Object Representation by Cores: Identifying and Representing Primitive Spatial Regions

TR94-048 (Version 3)
October, 1994

Christina A. Burbeck†¶*
Stephen M. Pizer†*

Medical Image Display and Analysis Group
†Department of Computer Science
¶Department of Psychology
The University of North Carolina
Chapel Hill, NC 27599-3175

The research reported herein was carried out with the partial support of NIH grant P01 CA 47982 and AFOSR grant F49620-92-J-085.

To appear *Vision Research*.

*Christina Burbeck and Stephen Pizer contributed equally to this research. Dr. Pizer has emphasized the computational aspects of the model and Dr. Burbeck its perceptual aspects.

Abstract

We propose a model of the spatial visual processes underlying the identification and representation of the shape of primitive spatial regions. We propose that a region's boundaries are sensed at multiple scales by boundariness detectors that give graded responses, that stimulated boundariness detectors of similar scale, σ, connect to one another across a distance that is proportional to their scale, and that they connect via cores, where a core encodes the middles and widths of the region and hence is a trace in (x,y,σ), i.e., 3D scale space.

INTRODUCTION

One of the more impressive feats that the human visual system performs is the identification of individual objects from the continuous distribution of light that falls on the retina. To accomplish this task, the observer uses information from the image to identify regions of interest on the basis of spatial changes in luminance, color, texture, motion, etc. He also interprets information from the image on the basis of prior experience to infer more complex shape properties: linking adjacent regions, linking regions that are not adjacent, and further segregating some regions into more subtle parts to create the best correspondence with representations that the observer has in memory. This is the process of object representation. The focus of the research reported here is the identification and representation of simple regions which can serve as a basis for object representation.

Specifically, we seek a process for identifying and representing primitive regions that satisfies the following requirements:

- the process must be physiologically plausible and justified by psychophysical results;
- the process must not require detailed prior knowledge about figural shape or location;
- the numbers of neural connections required must be appropriately economical;
- the resulting representation of each region must provide ready access to basic perceptual properties of shape; and
- the representation must be sufficiently rich and flexible to accommodate application of probabilities derived from prior experience.

The process of identifying primitive regions naturally begins with consideration of the sharp transitions in the visual scene: transitions of luminance, color, texture, velocity, etc. Single cell physiology of visual cortex supports the importance of such transitions. Hubel and Wiesel's important discovery of edge detectors [1968] was followed by discoveries of neurons sensitive to spatial change in other image properties, e.g., texture and motion [Van Essen, DeYoe, Olavarria, Knierim, Sari, Fox & Julesz, 1989; Nothdurft & Li, 1985; Nothdurft, Gallant & Van Essen, 1992; Sáry, Vogels & Orban, 1993]. We call all such detectors "boundariness" detectors. This terminology is intended to capture the idea that a particular detector does not by itself indicate an edge location, signaling instead a degree of stimulation or "boundariness". The collection of responses from many such detectors conveys the edge information. Although boundariness detectors can receive their inputs from a variety of sources, we assume that they share the common property of selectivity for spatial scale and orientation. Our research thus far has focused on luminance boundaries, but the ideas apply equally well to other types of boundaries.

How is information from boundariness detectors used to separate a spatial region from its background and to infer and represent its shape? Many current models of shape assume that objects are represented in terms of their component parts. These models typically fall into one of two classes. One class of models focuses on identifying defining regularities that permit economical descriptions of the components, e.g., generalized cylinders and geons [Marr & Nishihara, 1978; Biederman, 1987]. These models begin with an assumption about what the component shapes are and consequently result in a representation that is only an approximation. These models may be adequate for supporting object classification, but they are not adequate to support the subtle discriminations that observers can make. These models also provide no means of computing the components from an image and therefore cannot be implemented and rigorously tested. Finally, as will be argued later, they do not provide a representation that is rich enough to support the application of prior world knowledge.

Another type of model that aims at a component-based representation focuses on segmenting the boundaries in the image into components, typically by identifying boundary regions of high curvature [e.g., Hoffman & Richards, 1984; Beusmans, Hoffman & Bennett, 1987; Leyton, 1992]. These models have provided no means of representing the components once found, and thus are not yet models of shape representation.

The model that we propose addresses the major problems raised by these two classes of models and goes well beyond them: identifying the components of an object from computations on the image, creating a representation of those components that captures the perceptual information that the observer has available, and capturing some larger- and smaller-scale shape properties as well.

EXPERIMENTAL FOUNDATIONS

To segregate a region from its surroundings, the boundaries that define the region must be identified as being linked to one another. We base our work on the idea that a region's boundaries must be linked across the region itself. Tracking around the edge is not sufficient: it does not yield a representation of the region's shape, e.g., its width, curvature, or changes in either of these. Given the desire to create a representation that captures shape as the human observer sees it, what should be the basis for the connections between boundaries?

Knowing the shape of a region is equivalent to knowing the relative locations of its boundaries, and substantial experimental work has been done on how the human observer encodes such spatial relations. These results serve as the foundation of our model.

We know, and have known for more than 100 years, that — to a first approximation — the accuracy with which a human observer can judge the relative locations of two features scales with the separation between the features, i.e., Weber's law for size holds. Thus, for example, the threshold for judging the length of a line scales with its length, separation discrimination thresholds scale with separation, and bisection thresholds scale with the width being bisected. [For a nice historical account of studies of such judgments, see Wolfe, 1923; for a more recent review see Burbeck, 1991.] Thus, if a region's shape is encoded as the spatial relations between boundaries, then the accuracy with which a shape can be perceived will be constant relative to its size. In other words, the perceived shape will be zoom invariant (over an appropriate range).

More recent experimental results suggest a much stronger prediction. In a recent study, Burbeck and Hadden [1993] used a background line as a probe to investigate the area over which information is integrated in a separation discrimination task. Fig. 1a shows the stimulus. The separation discrimination targets were a pair of parallel lines. The background line was parallel to the targets and presented outside the target separation. The distance to the background line was varied. Typical data from this three-line task are shown in Fig. 1b. The effect of the background line depends both on its distance from the targets and on the mean separation between the targets. The perceived separation of the target lines increases whenever the distance to the background line is less than the mean separation between the targets.

This result suggests that the range over which position information is gathered (called the position integration area in that study) scales with the separation between the targets. That inference is further supported by their finding that the increase in this area with separation is sufficient to account for the corresponding increase in the separation discrimination threshold.

a)

