Communicated by Dana Ballard

Object Discrimination Based on Depth-from-Occlusion

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

We present a model of how objects can be visually discriminated based on the extraction of depth-from-occlusion. Object discrimination requires consideration of both the binding problem and the problem of segmentation. We propose that the visual system binds contours and surfaces by identifying "proto-objects"—compact regions bounded by contours. Proto-objects can then be linked into larger structures. The model is simulated by a system of interconnected neural networks. The networks have biologically motivated architectures and utilize a distributed representation of depth. We present simulations that demonstrate three robust psychophysical properties of the system. The networks are able to stratify multiple occluding objects in a complex scene into separate depth planes. They bind the contours and surfaces of occluded objects (for example, if a tree branch partially occludes the moon, the two "half-moons" are bound into a single object). Finally, the model accounts for human perceptions of illusory contour stimuli.

1 Introduction

In order to discriminate objects in the visual world, the nervous system must solve two fundamental problems: binding and segmentation. The binding problem (Barlow 1981) addresses how the attributes of an object—shape, color, motion, depth—are linked to create an individual object. Segmentation deals with the converse problem of how separate objects are distinguished. These two problems have been studied from the perspectives of both computational neuroscience (Marr 1982; Grossberg and Mingolla 1985; T. Poggio et al. 1988; Finkel and Edelman 1989) and machine vision (Guzman 1968; Rosenfeld 1988; Aloimonos and Shulman 1989; Fisher 1989). 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."

Object discrimination occurs at an intermediate stage of the transformation between two-dimensional (2D) image intensity values and visual recognition, and in general, depends on cues from multiple visual modalities. To simplify the problem, we restrict ourselves to discrimination based solely on occlusion relationships. In a typical visual scene, multiple objects may occlude one another. When this occurs, it creates a perceptual dilemma—to which of the two overlapping surfaces does the common border belong? If the border is, in fact, an occlusion border, then it belongs to the occluding object. This identification results in a stratification of the two objects in depth and a de facto discrimination of the objects. Consider the case of a tree branch crossing the face of the moon. We perceive the branch as closer and the moon more distant, but in addition, the two "half-moons" are perceptually linked into one object. The visual system supplies a virtual representation of the occluded contours and surfaces in a process Kanizsa (1979) has called "amodal completion." With this example in mind, we propose that the visual system identifies "proto-objects" and determines which proto-objects, if any, should be linked into objects. For present purposes, a proto-object is defined as a compact region surrounded by a closed, piecewise continuous contour and located at a certain distance from the viewer. The contour can be closed on itself, or more commonly, it can be closed by termination on other contours.

We will demonstrate how a system of interconnected, physiologically based neural networks can identify proto-objects, link them into objects, and stratify the objects in depth. The networks operate, largely in parallel, to carry out the following interdependent processes:

- discriminate edges
- segment and bind contours
- identify proto-objects (i.e., bind contours and surfaces)
- identify possible occlusion boundaries
- stratify occluding objects into different depth planes
- attempt to link proto-objects into objects
- influence earlier steps (e.g., contour binding) by results of later steps (e.g., object linkage).

The constructed networks implement these processes using a relatively small number of neural mechanisms (such as detecting curvature, and determining which surface is inside a closed contour). A few of the mechanisms used are similar to those of previous proposals (Grossberg and Mingolla 1985; Finkel and Edelman 1989; Fisher 1989). But our particular choice of mechanisms is constrained by two considerations. First, we utilize a distributed representation of depth—this is based on the example of how disparity is represented in the visual cortex (G. Poggio et al. 1988; Lehky and Sejnowski 1990). The relative depth of a particular object is represented by the relative activation of corresponding units in a *foreground* and *background* map. Second, as indicated above, we make extensive use of feedback (reentrant) connections from higher level networks to those at lower levels—this is particularly important in linking proto-objects. For example, once a higher level network has determined an occlusion relationship it can modify the way in which an earlier network binds contours to surfaces.

Any model of visual occlusion must be able to explain the perception of illusory (subjective) contours, since these illusions arise from artificially arranged cues to occlusion (Gregory 1972). The proposed model can account for the majority of such illusions. In fact, the ability to link contours in the foreground and background corresponds, respectively, to the processes of modal and amodal completion hypothesized by Kanizsa (1979). The present proposal differs from previous neural models of illusory contour generation (Ullman 1977; Grossberg and Mingolla 1985; von der Heydt et al. 1989; Finkel and Edelman 1989) in that it generates illusory objects—not just the contours. The difference is critical: a network which generates responses to the three sides of the Kanizsa triangle, for example, is not representing a triangle (the object) per se. To represent the triangle it is necessary to link these three contours into a single entity, to know which side of the contour is the inside, to represent the surface of the triangle, to know something about the properties of the surface (its depth, color, texture, etc.), and finally to bind all these attributes into a whole. This is clearly a much more difficult problem. We will describe, however, a simple model for how such a process might be carried out by a set of interconnected neural networks, and present the results of simulations that test the ability of the system on a range of normal and illusory scenes.

2 Implementation

Simulations of the model were conducted using the NEXUS Neural Simulator (Sajda and Finkel 1992). NEXUS is an interactive simulator designed for modeling multiple interconnected neural maps. The simulator allows considerable flexibility in specifying neuronal properties and neural architectures. The present simulations feature an interconnected system composed of 10 different network architectures, each of which contains one or more topographically organized arrays of 64 × 64 units. Two types of neuronal units are used. Standard neuronal units carry out a linear weighted summation of their excitatory and inhibitory inputs, and outputs are determined by a sigmoidal function between voltage and firing rate. NEXUS also allows the use of more complex units called PGN (programmable generalized neural) units that execute arbitrary functions or algorithms. A single PGN unit can emulate the function of a small circuit or assembly of standard units.

