T-junction at the bottom is locally confusing, our system gives the most plausible result through the lateral influence of the other two strong T-junctions. Without search and parameter tuning, our system gives the optimal solution shown in Fig. 8(c).

5 Conclusions

In this paper we propose a network for perceptual organization using temporal dynamics motivated by the fact that there are both on- and off-cell cells in the visual cortex. We demonstrate that the pair-wise representation can be treated as a surface representation, closing the gap between edge- and surface-based repre-

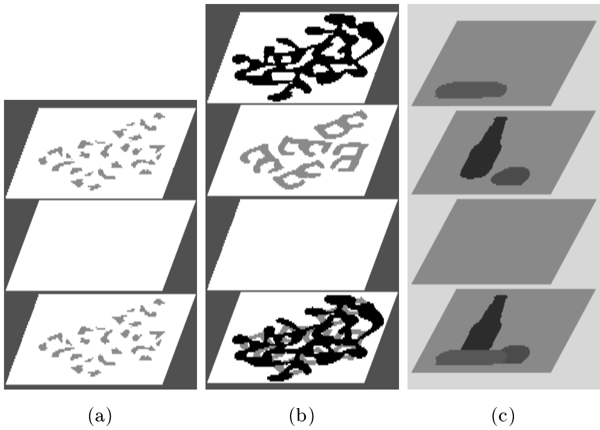


Figure 8: Surface completion results for images shown in Fig. 7.

sentations. The pair-wise representation also provides computational advantage over edge-based representation by overcoming the combinatorial problem common to existing models for perceptual organization [13] [3].

One of the critical advantages of our model is that it allows interactions among different modules dynamically and thus accounts for more context-sensitive behaviors. It is not clear to us whether there exists an energy function for the model. Nodes belonging to one region can be viewed as a Markov random field because the influence is defined locally in the network. However, the inhibition between paired nodes introduced in (2) complicates the system analysis.

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] I. Biederman, "Recognition-by-components: a theory of human image understanding," *Psychological Review*, vol. 94, pp. 115-147, 1987.

[2] A. S. Bregman, "Asking the 'What for' question in auditory perception," In *Perceptual Organization*, M. Kubovy and J R. Pomerantz, eds., Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey, pp. 99-118, 1981.

[3] 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.

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

[5] R. A. Hummel and S. W. Zucker, "On the foundations of relaxation labeling processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, pp. 267-287, 1985.

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

[7] X. Liu and D. L. Wang, "A boundary-pair representation for perception modeling," In *Proceedings of International Joint Conference on Neural Networks*, 1999.

[8] D. G. Lowe, *Perceptual Organization and Visual Recognition*, Kluwer Academic Publishers, Boston, 1985.

[9] 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.

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

[11] 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.

[12] M. Nitzberg, D. Mumford, and T. Shiota, *Filtering, Segmentation and Depth*, Springer-Verlag, New York, 1993.

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

[14] S. C. Zhu, "Embedding Gestalt laws in the Markov random fields," *IEEE Computer Society Workshop on Perceptual Organization in Computer Vision*, 1998.