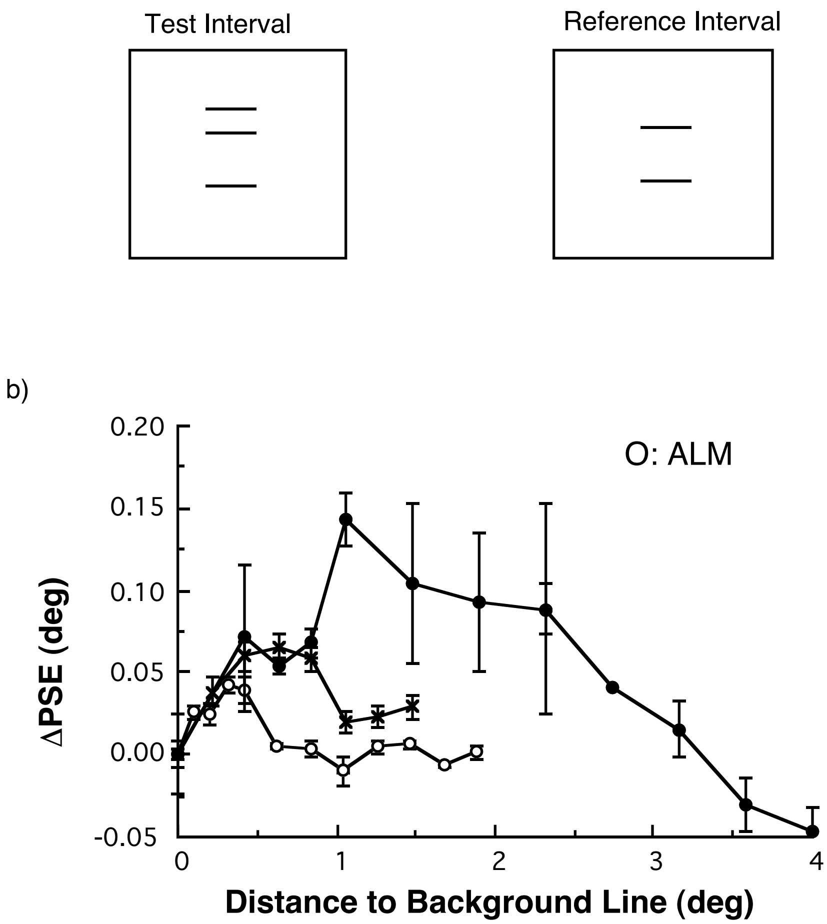


Fig. 1 a) Stimulus configuration used in Burbeck & Hadden (1993). The test and reference intervals were 100 ms in duration, and each interval was terminated by the presentation of a masking stimulus. b) Typical results obtained from this experiment. ΔPSE is the increase in the perceived target separation for the stimuli in the test interval relative to the reference interval. The distance to the background line is the distance between the top two lines in the test interval. Filled circles, 3.0° mean target separation; crosses, 1.5° mean target separation; open circles, 0.75° mean target separation.

Results of another study [Burbeck & Pizer, 1994] further support the idea that the scale of the boundariness detectors increases with increasing width of the region being encoded. This study used a quite different experimental paradigm: the task was bisection, and the stimuli used are shown in Fig. 2a. The frequency of the sinusoidal edge modulation was a parameter of the experiment. Two mean horizontal widths of the stimuli were used. In the condition described here, the edge modulation amplitude was 0.3 deg (peak-to-peak), the stimulus widths were 0.75° and 1.5°, and the stimulus length was 4°.

On each trial, a probe dot was placed near the center of the stimulus. The observer was asked to report whether the dot appeared to be to the left or right of the local left/right center of the stimulus. From his responses, we were able to infer the perceived modulation of the middle of the stimulus.

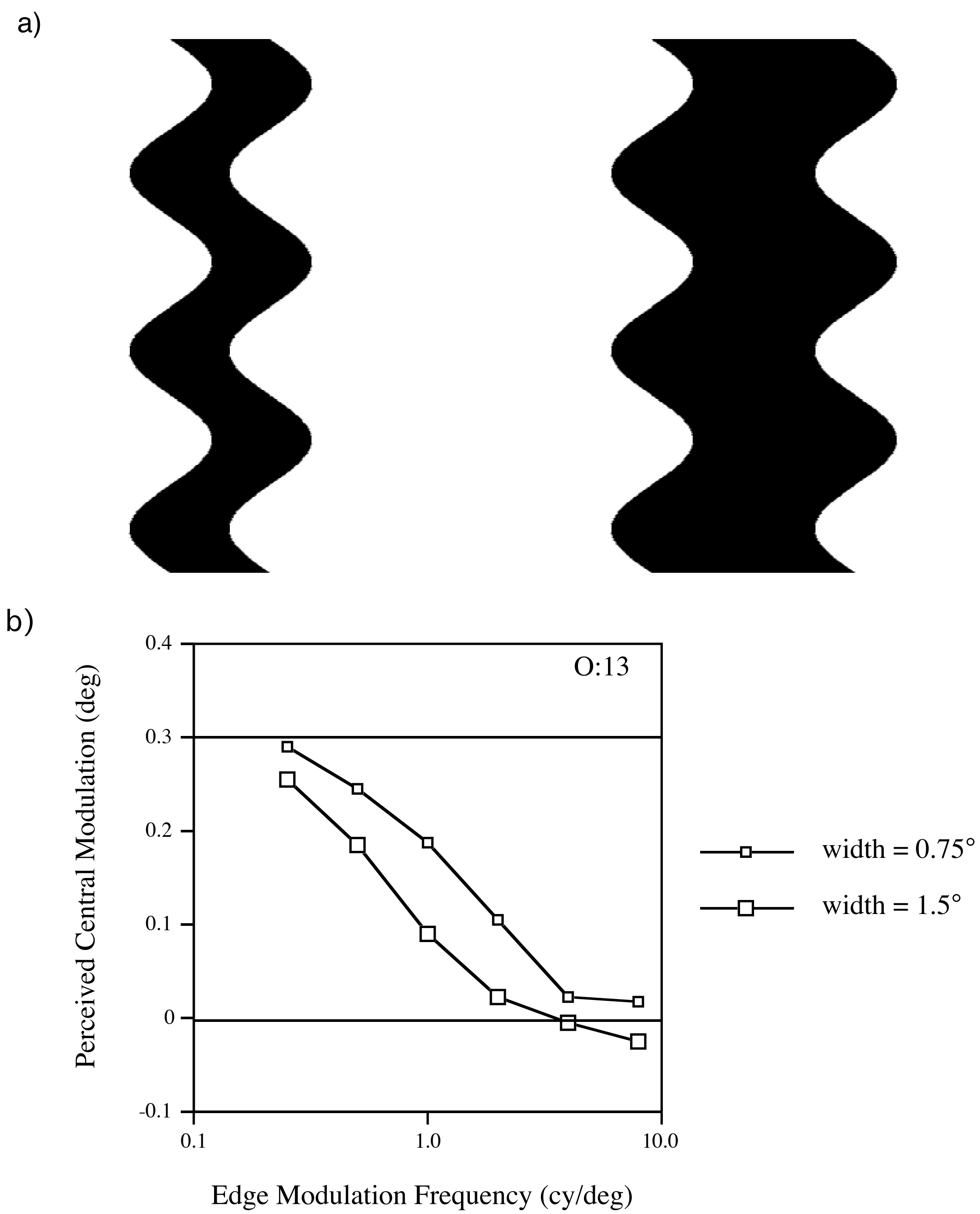


Fig. 2 a) Two of the sinusoidally edge-modulated stimuli used in Burbeck & Pizer [1994]. The edge frequency here is 0.75 cy/deg. The edge modulation amplitude is 0.3°. b) Perceived central modulation of the wiggly-edged object as a function of the frequency of the edge modulation and the width of the stimulus. The effect of the edge modulation depends on the width of the object.

The goal of these experiments was to determine whether the effect of the edge modulation on the judgment of the perceived center was affected by the width of the object. If the area over which boundary location information is integrated increases with increasing width, as inferred from the three-line study, then a given edge modulation frequency should cause the narrower object to appear more wiggly than the wider one (because the boundary information for the narrower stimulus is being gathered with a smaller aperture which can follow the edge modulation more faithfully).

Typical results are shown in Fig. 2b. The perceived central modulation decreases as the frequency increases — as more cycles of the edge fall within the relevant position integration area. More critically in the present context, the effect of the edge modulation depends on the width of the object. For a given edge modulation frequency, the wider object has a straighter perceived middle (i.e., a smaller perceived central modulation) than does the narrower object. This result supports the original conclusion: the area over which boundary information is gathered increases with the distance between the boundaries. Thus, to capture the percept, the effect of edge curvature must depend on the region's width.