PGN units are particularly useful in situations in which an intensive computation is being performed but the anatomical and physiological details of how the operation is performed *in vivo* are unknown. Alterna-

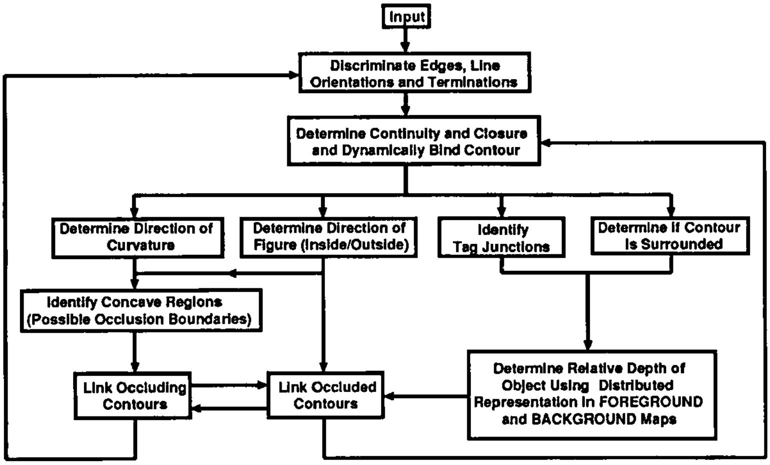

Input
Discriminate Edges, Line Orientations and Terminations
Determine Continuity and Closure and Dynamically Bind Contour
Determine Direction of Curvature
Determine Direction of Figure (Inside/Outside)
Identify Tag Junctions
Determine if Contour is Surrounded
Identify Concave Regions (Possible Occlusion Boundaries)
Link Occluding Contours
Link Occluded Contours
Determine Relative Depth of Object Using Distributed Representation in FOREGROUND and BACKGROUND Maps

Figure 1: Major processing stages in the model. Each process is carried out by one or more networks. Following early visual stages, information flows through two largely parallel pathways—one concerned with identifying and linking occlusion boundaries (left side) and another concerned with stratifying objects in depth (right side). Networks are multiply interconnected and note the presence of the two reentrant feedback pathways.

tively, PGN units can be used to carry out functions in a time-efficient manner; for example, to implement a one-step winner-take-all algorithm. The PGN units used in the present simulations can all be replaced with circuits composed of standard neuronal units, but this incurs a dramatic increase in processing time and memory allocation with minimal changes in functional behavior at the system level.

No learning is involved in the network dynamics. The model is intended to correspond to visual processing during a brief interval (less than 200 msec following stimulus presentation), and the interpretation of even complex scenes requires only a few cycles of network activity. The details of network construction will be described elsewhere; we will focus here on the processes performed and the theoretical issues behind the mechanisms.

3 Construction of the Model

The model consists of a number of stages as indicated in Figure 1. The first stage of early visual processing involves networks specialized for the detection of edges, line orientation, and line terminations (endstopping). As Ramachandran (1987) observed, the visual system must distinguish several different types of edges: we are concerned here with the distinction between edges due to surface discontinuities (transitions between different surfaces) and those due to surface markings (textures, stray lines, etc.). Only the former can be occlusion boundaries. The visual system utilizes several modalities to classify types of edges; we restrict ourselves to a single process carried out by the second processing stage, a network that determines which segments belong to which contours and whether the contours are closed.

When two contours cross each other, forming an "X" junction, there are several possible perceptual interpretations of which arms of the "X" should be joined. Our networks carry out the simple rule that discontinuities should be minimized—i.e., lines and curves should continue as straight (or with as much the same curvature) as possible. Similar assumptions underlie previous models (Ullman 1977), and this notion is in accord with psychophysical findings that discontinuities contain more information than continuous segments (Attneave 1954; Resnikoff 1989). We are thus minimizing the amount of self-generated information.

We employ a simple sequential process to determine whether a contour is closed—each unit on a closed contour requires that at least two of its nearest neighboring units also be on the contour. It is computationally difficult to determine closure in parallel. We speculate that, *in vivo*, the process is carried out by a combination of endstopped units and large-receptive field cells arranged in an architecture similar to that described in Area 17 (Rockland and Lund 1982; Mitchison and Crick 1982; Gilbert and Wiesel 1989). Once closure is determined, it is computationally efficient for the units involved to be identified with a "tag." Several of the higher level processes discussed below require that units responding to the same contour be distinguishable from those responding to different contours. There are several possible physiological mechanisms that could subserve such a tag—one possible mechanism is phase-locked firing (Gray and Singer 1989; Eckhorn et al. 1988). We have implemented this contour binding tag through the use of PGN units (Section 2), which are capable of representing several distinct tags. It must be emphasized, however, that the model is compatible with a number of possible physiological mechanisms.

Closed contours are a necessary condition to identify a proto-object, but sufficiency requires two additional components. As shown in Figure 1, the remaining determinations are carried out in parallel. One stage is concerned with determining on which side of the contour the figure lies, i.e., distinguishing inside from outside. The problem can be alternatively posed as determining which surface "owns" the contour (Koffka 1935; Nakayama and Shimojo 1990). This is a nontrivial problem that, in general, requires global information about the figure. The classic example is the spiral (Minsky and Papert 1969; Sejnowski and Hinton 1987) in

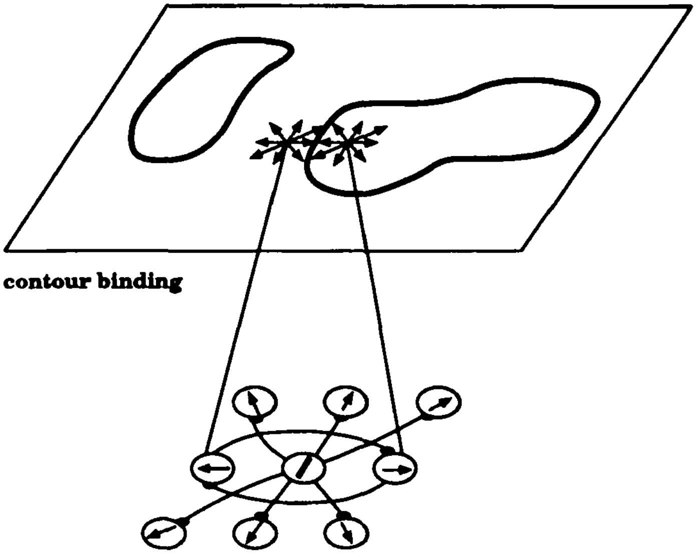


Figure 2: Neural circuit for determining direction of figure (inside vs. outside). Hypothetical visual stimulus consists of two closed contours (bold curves). The central unit of 3 × 3 array (shown below) determines the local orientation of the contour. Surrounding units represent possible directions (indicated by arrows) of the inside of the figure relative to the contour. All surrounding units are inhibited (black circles) except for the two units located perpendicular to local orientation of the contour. Units receive inputs from the contour binding map via dendrites that spread out in a stellate configuration, as indicated by clustered arrows (dendrites extend over long distances in map). Units inside the figure will receive more inputs than those located outside the figure. The two uninhibited units compete in a winner-take-all interaction. Note that inputs from separate objects are not confused due to the tags generated in the contour binding map.

