Object Segmentation and Binding within a Biologically-based Neural Network Model of Depth–from–Occlusion

Paul Sajda and Leif H. Finkel
Department of Bioengineering
University of Pennsylvania
Philadelphia, PA 19104-6392

Abstract

We address the problems of object segmentation and binding within a biologically-based network model capable of determining depth-from-occlusion. In particular, we discuss two subprocesses in our occlusion model most relevant to segmentation and binding: contour binding and direction of figure. We propose that these two subprocesses have intrinsic constraints which allow several underdetermined problems in occlusion processing and object segmentation to be uniquely solved. We present simulations that demonstrate the role these subprocesses play in discriminating objects and stratifying them in depth. Finally, we test our network on "illusory" stimuli, with the network's response indicating the existence of robust psychophysical properties in the system.

1 Introduction

Object discrimination is an intermediate visual process which operates upon the stimulus-based representations of early vision, and which, in turn, serves as the substrate for higher visual processes such as recognition and visual memory. Discrimination has two major components: segmentation and binding. Object segmentation deals with the problem of how separate objects are distinguished. Binding addresses the converse problems of 1) grouping points across the visual field and 2) linking object features–such as color, depth and motion–to create individual objects. These two problems have been studied from the perspectives of both computational neuroscience and machine vision. However, previous studies have not addressed what we consider to be the central issue: how does the visual system define an object–i.e., what constitutes a "thing".

We have developed a model of object discrimination for the specific case of 2D monochromatic planar shapes, where the major basis for segmentation is depth-from-occlusion. A detailed description of the model has been presented elsewhere [1]. Here we concentrate on two of the fundamental intermediate visual processes proposed by the model: *contour binding* and *direction of figure* (determination of inside vs. outside). Together these processes link contours and surfaces into proto-objects–the basic representational unit of an object within our model.

2 Overview of the Occlusion Model

Figure 1 is a schematic illustrating the flow of information within our neural network model. The system is organized into four main stages of processing; *feature extraction*, *segmentation and binding*, *depth processing*, and *completion processing*. These in turn are divided into subprocesses carried out by particular networks in the system. All networks are organized topographically and the operation of the different networks is integrated by extensive reentrant feedback.

In the first stage of the model low-level features are discriminated. Edges, oriented lines, line endings and junctions are identified using units having biologically-based connection fields. The next stage involves segmentation and binding. Here features are grouped to form what we call a *proto-object*. We define a proto-object as a compact, simply connected region surrounded by a closed, piecewise continuous contour and located at a certain depth. We consider a proto-object to be the precursor of an object, since feedback and higher level processing can group one or more proto-objects into a single object. The third stage of the model involves depth processing. We utilize a distributed representation of depth, based on the example of how disparity is represented in the visual cortex. Occlusion cues, identified in earlier stages of the model, act as "forces" for "pushing" proto-objects into different relative depth states. The final stage is responsible for linking occluded and occluding segments of contours. These links are integrated via feed-

0-8186-2855-3/92 $3.00

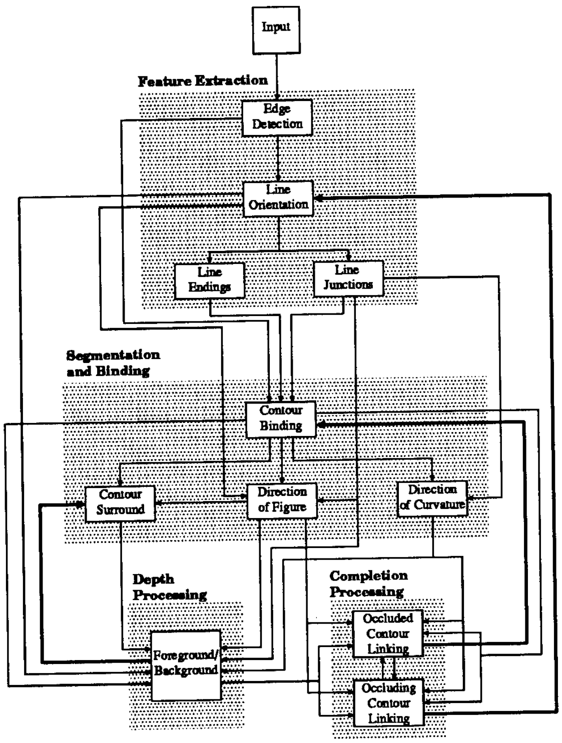


Figure 1: Organization and information flow in our model of depth–from–occlusion.

back pathways and serve to transform proto–objects into objects.