We implement this requirement by postulating that small boundariness detectors connect to one another over short distances and large boundariness detectors connect to one another over large distances. This is an economical means of providing connections at all scales while covering the entire visual space. It is also a basically sound idea. Theoretical analysis [Pizer, Burbeck, Coggins, Fritsch & Morse, 1994] suggests that an object-forming system that optimally avoids inter-figure interference across both space and scale and that is invariant to translation, rotation, and zoom [Koenderink, 1990a; ter Haar Romeny, Florack, Salden, & Viergever, 1993] must use boundariness detectors whose scale is proportional to the object's width. In the following, we propose a process by which this type of connection could by made.

THE MODEL: OBJECT REPRESENTATION BY CORES

Fig. 3 illustrates the basic idea of the model. Small scale boundariness detectors connect over short distances to one another and large scale detectors to one another over large distances. Rather than having the boundariness detectors connect to one another directly, however, we postulate that they connect via a representation of the middle of the region. Because connections are made only between same scale detectors, this middle also carries width information, namely the scale of the boundariness detectors that defined that middle location. We call this middle&width representation, the core. It serves as an reification of the connections between the boundaries.

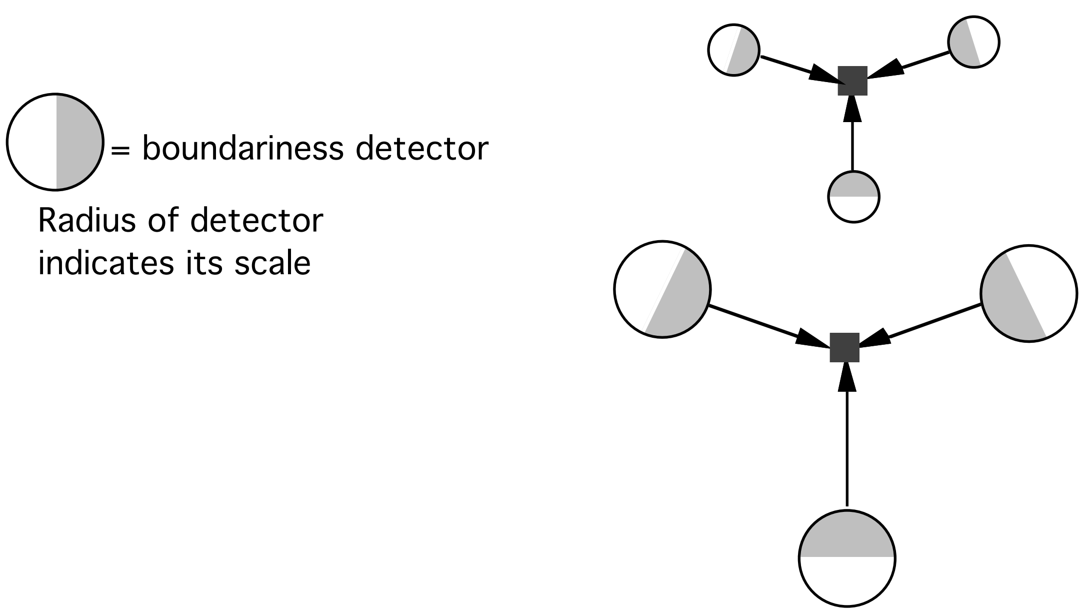


Fig. 3: Boundariness detectors of the same scale interact at a distance proportional to the scale of the detector and along directions normal to the optimal boundary orientation for each boundariness detector.

Fig. 4 shows how the boundaries of a simple region are connected. Each boundariness detector votes for a width that is proportional to its scale, and it casts its votes at a distance proportional to its scale and in the two directions orthogonal to its orientation. Each boundariness detector votes at a strength proportional to its degree of stimulation. Thus, boundariness detectors that catch only part of the boundary of a region will also vote, but more weakly than those more optimally located. A boundariness detector whose scale does not correspond to the local width of the region may vote strongly, but its votes will not be matched by complementary votes from the other side at that scale, because the locations of the votes will not coincide. Finally, the votes cast outside of the region will not coincide with other votes from this region, so they, too, will not be reinforced. (The problem of nearby regions is considered in the discussion section.) Only those votes that are at a scale and location for which there are matching votes from the other side of the region will be reinforced.

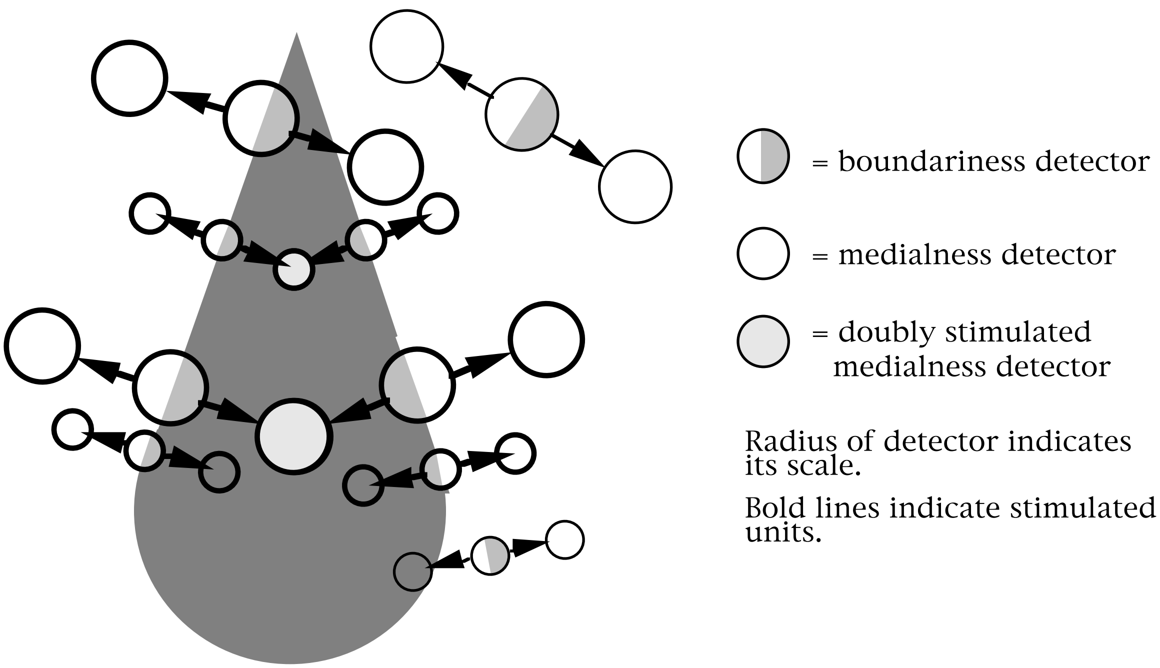


Fig. 4: Boundariness detectors combining (or failing to combine) to produce strong medialness on a teardrop-shaped region.

Adding up the votes of all of the boundariness detectors at each (location, scale) creates some amount of excitation at each (location, scale), resulting in a pattern of excitation in the three-dimensional space: (x,y,σ), two dimensions of space and one of scale. We call the excitation "medialness", where medialness is a measure of the likelihood that a given location is a middle at that scale. Regions of high medialness[1] typically correspond to traces in the 3D space. Each trace is a core. The more concentrated the medialness, the stronger the core. A core represents a region by its middles & widths. A summary of the mathematical description of the core model is given in the outset box.

---

[1]In our computational model, the cores are located using mathematics developed to locate ridges, where a ridge is a generalization of a local maximum. According to [Eberly, 1994] and [Eberly, Gardner, Morse, Pizer, & Scharlach, 1994] medialness ridges are places that are maximal in medialness in the two orthogonal cross-ridge scale-space directions in which medialness has largest cuvature, where all derivatives are taken according to scale-normalized distances.