which it is impossible to determine whether a point is inside or outside based on only local information. The mechanism we employ, as shown in Figure 2, is based on the following simple observation. Suppose a unit projects its dendrites in a stellate configuration and that the dendrites are activated by units responding to a contour. Then units located inside a closed contour will receive more activation than units located outside the contour. A winner-take-all interaction between the two units will

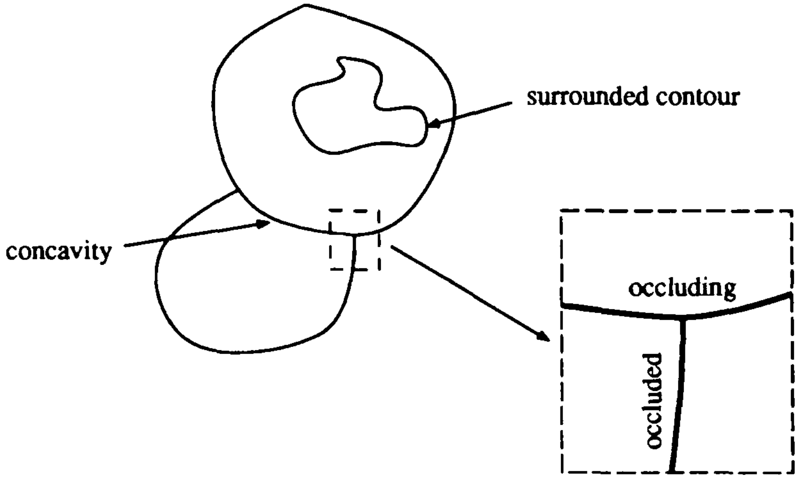


Figure 3: Primary cues for occlusion. Tag junctions (shown in the inset) signal a local discontinuity between occluding and occluded contours. Concave regions and surrounded contours suggest occlusion, but are not as reliable indicators as tag junctions. Additional cues such as accretion/deletion of texture (not considered here) are used *in vivo*.

determine which is more strongly activated, and hence which is inside the figure. As shown in Figure 2, it is advantageous to limit this competition to the two units that are located at positions perpendicular to the local orientation of the contour. As will be shown below (see Figs. 5–7), this network is quite efficient at locating the interior of figures. It also demonstrates deficiencies similar to those of human perception—for example, it cannot distinguish the inside from the outside of a spiral. The mechanism depends on the contour binding carried out above. Each unit only considers inputs with the appropriate tag—in this way, inputs from separate contours in the scene are not confused.

Identification of a proto-object also requires that the relative depth of the surface be determined. This is carried out chiefly through the use of tag junctions. As shown in Figure 3, a tag junction is formed by the termination of an occluded boundary on an occluding boundary. Tag junctions generally correspond to T-junctions in the image, however, they arive from discontinuities in the binding tags and are therfore associated with surface discontinuities as well. Note that tag junctions are identified at an intermediate stage in the sytem (see Fig. 1) and are not constructed directly from end-stopped units in early vision. This accords with the lack of physiological evidence for "junction" detectors in striate cortex.

In this model, tag junctions serve as the major determinant of relative depth. At such junctions, there is a change in the binding (or ownership) of contours, and it is this change which produces the discontinuity in perceived depth. Depth is represented by the relative level of activity in two topographic maps (called *foreground* and *background*). The closest object maximally activates *foreground* units and minimally activates *background* units; the most distant object has the reverse values, and objects located at intermediate depths display intermediate values. The initial state of the two maps is such that all closed contours lie in the background plane. Depth values are then modified at tag junctions—contours corresponding to the head of the "T" are pushed toward the foreground. Since multiple objects can overlap, a contour can be both occluding and occluded—therefore, the relative depth of a contour is determined in a type of push–pull process in which proto-objects are shuffled in depth. The contour binding tag is critical in this process in that all units with the same tag are pushed forward or backward together. (In the more general case of nonplanar objects, the alteration of depth values would depend on position along the contour.)

Tag junctions arise in cases of partial occlusion; however, in some instances, a smaller object may actually lie directly in front of a larger object. In this case, which we call "surround" occlusion, the contour of the occluded object surrounds that of the occluding object. As shown in Figure 1, a separate process determines whether such a surround occlusion is present, and in the same manner as tag junctions, leads to a change in the representation of relative depth. The network mechanism for detecting surround occlusion is almost identical to that discussed above for determining the direction of figure (see Fig. 2). Note that a similar configuration of two concentric contours arises in the case of a "hole." The model is currently being extended to deal with such non-simply connected objects.

These processes—contour binding, determining direction of the figure, and determination of relative depth—define the proto-object. The remainder of the model is concerned with linking proto-objects into objects. The first step in this endeavor is to identify occlusion boundaries. Since occlusion boundaries are concave segments of contours, such segments must be detected (particularly, concave segments bounded by tag junctions). Although many machine vision algorithms exist for determining convexity, we have chosen to use a simple, neurally plausible mechanism: at each point of a contour, the direction of figure is compared to the direction of curvature [which is determined using endstopped units (Dobbins et al. 1987)]. In convex regions, the two directions are the same; in concave regions, the two directions are opposed. A simple AND mechanism can therefore identify the concave segments of the contours.