2.1 Contour Binding

Contour binding groups points across the visual field by propagating a unique "tag" to all units responding to the same occluding or object contour. In general, this labeling is governed by the rule that units tend to have the same tag as their neighbors. However, at junctions a problem arises in that it is ambiguous which neighboring units should influence one another. Consider figure 2, which is a simple line drawing with two types of junctions (three and four segment) shown in the insets. Note that for junction a, possible groupings include (A_1OA_2, A_3O), (A_1OA_3, A_2O), and (A_2OA_3, A_1O). Configuration b has even more combinations. We propose that a local binding mechanism links segments at junctions so as to minimize the discontinuity at the junction. This process constrains the problem, producing a solution that conforms to human perception. The constraint is based

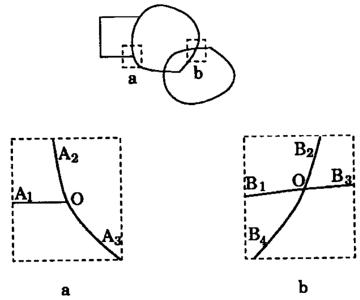


Figure 2: Possible segment pairings at two types of junctions. **a** "T" junction **b** "X" junction.

on simple principles from information theory [4].

The information I_i, at a point on a contour can be expressed as;

$$I_i = log_2(\frac{\pi}{\alpha_i})\ bits \qquad (1)$$

where α_i $(0 < \alpha_i \leq \pi)$ is the angle between line segments formed by joining neighboring points on the contour. The total information I at a junction, found by summing over the i angles created at the junction, is ambiguous since grouping of segments is ambiguous. The system resolves this problem by constraining the tagging so that the configuration which introduces the least amount of information (i.e. minimizes the amount of self–supplied information introduced by the tagging process) is realized. For example, in figure 2b the system would label the segments (B_4O) and (OB_2) with one tag and (B_1O) and (OB_3) with another tag since this configuration results in the minimum information. Given that smaller angles create larger discontinuities, then minimizing self-generate information is analogous to minimizing total discontinuity.

These same arguments apply to junctions of curved contour segments, where minimizing the self-generated information constrains the system to choose the configuration with the minimum curvature. Note that such a constraint might be implemented in biological vision through orientation and directionally specific connectivity. In fact there is anatomical data [5] suggesting that such specific connectivity may exist.

We present simulation results illustrating the role of contour binding within our complete model of depth–from–occlusion. Simulations were carried out using NEXUS, an interactive simulation environment designed for simulating large-scale neural networks [6]. Figure 3A shows a scene presented to the system. Fig-

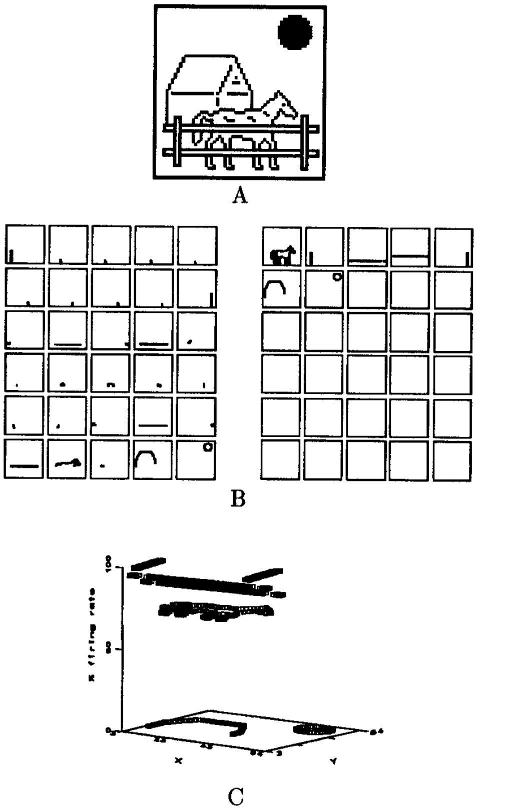


Figure 3: Simulation illustrating discrimination and grouping of occlusion contours. **A** The 64x64 input stimulus. **B** Discrimination and linking of occlusion contours after the first (left) and fifth (right) cycle. **C** Activity in the *foreground* depth network.

ure 3B displays the contour binding tags assigned to different scene elements (on the first and fifth cycles). Each box represents elements with a common tag and the ordering of the boxes is arbitrary. On the first cycle continuous segments of contour are given the same tag. The fifth cycle uses feedback from occluded contour linking (see figure 1) to bind discontinuous proto–objects, separated via occlusion, into objects–thus transforming the proto–objects to objects. Figure 3C plots the firing rate of units in the *foreground* network and illustrates how the system successfully stratifies the fence, horse, house and sun. A competition process utilizes occlusion cues (T-junctions, surround occlusion and concavities) to force points on a contour to different relative depths, while a cooperative process uses information from contour binding

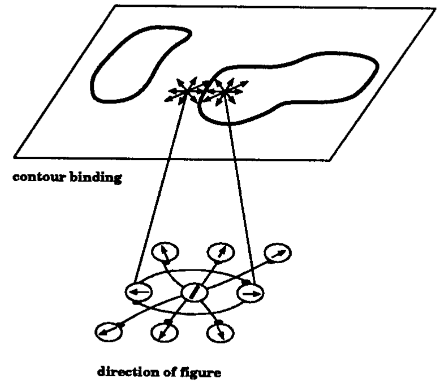


Figure 4: Network architecture for determining direction of figure.

to constrain points belonging to the same occlusion contour to have the same relative depth.