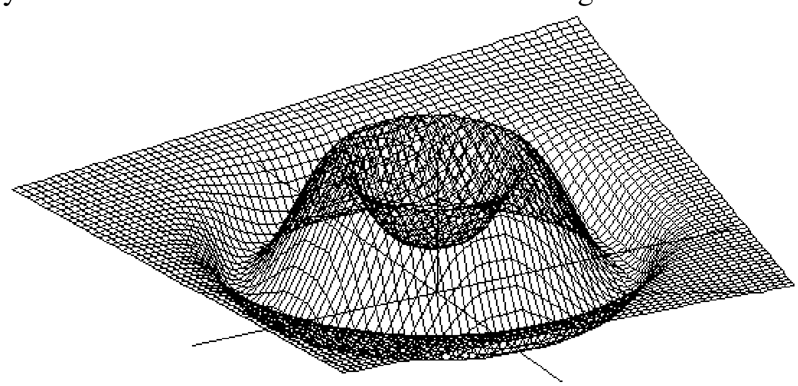
Mathematical Summary of Core Model

Scale space consists of points $(\vec{x},\sigma) \in \Re^2 \times \Re^+$.

Boundariness is a real function B of position $\vec{x}$ × scale σ × orientation $\vec{u}$: $B: \Re^2 \times \Re^+ \times S^2 \to \Re$, where S^2 is the set of all unit-length vectors. The following are three examples of boundariness based on change in luminance:

1) B is a Gaussian directional derivative of the image in the $\vec{u}$ direction: $B(\vec{x};\sigma;\vec{u}) = \left[D_{\vec{u}}G(\vec{x};\sigma)\right] * I(\vec{x})$, where $G(\vec{x};\sigma)$ is a Gaussian with standard deviation σ and $I(\vec{x})$ is the image provided to cortical area V1.
2) $B(\vec{x};\sigma;\vec{u}) = \left|\left[D_{\vec{u}}G(\vec{x};\sigma)\right] * I(\vec{x})\right|$, the previous result but without sensitivity to contrast polarity.
3) B is a Gabor function with odd phase, with a frequency that is a low enough multiple of its scale to provide few significant lobes, and oriented in direction $\vec{u}$, convolved with $I(\vec{x})$: $B(\vec{x};\sigma;\vec{u}) = \left\{G(\vec{x};\sigma)\sin\left[2\pi\nu\sigma(\vec{x} \bullet \vec{u})\right]\right\} * I(\vec{x})$ for $\nu < 0.25$.

Medialness is a real function of position $\vec{y}$ × scale σ: $M: \Re^2 \times \Re^+ \to \Re$. It integrates the responses of boundariness detectors whose scale is σ and whose centers lie on a circle with center $\vec{y}$ and radius proportional to σ with constant of proportionality $1/\rho$. Formally, $M(\vec{y};\sigma) = \int_{S^2} B\left(\vec{y} - \frac{\sigma}{\rho}\vec{u};\sigma;\vec{u}\right) d\vec{u}$. If boundariness is defined as in example 1 above, then medialness is obtained by convolving the image by an appropriately scaled version of the kernel shown in the figure below.



A core is a locus in $(\vec{y};\sigma)$ defined by $\textbf{core}(\vec{y};\sigma) \equiv \underset{(\vec{y};\sigma)}{\text{ri}\not{d}\text{ge}}\, M(\vec{y};\sigma)$, where *ridge* denotes an operator that identifies a maximum in *n-1* specified directions in *n*-space and $\text{ri}\not{d}\text{ge}$ indicates a ridge in scale space, in which all distances are relative to σ. (Eberly [1994] discusses such possibilities for ridges.) Thus, cores are 1-manifolds in scale space, with each point on the core giving medial position $\vec{x} \in \Re^2$ and scale σ (∝ half-width) of a figure.

The voting process is the means by which a boundariness detector contributes to solving the problem of segregating a region from its background. It effectively seeks out the other side of a region at its scale. This seeking avoids the problem of having to know the scale of the region, *a priori*. The creation of a core by this voting process defines a spatial region, but the region so defined may not correspond to a component. It may represent larger- or smaller-scale shape properties (as discussed below and shown in Fig. 7). Thus, we use the term figure to mean the spatial region defined by a single core. A figure may correspond to what an observer would term an object, but as noted above, prior experience also contributes to the determination of what is judged to be an object.

An important property of the core is that the spatial resolution with which its location is represented is proportional to the scale of the associated boundariness detectors. Thus, the resolution is proportional to the figure's width at that location — just as the bisection threshold is proportional to the width being bisected. The width is also represented with a resolution proportional to itself (cf. the scaling of separation discrimination thresholds). Thus, as illustrated in Fig. 5, this (middle, width) track can be thought of as a fuzzy medial axis. The location of the core represents the figure's middle. The spread of the core represents the width of the figure.

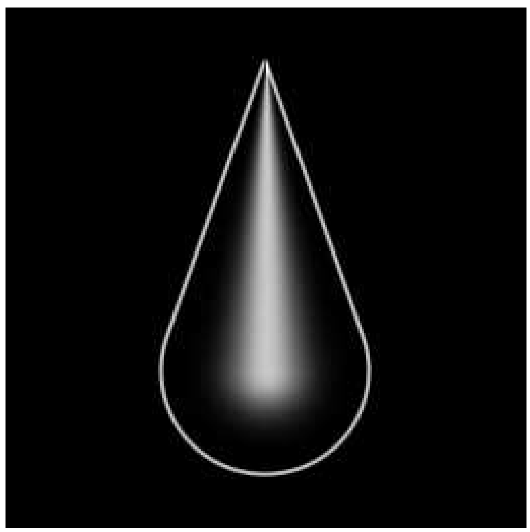


Fig. 5: A teardrop-shaped figure (shown here by its boundary) and a representation of its core, with core strength being indicated by intensity. Because a core is a locus in 3-space, it is awkward to represent in 2D. We use two conventions: 1) a fuzzy core in the image plane, where the core's width indicates the figure's width, as in this figure, and 2) a trace in scale space, where the height indicates the width (Fig. 7). We do not yet know what the width of a given core might be; we know only that for this to be an accurate representation of the visual percept, it must be proportional to the figure's width. Thus the exact width (or height) depicted in the figures is arbitrary.

The proportionality between the contributing boundariness detectors and the local width of the figure means that a given core is insensitive to a protrusion or indentation whose scale is small relative to that local width. Consider the objects in Fig. 6. They have a similar shape at their largest scales despite having widely differing boundary characteristics. Each of these objects would, in our model, generate a long smooth core down the middle. The edge variations in the jagged-edged object would be represented by other smaller-scale cores.

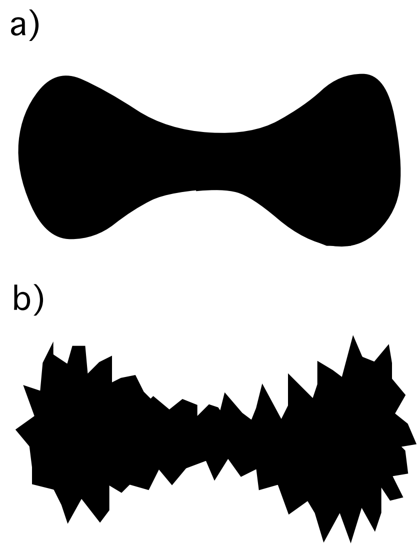


Fig. 6: Figures that differ at small scale but are similar at larger scales.