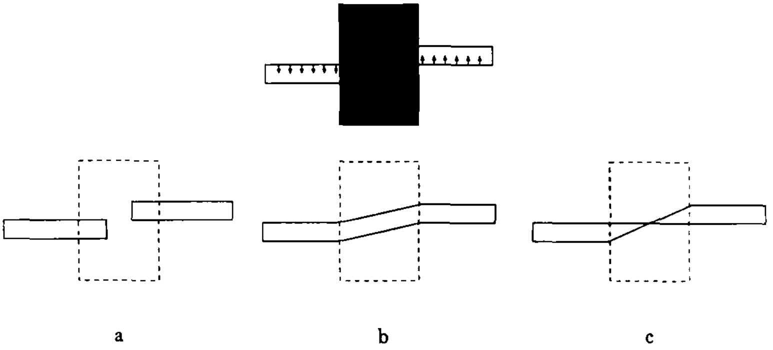


Figure 4: Linking of occluded contours. Three possible perceptual interpretations (below) of an occlusion configuration (above) are shown. Small arrows indicate direction of figure (inside/outside). Collinearity cannot be the sole criterion for linking occluded edges. Consistency in the direction of figure between linked objects rules out perception c.

Once occlusion borders are identified, proto-objects can be linked by trying to extend, complete, or continue occluded segments. Linkage most commonly occurs between proto-objects in the background, i.e., between spatially separated occluded contours. For example, in Figure 3, the occluded contours which terminate at the two tag junctions can be linked to generate a virtual representation of the occluded segment. Since it is impossible to know exactly what the occluded segment looks like, and since it is not actually "perceived," we have chosen not to generate a representation of the occluded segment. Rather, a network link binds together the endpoints of the two tag junctions. In the case where multiple objects are occluded by a single object, the problem of which contours to link can become complex. As shown in Figure 4, one important constraint on this process is that the directions of figure be consistent between the two linked proto-objects.

Another condition in which proto-objects can be linked involves the joining of occluding contours, i.e., of proto-objects in the foreground. This phenomenon occurs in our perception of illusory contours, for example, in the Kanizsa triangle (Kanizsa 1979) or when a gray disc is viewed against a background whose luminance changes in a smooth spatial gradient from black to white (Marr 1982; Shapley and Gordon 1987). In this case, a representation of the actual contour is generated. The conditions for linkage are that the two contours must be smoothly joined by a line or curve, and that the direction of figure be consistent (as in the case of occluded contours above).

The major difference between these two linking or completion processes is that contours generated in the foreground are perceived while those in the background are not. However, the same mechanisms are used in both cases. We have elected to segregate the foreground and background linking processes into separate networks for computational simplicity—it is possible, however, that *in vivo* a single population of units carries out both functions.

Regardless of the implementation, the interaction between ongoing linking processes in the foreground and background is critical. Since these links are self-generated by the system (they do not exist in the physical world), they must be scrutinized to avoid false conjunctions. The most powerful check on these processes is their mutual consistency—an increased certainty of the occluded contour continuation being correct increases the confidence of the occluding contour continuation, and vice versa. For example, in the case of the Kanizsa triangle, the "pac-man"-like figures can be completed to form complete circles by simply continuing the contour of the pac-man. The relative ease of completing the occluded contours, in turn, favors the construction of the illusory contours, which correspond to the continuations of the occluding contours. In fact, we believe that the interaction between these two processes determines the perceptual vividness of the illusion.

The final steps in the process involve a recurrent feedback (or reentry, Finkel and Edelman 1989) from the networks that generate these links back to earlier stages so that the completed contours can be treated as real objects. Note that the occluded contours feedback to the contour binding stage, not to the line discrimination stage, since in this case, the link is virtual, and there is no generated line whose orientation, etc., can be determined. The feedback is particularly important for integrating the outputs of the two parallel paths. For example, once an occluding contour is generated, as in the illusory contours generated in the Kanizsa triangle, it creates a new tag junction (with the circular arc as the "tail" and the illusory contour as the "head" of the "T"). On the next iteration through the system, this tag junction is identified by networks in the other parallel path of the system (see Fig. 1), and is used to stratify the illusory contour in depth.

4 Results of Simulations

4.1 Linking Proto-objects. We present the results of three simulations which illustrate the ability of the system to discriminate objects. Figure 5 shows a visual scene that was presented to the system. The early networks discriminate the edges, lines, terminations, and junctions present in the scene. Figure 5A displays the contour binding tags assigned to different scene elements (on the first and fifth cycle of activity). Each box represents active units with a common tag, different boxes represent different tags, and the ordering of the boxes is arbitrary. Note that on the first cycle of activity, discontinuous segments of contours are given separate tags. These tags are changed by the fifth cycle as a result of feedback from the linking processes.

Figure 5B shows the output of the *direction of figure* network, for a small portion of the input scene (near the horse's head). The direction of the arrows indicates the direction of figure determined by the network. The correct direction of figure is determined in all cases: for the horse's head, and for the horizontal and vertical posts of the fence. Once the direction of figure is identified, occluded contours can be linked (as in Fig. 4), and proto-objects combined into objects. This linkage is what changes the contour binding tags, so that after several cycles (Fig. 5A, right), separate tags are assigned to separate objects—the horse, the gate posts, the house, the sun.

The presence of tag junctions (e.g., between the horse's contour and the fence, between the house and the horse's back) is used by the system to force various objects into different depth planes. The results of this process are displayed in Figure 5C, which plots the firing rate (percent of maximum) of units in the *foreground* network. The system has successfully stratified the fence, horse, house, and sun. The actual depth value determined for each object is somewhat arbitrary, and can vary depending on minor changes in the scene—the system is designed only to achieve the correct relative ordering, not absolute depth. Note that the horizontal and vertical posts of the fence are perceived at different depths—this is because of the tag junctions present between them; in fact, the two surfaces do lie at slightly different depths. In addition, there is no way to determine the relative depth of the two objects in the background, the house and the sun, because they bear no occlusion relationship to each other. Again, this conforms to human perceptions, e.g., the sun and the moon appear about the same distance away. The system thus appears to process occlusion information in a manner similar to human perception.

4.2 Gestalt Psychology of a Network. The system also displays a response consistent with human responses to a number of illusory stimuli. Figure 6 shows a stimulus, adapted from an example of Kanizsa (1979), which shows that preservation of local continuity in contours is more powerful than global symmetry in perception (this is contrary to classical Gestalt theory—e.g., Koffka 1935). As shown in the middle panels, there are two possible perceptual interpretations of the contours—on the left, the two figures respect local continuity (this is the dominant human perception); on the right, the figures respect global symmetry.

Figure 6A shows the contour binding tags assigned by the system to this stimulus, and Figure 6B shows the direction of figure that was determined. Both results indicate that the network makes the same perceptual interpretation as a human observer.