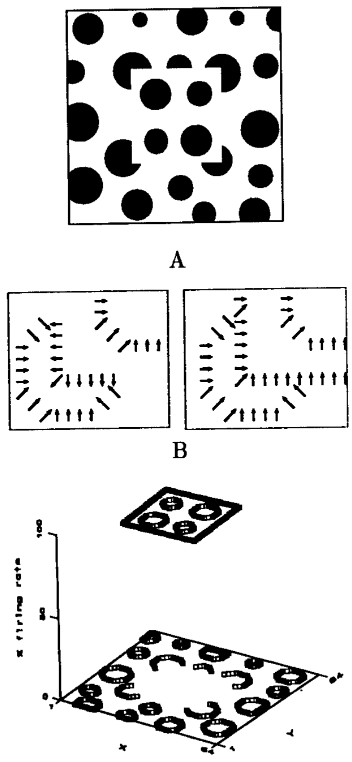
2.2 Direction of Figure

Although contour binding can group individual points on a contour, a mechanism is needed to implicitly identify object surfaces. The task, often formulated as "the figure/ground problem", is to determine which side of the contour is the "inside" of the object and which is the "outside". The problem can also be restated as determining contour "ownership".

The mechanism we employ in our model is shown in figure 4 and is based on the following simple observation. Suppose a unit projects its dendrites in a stellate configuration and that these dendrites are activated by units responding to a contour. Then if a given unit is inside a contour, more of its dendrites will be activated than if it is outside the contour. A winner–take–all interaction between two such units will determine which is more strongly activated, and hence which is the inside of the figure. As shown in figure 6, it is advantageous to limit this competition to the two units which are located at positions directly perpendicular to the local orientation of the contour.

The determination of direction of figure relies on processing within the contour binding network. Each direction of figure unit only considers inputs with the appropriate tag, so that inputs from separate contours in the scene are not confused.

The final simulation shows the ability of the system to generate illusory contours and to use illusory objects in a veridical fashion. The stimulus is adapted from Kanizsa [2], and shows a perceptually vivid, illusory white square in a field of black discs. The illusory



A B C

Figure 5: Simulation of an illusory contour stimulus. **A** The input stimulus. **B** View of direction of figure before and after the illusory contour is generated by the system. **C** *foreground* map activity illustrating the system's determination of relative depth.

square (in figure 5A) appears to be closer to the viewer than the background, and in addition, the four discs which lie inside its borders appear to be pulled forward by the square. This is an example of what we call "occlusion capture", an effect related to Ramachandran's motion and stereo capture phenomenon [3], in which the illusory square has "captured" the discs within its borders, thus pulling them into the foreground.

Figure 5B shows the direction of figure for a portion of the display near the bottom left edge of the illusory square. After the first cycle, the system identifies the direction of figure for the "L"-shaped edge to be toward the interior of the disc, and thus belonging to that proto-object. After three cycles the "L" shaped edge has changed ownership and is now considered part of the illusory square. Finally, figure 5C displays the firing rate in the *foreground* network, showing the relative depth discriminated by the system.

3 Summary

We have presented a neural network model capable of determining depth–from–occlusion, and have focused on two subprocesses within the model related to object segmentation and binding. We have discussed network constraints for binding points on the same object contour. In addition, we have addressed issues for identifying the "ownership" of contours by implicitly defined object surfaces. Our neural network model is capable of segmenting simple images and stratifying objects in depth. The system's response to several classes of illusory stimuli are qualitatively consistent with human perception. Our current research involves extending the model to handle more complex input images and performing quantitative comparisons between the network's performance and human psychophysics.

Acknowledgements

This research was supported by grants from The Office of Naval Research (N00014-90-J-1864), The Whitaker Foundation, and The McDonnell-Pew Program in Cognitive Neuroscience.

References

[1] Finkel, L.H., and Sajda, P., "Object Discrimination based on Depth-from-Occlusion," *Neural Computation* [in press].

[2] Kanizsa, K *Organization in Vision,* Praeger, New York, 1989.

[3] Ramachandran, V.S., "Capture of stereopsis and apparent motion by illusory contours," *Perception and Psychophysics*, Vol. 39, pp. 361-373, 1986.

[4] Resnikoff, H., *The Illusion of Reality,* Springer–Verlag, New York, 1989.

[5] Rockland, K. and Lund, J., "Widespread periodic intrinsic connections in the tree shrew visual cortex," *Science* , Vol. 215, pp. 1532-1534, 1982.

[6] Sajda, P., and Finkel, L.H., "NEXUS: A simulation environment for large-scale neural systems," *Simulation* [in press].