The core representation naturally encodes important aspects of figural shape. Specifically, the core description separates curvature (i.e., bending) from changes in width (i.e., bulging or compressing) [cf. Blum & Nagel, 1978], a distinction that requires comparison of one side of the region with the other. The spatial derivatives[2] of the core (i.e., the derivatives of the projection of the core onto the spatial plane) yield curvature. The scale derivatives of the core (i.e., the derivatives of the vertical direction in the 3D (x,y,σ) space) indicate the way in which the figure's width is changing. Thus, the core provides a solid basis for analysis of the shape of individual figures.

Multiple Cores

Our mathematical and computational studies [Fritsch, 1993; Morse, Pizer, & Liu, 1993] have shown that the core is normally unbranching, so even simple regions will in general induce many cores. Consider the saw-shaped region shown in Fig. 7: it has a vertical core and a horizontal core at the largest scales, each with approximately constant width (scale); it also has cores for each corner, each sawtooth, and each inter-tooth indentation.

[2] These derivatives need to be taken in scale-space geometry, i.e., according to scale-normalized distances. The required mathematics are laid out in [Eberly, 1994] and [Eberly, Gardner, Morse, Pizer, & Scharlach, 1994].

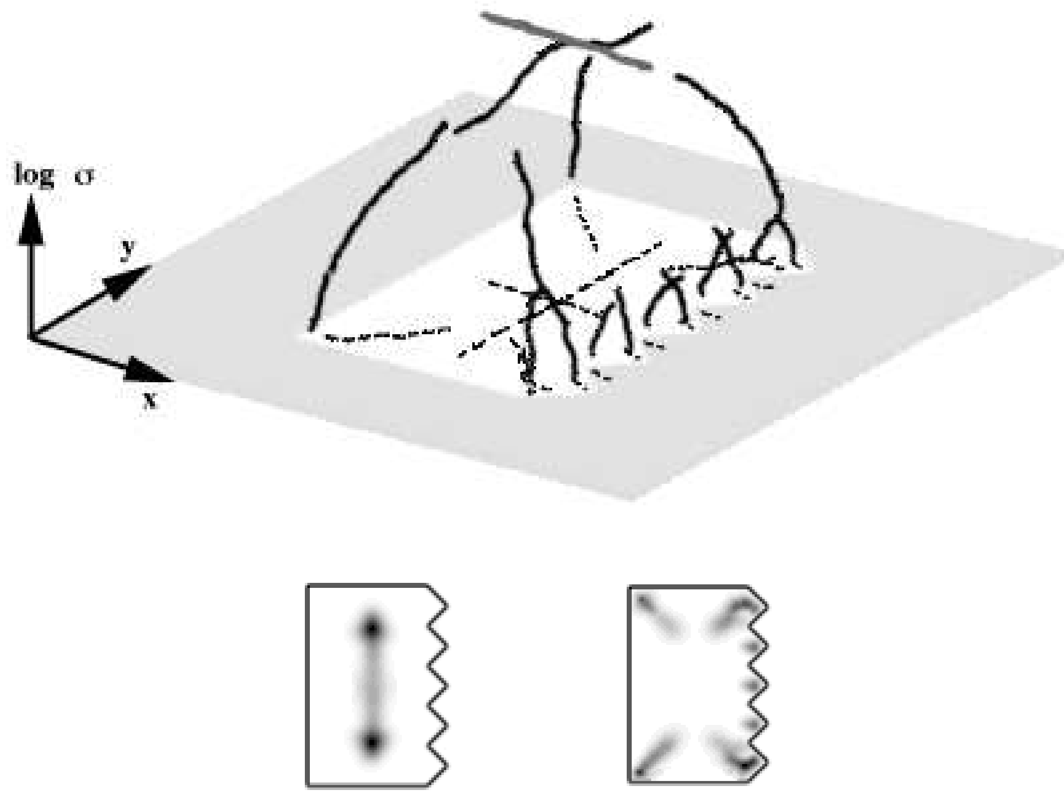


Fig. 7: Cores computed from a rectangle with one saw-tooth edge. The region is shown in white for clarity; the cores were calculated on a uniform filled region. In the top diagram the cores are shown as curves in scale space, and dashed lines show core center locations projected onto the region. The height at which a core is portrayed represents its scale, which in turn captures the width of the object at the corresponding location. The cores that bisect the rectangle in the horizontal and vertical directions are the two cores at the top of the upper figure. Each is approximately constant in scale because the rectangle has parallel sides (at that scale). The horizontal core is shown by the lighter curve. The two cores do not cross in scale space: the horizontal one is higher in scale. Four other cores arise from the four corners. These increase in width from zero as one moves toward the interior of the rectangle. Smaller scale cores arise from the protruding saw-teeth and from the indentations between them. In the bottom diagrams, selected cores are portrayed as blurs on the region itself.

Fig. 8 shows the cores of some simple shapes to help illustrate how regions can be represented by their cores. The scale of the core is symbolized by the width of the line (or dot) representing it. Cores of different scales are shown on different copies of the regions for clarity. A circle has the strongest core, i.e., the sharpest ridge (peak, in this case) of medialness, receiving same-scale votes from locations all around its perimeter. A square is a circle with corners. Corners have linear cores that slant upward in scale space The magnitude of the upward slope determines the angle of the corner. Rectangles are oblongs with corners. Other shapes may be represented by cores that curve in space and/or scale, by cores outside of them which represent indentations, and by smaller scale cores which represent smaller subparts, indentations, or protrusions.