4.3 Occlusion Capture. 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, again, adapted from Kanizsa (1979), and shows a perceptually vivid, illusory white square in a field of black discs. The illusory square appears to be closer to the viewer than the background, and, in addition, the four discs that lie inside its borders also appear closer than the background (some viewers perceive the four internal discs to be even closer than the illusory square). This is an example of what we call "occlusion capture," an effect related to the capture phenomena involving motion, stereopsis, and other submodalities (Ramachandran and Cavanaugh 1985; Ramachandran 1986). In this case, the illusory square has "captured" the discs within its borders and they are thus pulled into the foreground.

Figure 7A shows the contour binding tags after one (left) and three (right) cycles of activity. Each disc receives a separate tag. After the responses to illusory square are generated, the illusory contours are fed back to the contour binding network and given a common tag. Note that the edges of the discs occluded by the illusory square are now given the same tag as the square, not the same tags as the discs.

The change in "ownership" of the occluded edges of the discs is the critical step in defining the illusory square as an object. For example, Figure 7B shows the output of the *direction of figure* network after one and three cycles of activity. The large display shows that every disc is identified as an object with the inside of the disc correctly labeled in each case. The two insets focus on a portion of the display near the bottom left edge of the illusory square. At first, the system identifies the "L"-shaped angular edge as belonging to the disc, and thus the direction of figure arrows point "inward." After three cycles of activity, this same "L"-shaped edge is identified as belonging to the illusory square, and thus the arrows now point toward the inside of the square, rather than the inside of the disc. This change in the ownership of the edge results from the discrimination of occlusion—the edge has been determined to

Figure 5: *Facing page.* Object discrimination and stratification in depth. Top panel shows a 64 × 64 input stimulus presented to the system. (A) Spatial histogram of the contour binding tags (each box shows units with common tag, different boxes represent different tags, and the order of the boxes is arbitrary). Initial tags shown on left; tags after five iterations shown on right. Note that linking of occluded contours has transformed proto-objects into objects. (B) Magnified view of a local section of the direction of figure network corresponding to portion of the image near horse's nose and crossing fence posts. Arrows indicate direction of inside of proto-objects as determined by network. (C) Relative depth of objects in scene as determined by the system. Plot of activity (% of maximum) of units in the foreground network after five iterations. Points with higher activity are "perceived" as being relatively closer to the viewer.

be an occlusion border. The interconnected processing of the system then results in a change in the direction of figure and of the continuity tags associated with this edge. The illusory square is perceived as an *object*. Its four contours are bound together, the contours are bound to the internal surface, and the properties of the surface are identified.

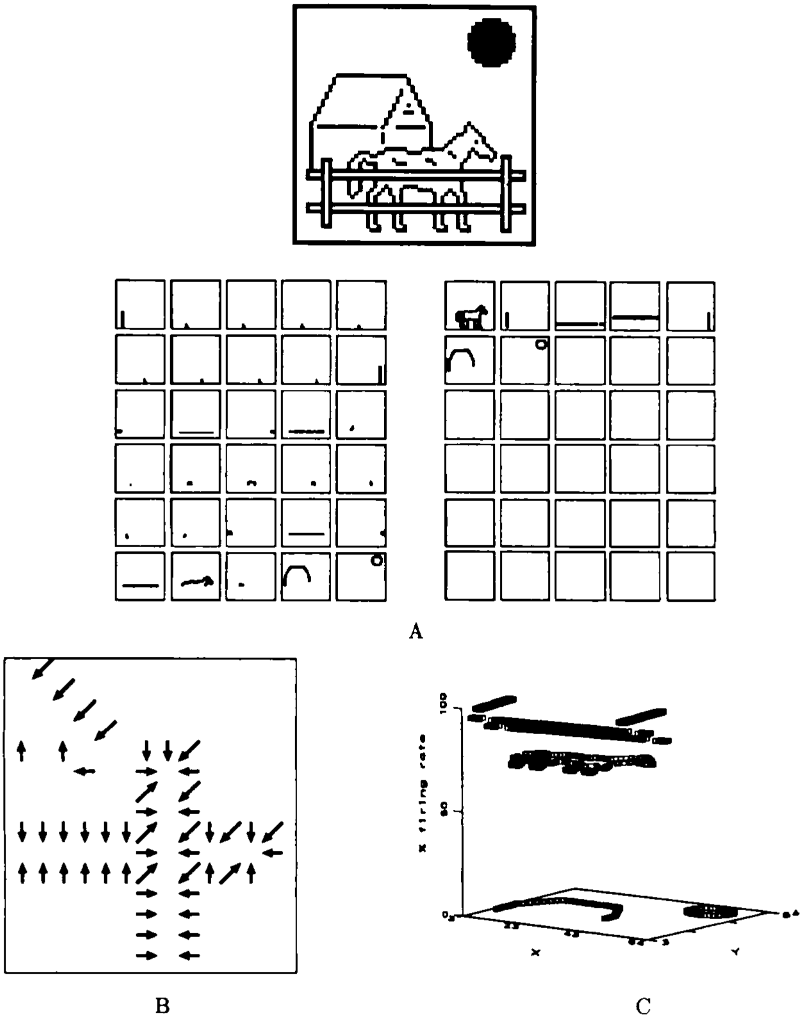


Figure 7C displays the firing rate of units in the *foreground* map (as in 5C), thus showing the relative depths discriminated by the system. The discs are placed in the background, the illusory square and the four discs within its borders are located in the foreground. In this case, the depth cue which forces the internal discs to the foreground is not due to tag junctions, but rather to surround occlusion (see Figure 3). Once the illusory square is generated, the contours of the discs inside the square are surrounded by that of the square. The fact that the contour is "illusory" is irrelevant; once responses are generated in the networks responsible for linking occluding contours and are then fed back to earlier networks, they are indistinguishable from responses to real contours in the periphery. Thus the system demonstrates occlusion capture corresponding to human perceptions of this stimulus.

5 Discussion

In most visual scenes, the majority of objects are partially occluded. Our seamless perception of the world depends upon an ability to complete or link the spatially separated, non-occluded portions of an object. We have used the idea that the visual system identifies proto-objects (which may or may not be objects) and then attempts to link these proto-objects into larger structures. This linking process is most apparent in the perception of illusory contours, and our model can account for a wide range of these illusions.