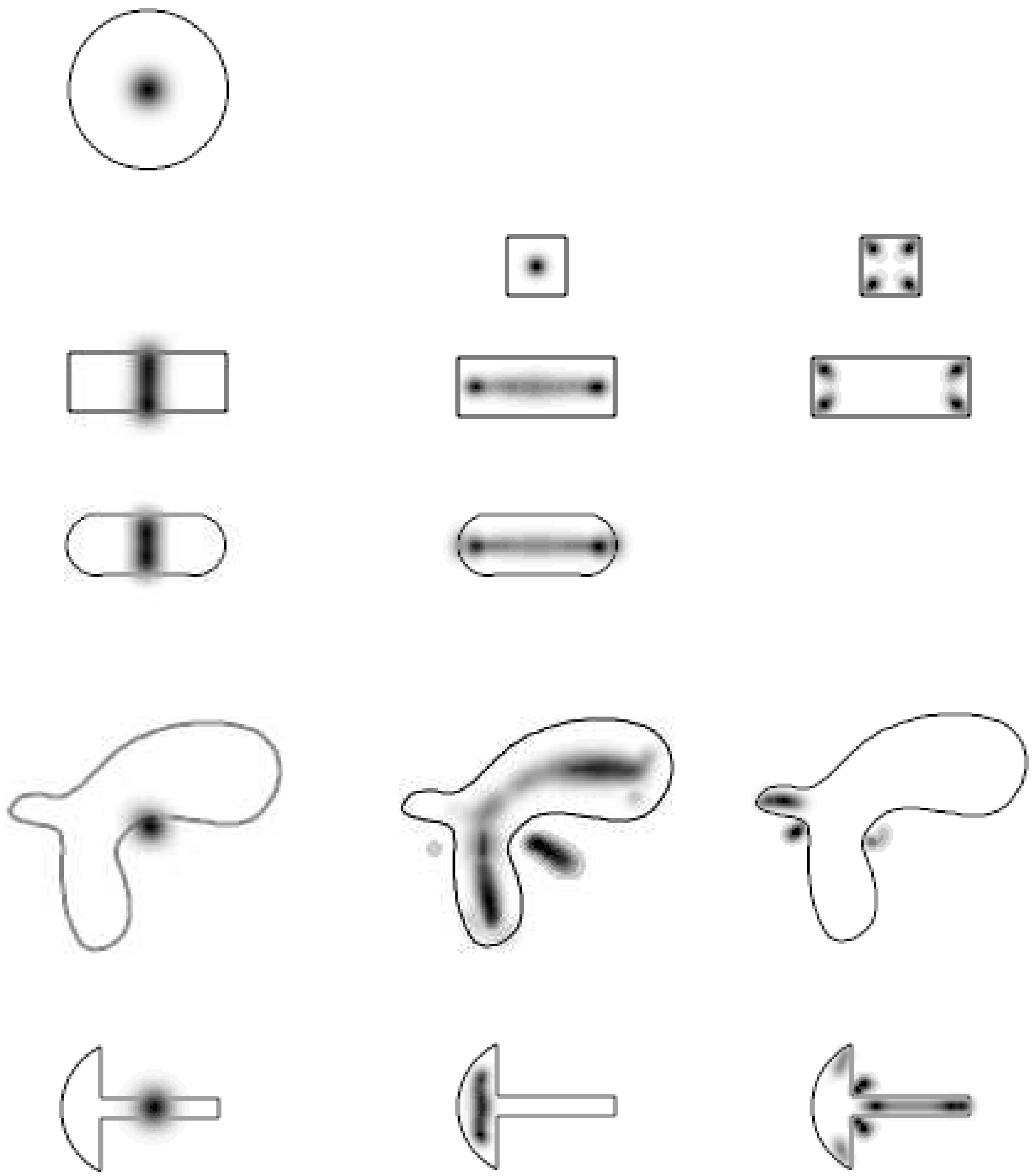


Fig. 8: Some simple shapes and their cores. The left column shows cores at large scale, the center column cores at medium scale, and the right column those cores that include small scales. The results at medium scale in row 5 and at medium and small scale in the bottom row were calculated with excitation along the core (see Fig. 10 and related text).

Using Cores

Thus far we have laid out the basic ideas of a model of how simple connected spatial regions can be found and represented, and we have provided some demonstrations of the cores that result from the associated computations (computational details can be found in [Morse, Pizer & Liu, 1993]). Cores have also been successfully computed for more complex scenes in which there is luminance variation within the regions themselves [Morse, et al, 1993; Fritsch, Pizer, Morse, Eberly & Liu, 1994].

While such computability is essential, another crucial test of the model is its ability to support the rich variety of interpretations that can be made of a given region, depending on its context. For example, consider Shimojo, Silverman and Nakayama's [1989] elegant demonstration of the primacy of spatial organization in the perception of what we think of as one of the most basic visual properties: motion. Fig. 9 shows one type of configuration they used. The stripes in the horizontal rectangular regions were presented in motion and were seen to be moving horizontally because of the elongation of the rectangles in that direction [Nakayama & Silverman, 1988]. This percept was dramatically altered, however, by the addition of stereo cues that brought the horizontal dividing strips into a more proximal plane: the striped regions were then seen to move in the vertical direction, following the elongation of the single large vertical rectangle. Any model of shape representation must be able to cope with such phenomena. Similarly, it should be able to accommodate at least some effects of occlusions and illusory contours. In general, the perception that one region lies in front of another causes the nearer region's boundaries to disassociate themselves from the more distal region, which may then be seen to be completed behind the occluding figure. These profound perceptual effects must be made manifest in the representations of the contributing regions.

Fig. 9: Modification of the barber-pole illusion. When the diagonal stripes move, they appear to move horizontally. When the strips intervening between the three horizontal rectangles are presented in a plane in front of the striped regions, however, the stripes appear to move downward [Shimojo, Silverman & Nakayama, 1989]

How does the core model handle such situations? To do this, some additional details need to be added to the basic ideas of the model. First, we posit that there are excitatory and inhibitory connections along the core. In the terminology of the model, medialness detectors excite nearby medialness detectors of similar scale in the direction of the core and inhibit nearby medialness detectors of sufficiently different scale. The inhibition sharpens the core, yielding higher spatial resolution (still proportional to the width, but with a different constant of proportionality). Excitation along the core's direction allows the core to bridge gaps and weaknesses in the boundary.

Excitation along the core can be accomplished without having oriented medialness detectors. An example of how this could be done is given in Fig. 10. The important point here is that cores seek to extend themselves unless specifically terminated.

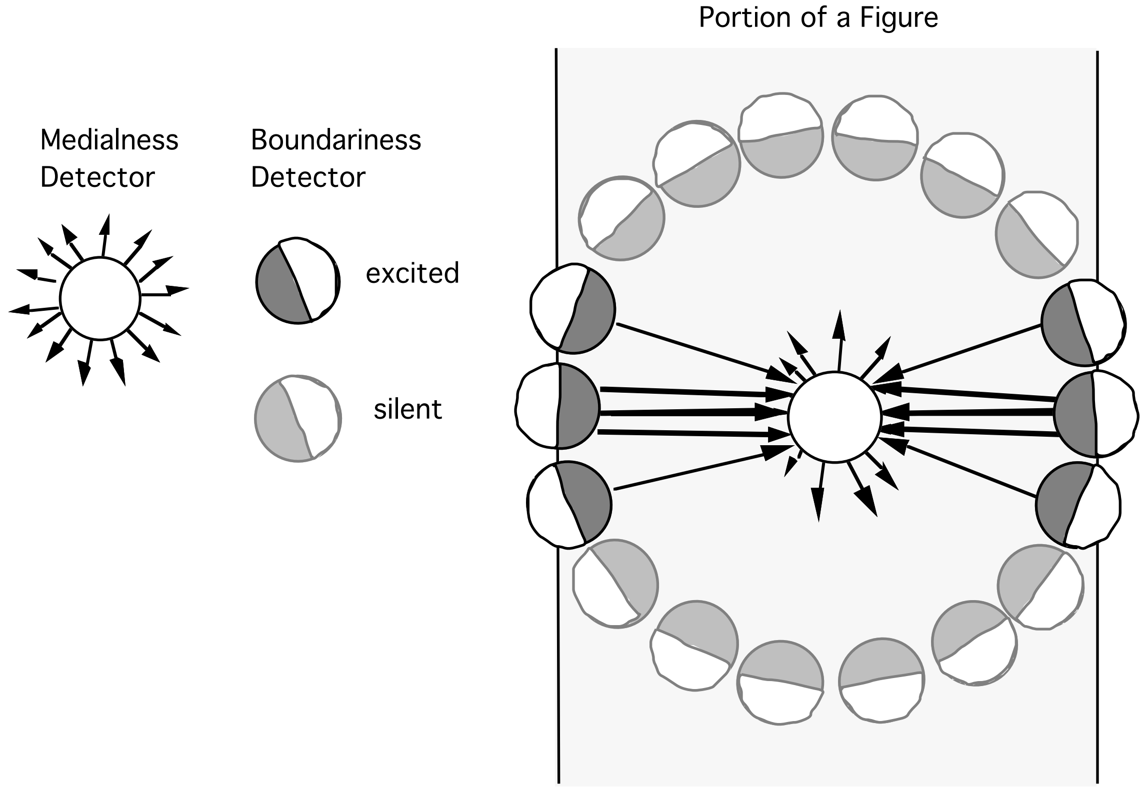


Fig. 10: Representation of possible excitatory and inhibitory connections that result in enhancement of the core. A medialness detector is postulated to have excitatory connections to all nearby medialness detectors of the same scale (and inhibitory connections to medialness detectors of substantially different scales). The arrows pointing outward from the medialness detector represent the excitatory connections. The length of each outward arrow symbolizes its strength. The other medialness detectors are not shown in this figure for clarity of presentation. All boundariness detectors that can contribute to a given medialness detector are shown by the ring of bipartite circles. Those that are excited vote for the medialness detector, as indicated by the centrally-pointing arrows, and they inhibit the excitatory connections from the medialness detector to other same-scale medialness detectors in the region near which they attach, as shown by the reduction in the arrows pointing outward from the medialness detector. The result is that this medialness detector will excite most strongly those medialness detectors that are of similar scale and that are located in a direction that is consistent with the angle of the boundariness detectors contributing to the excitation of this medialness detector. In this case, this would be medialness detectors immediately above and below the one shown in the drawing.

We postulate next that the distinctive cores formed by corners (i.e., straight cores of rapidly increasing scale) signal termination to the main core to which they point in scale space. A rounded end acts in the same way by virtue of the strength of the localized core it generates. In the absence of such terminators, the core tries to seek an extension of itself, fading away gradually in space if unsuccessful.

To account for percepts arising from occluding contours (whether real or illusory [Kellman and Shipley, 1991]), we postulate that cores that arise from boundaries that are seen or inferred to be in different depth planes are suppressed. When that information is available in the initial parallel processing of the image [Enns & Rensink, 1991], those cores are never formed. For example, the corner cores created by a T-junction would be suppressed because of the assignment of the orthogonal boundaries to different depth planes (on the basis of prior knowledge). This would free the central core to seek its extension.

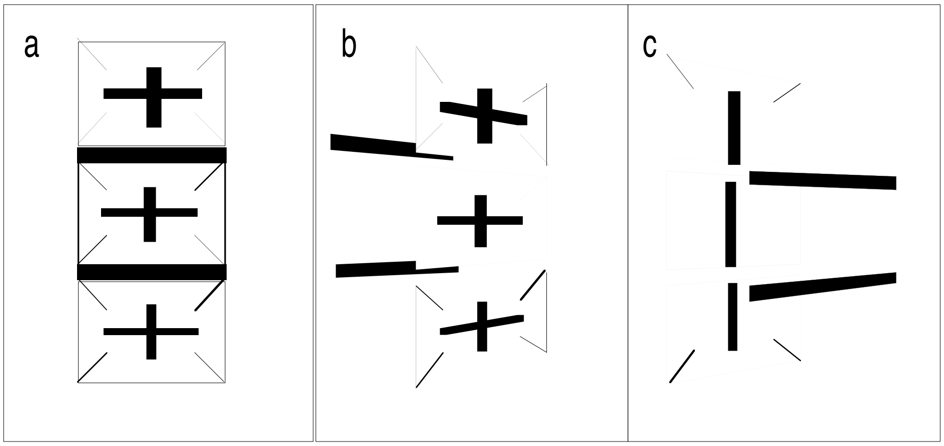


Fig. 11: Schematic representation of the cores of the three horizontal rectangles from Fig. 9. Here the third dimension represents depth not scale. Fig. 11a shows the complete set of cores that would be generated for each horizontal rectangle. Fig. 11b shows the cores that would remain if the intervening strips were seen to be behind the rectangles: the cores are unchanged. Fig. 11c shows the cores that would result from perceiving the intervening strips as being in front of the rectangles: the vertical cores extend toward one another, the horizontal cores are suppressed, and some of the corner cores are suppressed.

Fig. 11 illustrates application of these rules in the case of the example from Fig. 9. The diagonal stripes have been removed for clarity of presentation. The cores are shown schematically by the crossed lines in the centers and the diagonal lines in the corners of the three horizontal rectangles. Fig. 11a shows the core representation that would be created on the basis of the local information within each of the three horizontal rectangles. (Other cores arise from the intervening strips and from the overall object.) When the intervening strips are presented at a farther distance than the rectangles, as shown in Fig. 11b, the representations of the three rectangles is unchanged. When the intervening strips are presented in front of the rectangles, however, as shown in Fig. 11c, application of these rules would suppress the horizontal cores and the interior corner cores as shown. This would allow the vertical cores to cross the intervening strips to create a single connected core representing the larger vertical rectangle, while suppressing the perception of three individual horizontal rectangles.

Although these particular rules are speculative at this point and remain to be tested, the example shows that the core representation has sufficient richness to support the multiple interpretations that are possible for even a simple visual region. The inherent flexibility of the visual system in interpreting the visual input must be considered when evaluating models of how that visual input is represented. The visual system must be able to operate readily on the representations of primitive spatial regions to incorporate information inferred or gained directly from other parts of the image. This ability to support multiple interpretations constitutes an important test for models of shape representation. In the following section, we compare the core model with other models of shape representation, using this richness-of-representation requirement as well as more conventional criteria.

RELATION TO OTHER MODELS

Multiscale Representations

Any discussion of the relationship between the core model and other spatial models in the literature must begin with acknowledgment of the long and well-developed history of the idea of multiple spatial scales in human vision. The seminal studies in this area [e.g., Blakemore & Campbell, 1969] suggested that the human visual system responds selectively to the spatial frequency of the stimulus, and subsequent research refined this idea with the suggestion of more local scale-selective processors [Daugman, 1980; Koenderink & van Doorn, 1982; Watson, 1987]. The basic idea, that spatial scale is important in human visual processing, has been supported and extended by substantial psychophysical, physiological, and theoretical work. For a review of some of the spatial frequency ideas see [DeValois & DeValois, 1980; Kelly & Burbeck, 1984].

Our model rests heavily on the idea of scale-selectivity. It begins with self-similar arrays of boundariness detectors at multiple scales, similar to Burt and Adelson's pyramid [1983] and to Koenderink's model [1984]. Our model proposes that the scale of the boundariness detector determines the scale at which it will communicate with other boundariness detectors: large-scale boundariness detectors link to one another across large distances; small-scale boundariness detectors link to one another across small distances. Thus the scale of the boundariness detector determines the role that it will play in the extraction and representation of significant spatial regions, and ultimately, objects.

Edge-Based Models of Shape

We have used the term "boundariness" to mean a graded response to the spatial changes in luminance (or other features of the stimulus) that occur near an object edge. We use the term "edge", on the other hand, to mean a spatial locus bounding an object. Core-based analysis of an image uses boundariness as its input. It thus has the unusual advantage that the visual system does not have to begin by finding the edges of the region.

Strictly edge-based models of shape perception, on the other hand, begin with the assumption that the edge has been found, presumably by connecting the loci of high boundariness into a closed curve. In these models, the connected closed curve, or its decomposition into component parts, is the basic representation of the object.

Pursuing this approach, Grossberg [originally in 1985, most up-to-date position in 1994] proposed a neural network model of vision that is based on such connections together with filling-in operations. His proposal emphasizes the importance of object boundaries being closed. The most important distinction between his model and ours is in the way in which regions are found and represented. Whereas Grossberg's model is based on the locus of the edge, ours is based on a representation of the spatial relationships between opposite boundaries of the region.

Representing objects in terms of their absolute edge locations has serious weaknesses. The task of finding edges is itself difficult, especially in low signal-to-noise conditions or if the edges are blurred or occluded. To cope with this problem, Grossberg (and others) proposes that larger scale edge detectors be used to preserve continuity. With this approach, the size of gap in the boundary that can be spanned is independent of the size of the object itself. Core-based analysis, on the other hand, predicts that the size of the spannable gap will covary with the width of the object itself (assuming that the aspect ratio is kept constant). Experimental results on subjective contours indicate that there is some such scaling [Shipley & Kellman, 1992], but the object-size dependency has not yet been explicitly tested.

Edge-based models are subject to another serious criticism: the resulting representation does not carry shape information in a rapidly accessible manner. Thus, there is no natural means of assessing similarities and differences in shape. Storing edge information directly is not economical because there is no dependence on the scale of the object. Further, there is no means of directing attention to one portion of the object because the spatial organization of the object has not been found. In short, encoding edge information as such merely postpones the real problems of shape representation and segmentation of the image into regions, i.e., into areas that probably belong to a single object.