This model builds upon previous neural, psychological, and machine vision studies. Several models of illusory contour generation (Ullman 1977; Peterhans and von der Heydt 1989; Finkel and Edelman 1989) have used related mechanisms to check for collinearity and to generate the illusory contours. Our model differs at a more fundamental level—we are concerned with objects not just contours. To define an object, surfaces must also be considered. For example, in a simple line drawing, we perceive an interior surface despite the fact that no surface properties are indicated. Thus, the model must be capable of characterizing a surface—and it does so, in a rudimentary manner, by determining the direction of figure and relative depth. Nakayama and Shimojo (1990) have approached the problem of surface representation from a similar viewpoint. They discuss how contours and surfaces become associated, how T-junctions serve to stratify objects in depth, and how occluded surfaces are amodally completed. Nakayama's analysis concentrates on the external "ecological" constraints on perception. In addition to these Gibsonian constraints, we emphasize the importance of *internal* constraints imposed by physiological mechanisms and neural architectures. Nakayama has also explored the interactions between occlusion and surface attributes. A more complete model must consider such surface properties such as color, brightness, texture, and surface orientation. The examination of

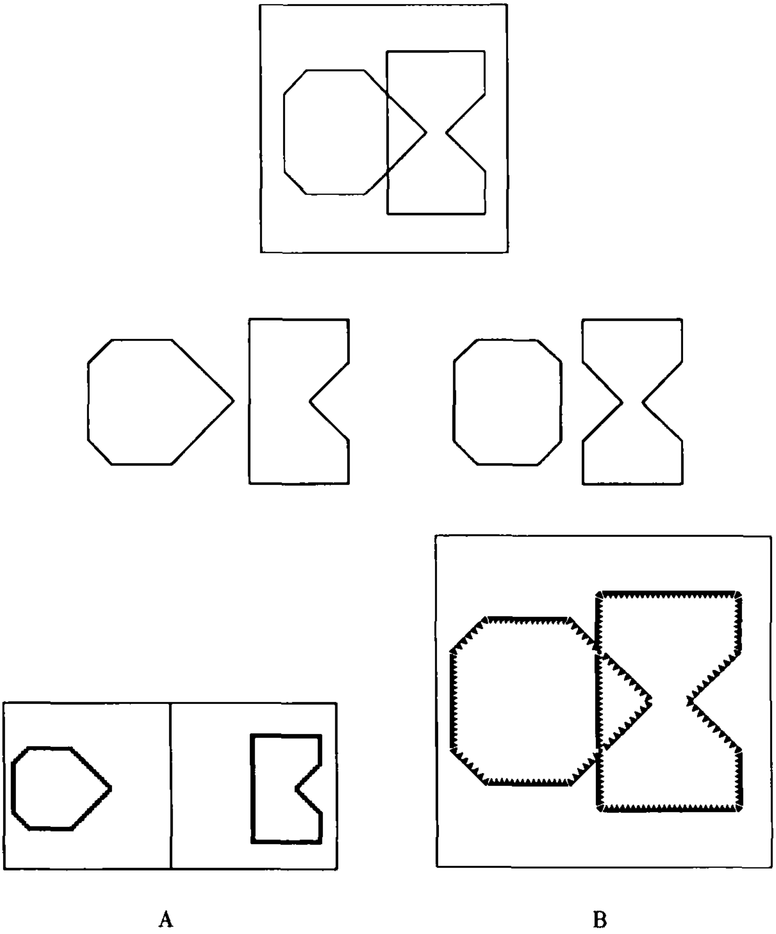


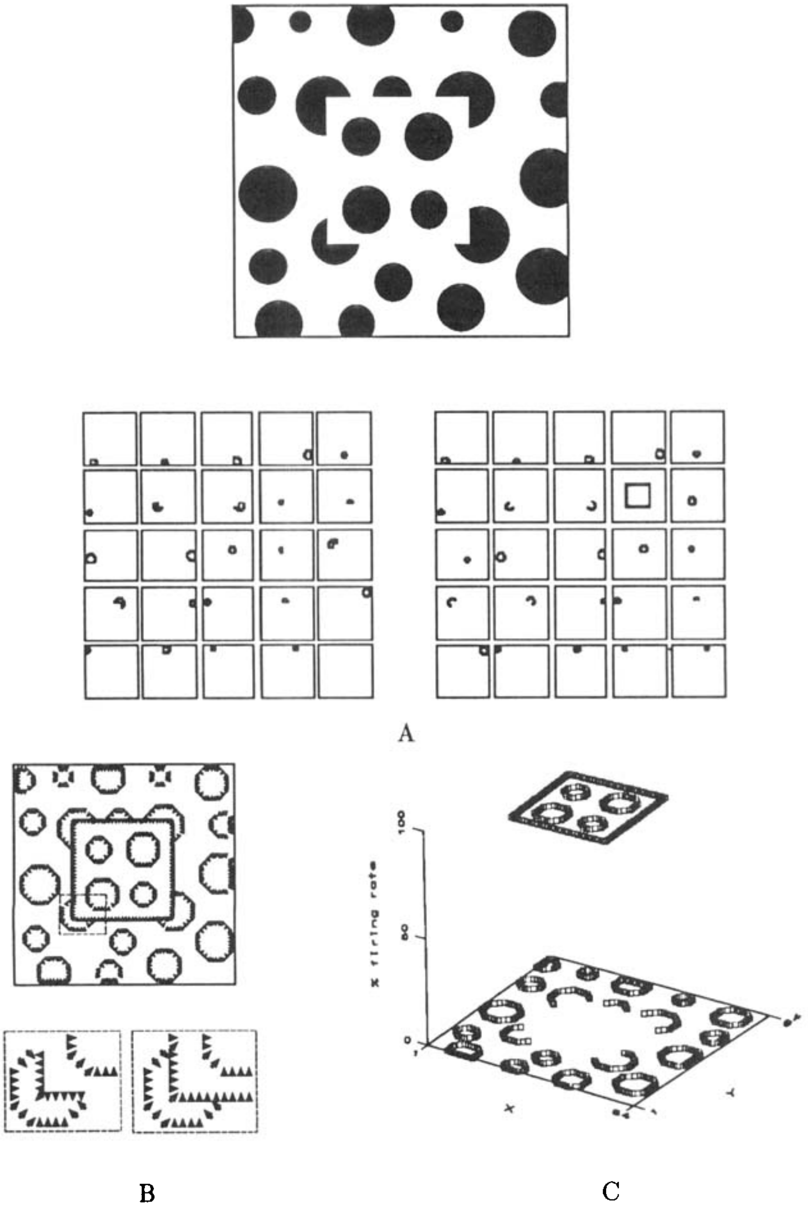
Figure 6: Minimization of ambiguous discontinuities. Upper panel shows an ambiguous stimulus (adapted from Kanizsa 1979), two possible perceptual interpretations of which are shown below. The interpretation on the left is dominant for humans, despite the figural symmetry of the segmentation on the right. Stimulus was presented to the system, results shown after three iterations. (A) Spatial histogram showing the contour binding patterns (as in 5A). The network segments the figures in the same manner as human perception. (B) Determination of direction of figure confirms network interpretation (note at junction points, direction of figure is indeterminate).

how surface features might interact with contour boundaries has been pioneered by Grossberg (1987). Finally, in some regards, our model constitutes the first step of a "bottom-up" model of object perception (Kanizsa 1979; Biederman 1987). It is interesting that regardless of one's orientation (bottom-up or top-down) the constraints of the physical problem result in certain similarities of solution as witnessed by the analogies present with AI based models (Fisher 1989).

One of the most speculative aspects of the model is the use of tags to identify elements as belonging to the same object. Tags linking units responding to the same contour are used to determine the direction of figure and to change the perceived depth of the entire contour based on occlusion relationships detected at isolated points (the tag junctions). It is possible to derive alternative mechanisms for these processes that do not depend on the use of tags, but they are conceptually inelegant and computationally unwieldy. Our model offers no insight as to the biophysical basis of such a tag. However, the model does suggest that there should be a relatively small number of tags, on the order of 10, since this number corresponds to the number of objects that can be simultaneously discriminated. This constraint is consistent with several possible mechanisms: tags represented by different oscillation frequencies, tags represented by different phases of firing, or tags represented by firing within discrete time windows (e.g., the first 10 msec of each 50 msec interval). The number of distinct tags generated by these various mechanisms may depend on the integration time of the neuron, or possibly on the time constant of a synaptic switch, such as the NMDA receptor.

At the outset, we discussed the importance of both binding and segmentation for visual object discrimination. Our model has largely dealt with the segmentation problem, however, the two problems are not entirely independent. For example, the association of a depth value with the object discriminated is, in essence, an example of the binding of an attribute to an object. Consideration of additional attributes makes the

Figure 7: *Facing page.* Occlusion capture. Upper panel shows stimulus (adapted from Kanizsa 1979) in which we perceive a white illusory square. Note that the four black discs inside the illusory square appear closer than the background. A 64 × 64 discrete version of stimulus was presented to the network. (A) Spatial histogram (as in 5A) of the initial and final (after three iterations) contour binding tags. Note that the illusory square is bound as an object. (B) Direction of figure determined by the system. Insets show a magnified view of the initial (left) and final (right) direction of figure (region of magnification is indicated). Note that the direction of figure of the "mouth" of the pac-man flips once the illusory contour is generated. (C) Activity in the *foreground* network (% of maximum) demonstrates network stratification of objects in relative depth. The illusory square has "captured" the background texture.


A
100
50
0
% firing rate
x
y
64
B
C

problem more complex, but it also aids in the discrimination of separate objects (Damasio 1989; Crick and Koch 1990). For example, we have only considered static visual scenes, but one of the major cues to the linking process is common motion of proto-objects. During development, common motion may, in fact, play the largest role in establishing our concept of what is an object (Termine et al. 1987).

Object definition also clearly depends on higher cognitive processes such as attention, context and categorization (Rosch and Lloyd 1978). There is abundant evidence that "top-down" processes can influence the discrimination of figure/ground as well as the perception of illusory figures (Gregory 1972). The examples considered here (e.g., Figs. 5–7) represent extended visual scenes, and perception of these stimuli would require multiple shifts of gaze and/or attention. The representation of such a scene in intermediate vision is thus a more dynamic entity than portrayed here. The processes we have proposed are rapid (all occur in several cycles of iteration), and thus might be ascribed to preattentive perception. However, such preattentive processing sets the stage for directed attention because it defines segmented *objects* localized to particular spatial locations. Furthermore, the process of binding contours, surfaces, and surface features may be restricted to one or two limited spatial regions at any one time. Thus, feature binding may be a substrate rather than a result of the attentional process.

We have implicitly assumed that object discrimination is a necessary precursor to object recognition. Ullman (1989) has developed a model of recognition that demonstrates that this need not *logically* be the case. The question of whether you have to know that something is a "thing" before you can recognize what kind of thing it is remains to be determined through psychophysical experiment. It is appealing, however, to view object discrimination as the function of intermediate vision, i.e., those processes carried out by the multiple extrastriate visual areas. In this view, each cortical module develops invariant representations of aspects of the visual scene (motion, color, texture, depth) and the operations of these modules are dynamically linked. The consistent representations developed in intermediate vision then serve as the substrate for higher level cognitive processes.

In conclusion, we have shown that one can build a self-contained system for discriminating objects based on occlusion relationships. The model is successful at stratifying simple visual scenes, for linking the representations of occluded objects, and at generating responses to illusory objects in a manner consistent with human perceptual responses. The model uses neural circuits that are biologically based, and conforms to general neural principles, such as the use of a distributed representation for depth. The system can be tested in psychophysical paradigms and the results compared to human and animal results. In this manner, a computational model that is designed based on physiological data and tested in comparison to psychophysical data offers a powerful paradigm for bridging the gap between neuroscience and perception.

Note Added in Proof The recent findings of dynamic changes in receptive field structure in striate cortical neurons by Gilbert and Wiesel (1992) indicates that long-range connections undergo context-dependent changes in efficacy. Such a mechanism may provide the biological basis for the direction of figure and linkage mechanisms proposed here. [Gilbert, C. D., and Wiesel, T. N. 1992. Receptive field dynamics in adult primary visual cortex. *Nature* **356**, 150–152.]

Acknowledgments

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

References

Aloimonos, J., and Shulman, D. 1989. *Integration of Visual Modules*. New York, Academic Press.

Attneave, F. 1954. Some informational aspects of visual perception. *Psych. Rev.* **61**, 183–193.

Barlow, H. B. 1981. Critical limiting factors in the design of the eye and visual cortex. *Proc. R. Soc. (London)* **B212**, 1–34.

Biederman, I. 1987. Recognition by components: A theory of human image understanding. *Psych. Rev.* **94**, 115–147.

Crick, F., and Koch, C. 1990. Towards a neurobiological theory of consciousness. *Semin. Neurosci.* **2**, 263–275.

Damasio, A. R. 1989. The brain binds entities and events by multiregional activation from convergence zones. *Neural Comp.* **1**, 1223–1232.

Dobbins, A. S., Zucker, S. W., and Cynader, M. S. 1987. Endstopping in the visual cortex as a neural substrate for calculating curvature. *Nature (London)*, **329**, 438–441.

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., and Reitboeck, H. 1988. Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biol. Cybernet.* **60**, 121–130.

Finkel, L., and Edelman, G. 1989. Integration of distributed cortical systems by reentry: A computer simulation of interactive functionally segregated visual areas. *J. Neurosci.* **9**, 3188–3208.

Fisher, R. B. 1989. *From Objects to Surfaces*. John Wiley & Sons, New York.

Gilbert, C. D., and Wiesel, T. N. 1989. Columnar specificity of intrinsic connections in cat visual cortex. *J. Neurosci.* **9**, 2432–2442.

Gray, C. M., and Singer, W. 1989. Neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 1698–1702.

Gregory, R. L. 1972. Cognitive contours. *Nature (London)* **238**, 51–52.

Grossberg, S. 1987. Cortical dynamics of three-dimensional form, color, and brightness perception. I: Monocular theory. *Percept. Psychophys.* **41**, 87–116.

Grossberg, S., and Mingolla, E. 1985. Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychol. Rev.* **92**, 173–211.

Guzman, A. 1968. Decomposition of a visual scene into three-dimensional bodies. *Fall Joint Comput. Conf.* **1968**, 291–304.

Kanizsa, G. 1979. *Organization in Vision.* Praeger, New York.

Koffka, K. 1935. *Principles of Gestalt Psychology.* Harcourt, Brace, New York.

Konig, P., and Schillen, T. 1991. Stimulus-dependent assembly formation of oscillatory responses: I. Synchronization. *Neural Comp.* **3**, 155–166.

Lehky, S., and Sejnowski, T. 1990. Neural model of stereoacuity and depth interpolation based on distributed representation of stereo disparity. *J. Neurosci.* **7**, 2281–2299.

Livingstone, M. S., and Hubel, D. 1988. Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science* **240**, 740–749.

Marr, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* W. H. Freeman, San Francisco.

Minsky, M., and Papert, S. 1969. *Perceptrons.* The MIT Press, Cambridge, MA.

Mitchison, G., and Crick, F. 1982. Long axons within the striate cortex: Their distribution, orientation, and patterns of connections. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3661–3665.

Nakayama, K., and Shimojo, S. 1990. Toward a neural understanding of visual surface representation. *Cold Spring Harbor Symp. Quant. Biol.* **LV**, 911–924.

Peterhans, E., and von der Heydt, R. 1989. Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. *J. Neurosci.* **9**, 1749–1763.

Poggio, G. F., Gonzalez, F., and Krause, F. 1988. Stereoscopic mechanisms in monkey visual cortex: Binocular correlation and disparity selectivity. *J. Neurosci.* **8**, 4531–4550.

Poggio, T., Gamble, E. B., and Little, J. J. 1988. Parallel integration of vision modules. *Science* **242**, 436–440.

Ramachandran, V. S. 1987. Visual perception of surfaces: A biological theory. In *The Perception of Illusory Contours,* S. Petry and G. E. Meyer, eds., pp. 93–108. Springer-Verlag, New York.

Ramachandran, V. S. 1986. Capture of stereopsis and apparent motion by illusory contours. *Percept. Psychophys.* **39**, 361–373.

Ramachandran, V. S., and Cavanaugh, P. 1985. Subjective contours capture stereopsis. *Nature (London)* **317**, 527–530.

Resnikoff, H. L. 1989. *The Illusion of Reality.* Springer-Verlag, New York.

Rockland, K. S., and Lund, J. S. 1982. Widespread periodic intrinsic connections in the tree shrew visual cortex. *Science* **215**, 1532–1534.

Rosch, E., and Lloyd, B. B. 1978. *Cognition and Categorization.* Lawrence Erlbaum, Hillsdale, NJ.

Rosenfeld, A. 1988. Computer vision. *Adv. Comput.* **27**, 265–308.

Sajda, P., and Finkel, L. 1990. *NEXUS: A neural simulation environment.* University of Pennsylvania Tech. Rep.

Sajda, P., and Finkel, L. 1992. NEXUS: A simulation environment for large-scale neural systems. *Simulation*, in press.

Sejnowski, T., and Hinton, G. 1987. Separating figure from ground with a Boltzmann machine. In *Vision, Brain and Cooperative Computation*, M. Arbib and A. Hanson, eds., pp. 703–724. The MIT Press, Cambridge, MA.

Shapley, R., and Gordon, J. 1987. The existence of interpolated illusory contours depends on contrast and spatial separation. In *The Perception of Illusory Contours*, S. Petry and G. E. Meyer, eds., pp. 109–115. Springer-Verlag, New York.

Termine, N., Hrynick, T., Kestenbaum, T., Gleitman, H., and Spelke, E. S. 1987. Perceptual completion of surfaces in infancy. *J. Exp. Psychol. Human Percept.* **13**, 524–532.

Ullman, S. 1989. Aligning pictorial descriptions: An approach to object recognition. *Cognition* **32**, 193–254.

Ullman, S. 1977. Filling-in the gaps: The shape of subjective contours and a model for their generation. *Biol. Cybernet.* **25**, 1–6.

von der Heydt, R., and Peterhans, E. 1989. Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. *J. Neurosci.* **9**, 1731–1748.

Received 22 November 1991; accepted 6 April 1992.