Another class of edge-based models [Hoffman & Richards, 1984; Leyton, 1992; Beusmans, Hoffman & Bennett, 1987] focuses on the importance of regions of high curvature on the edge. These regions are seen to be important shape indicators, marking the breaks between components [Hoffman & Richards, 1984 ; Beusmans, Hoffman & Bennett, 1987] and the ends of symmetric axes [Leyton, 1987]. As important as these regions are, identifying them is not by itself sufficient; one also needs an efficient and robust scheme for explicitly encoding the shape of the adjoining region. As Hoffman & Richards [1984] note, this type of approach does not provide such a representation.

Symmetric Axis Models

The idea of representing objects by their central axes and associated width functions has been proposed previously, but the methods of finding the axes and widths of an object that have been suggested and the nature of the representations that resulted differ from ours in profound ways. Blum's work [1967, 1973, 1978] laid the foundations for all of the subsequent models, so we focus on his work. Blum proposed that objects are represented by their symmetric axes. The symmetric axis is defined to be the centers of all circles that are doubly tangent to the object edge whose interiors lie entirely within the object. This model was intended to be a model of shape, in general, and also a model of how the visual system operates, but it requires too many neurons to be plausible physiologically because each edge neuron, at fine spacing, must be attached to every axis neuron, at fine spacing. Nevertheless, results from two experimental studies [Frome, 1972; Psotka, 1978] support the idea that this medial analysis does approximately characterize important aspects of the human shape perception. So the idea remains viable, although the specific implementation Blum suggested seems unlikely.

There has been much mathematical study of and algorithm development for Blum's symmetric axis and for the generalization he later proposed [1973] in which the axis is the locus of all circles that are doubly tangent to the edge — not just those whose interiors are contained in the object. Other modifications have also been proposed. Brady [1983] proposed a medial axis that is the locus of centers of the chords connecting the tangent positions of circles doubly tangent to the edge. Leyton [1992] proposed a medial axis that is the locus of centers of the shorter circular arc connecting those tangent points. Leyton's axis terminates at the edge, and axes can form on either the inside or the outside of the edge.

As models of visual shape representation, the axis models described above all have serious weaknesses, however. First, they all require that the edge be found rapidly and with a high degree of accuracy before the shape analysis can begin, ignoring the difficult problem of image segmentation. Second, in contrast to the behavior of cores [Morse, Pizer & Fritsch, 1994] the set of doubly tangent circles is extremely sensitive to small changes in the object edge. For example, the smallest dimple changes the basic structural representation of the object. Third, having an axis only in the long direction of a region weakens the generality of its representation (e.g., see Fig. 11 above).

Process-Oriented Models of Shape

The symmetric axis idea has also led to intriguing process-oriented models of shape in which the inspiration derives from how the object might have gotten it shape. Leyton's symmetric axis model, mentioned above, is intended to allow object shape to be described in terms of the results of symmetric deformations of a simple object. The symmetric axis touches the boundary at the point of force and the force operates in the direction of the the axis. Kimia, Tannenbaum, & Zucker [1994] have also developed a process-oriented model of shape that includes the symmetric axis as one of its descriptors. In their model, the boundary curves evolve according to a variant of the diffusion equation — making this model similar to ours in some respects. The meeting of two evolving boundary curve sections (from the two sides of the object) defines a medial axis for the object. Because these curve sections blur as they move toward the center of the object, the resolution of the resulting medial representation scales with the object width, as in the core model. Because their model is edge-based, however, the first curves to meet are those from opposing sides of the edge detail. These "shocks", as they are termed, propagate onward and contribute, via higher order shocks, to the medial representation of the object as a whole. Thus, in their model, boundary detail is represented before the overall object shape is. In core-based analysis, the large-scale core is found first and detail is represented as desired. It would be interesting to know whether their model, in which the time of shock formation of various orders depends on object width, could account for the specific width-sensitive behavior of perceived wiggliness [Burbeck & Pizer, 1994].

Cores and Components

It has been proposed by others that objects are represented by simple components, e.g., generalized cylinders, and their relationships [Marr & Nishihara, 1978; Biederman, 1987; and others]. The term component is typically used to mean a simpler part of an object whose join to another component is marked by one or two regions of high curvature in the boundary [Hoffman & Richards, 1984]. Biederman advanced the popularity of this approach with his proposal that the visual system approximates object components by a small set of geometrically simple ones, i.e., ones with simple symmetric axes. Much as language is composed of phonemes, so object recognition was to be based on a small set of basic elements.

When considering the relationship of Biederman's model, or any component based model, to ours, it is important to be clear about what is being modeled. Biederman [1987] proposes that a few three dimensional components, whose characteristics are inferred from a combination of 2D shape and non-accidental features, are sufficient to account for the recognition of objects. His tests were all done on familiar objects, although he posits that his model applies equally to unfamiliar objects. His model is explicitly not intended to account for the encoding of high-resolution information about a specific instance of an object. The core model, on the other hand, is a model of how information about the shape of a specific region in an image can be encoded directly from the image information. The shapes of regions are encoded accurately (at scale) rather than being approximated by one of a limited set of components. Thus, object representation by cores is a model of one part of shape representation, namely the finding and representation of regions in the image, whereas Biederman's Recognition by Components is a model of shape recognition. His model doesn't find components; it uses them and predicts recognition. The core model finds regions and represents them in a way that is useful for subsequent interpretation.

MIRAGE

As models of scene analysis, the core model and the MIRAGE model [Watt & Morgan, 1985; Watt, 1988] have some interesting similarities that have been accentuated by more recent research. A key feature of the MIRAGE model is that it includes local scale-space analysis of the luminance distribution followed by analysis in terms of the locations of important features. The core model operates similarly: boundariness detectors identify regions of important transitions, and the core encodes the spatial relations between those regions. The importance of scale-selective information at a local level and position relations at a more global level has been nicely demonstrated recently by Morgan, Ross and Hayes [1991]. An important difference between the models is that, after an initial parallel scale-dependent analysis, MIRAGE brings the information from the various scales back together in a single representation. The core model keeps the information from the multiple scales separate, in distinct cores.

DISCUSSION

Controlling the Formation of Cores

As described above, a core could be created for every pair of roughly parallel boundaries in the scene. This is clearly undesirable: boundaries with similar characteristics but opposite directions (e.g., white to black with black to white) should connect most strongly. Further, high-resolution representations of shape are not available for all regions in the scene simultaneously. Palmer [1990] has shown that length judgments cannot be performed with optimal accuracy simultaneously at several locations. Burbeck and Yap [1990] have shown that even the two distance judgments required for a bisection task cannot be made simultaneously at highest resolution. Thus, it is reasonable to suppose that, if cores are the basis for high resolution spatial judgments, they are not formed in parallel across the scene. Instead, at least some attentional control is required.

Interference from Neighboring Objects

Because the size of the relevant boundariness detectors scales with the size of the region being encoded, boundary information will typically not be gathered at the highest spatial resolution. Instead, a substantial area near the region's boundary will be included in the analysis. This raises the possibility of neighboring boundaries interfering with the representation of a given region. The results of the 3-line task discussed above (see Fig. 1), indicate that the range over which such interference can occur is quite large indeed. The probe line affected the perceived target separation whenever it was closer to the target line than the targets were to each other. This type of interference poses problems for accurate segregation and representation of regions with nearby neighbors.

The results of that experiment [Burbeck & Hadden, 1993] also point to a solution to the problem, however. The probe line was found to have a considerably larger effect with a 100 ms exposure duration than it did with a 500 ms exposure duration. The range of the effect did not change, but its magnitude did. The contribution of the adjacent line to the perceived target separation was apparently attenuated over time. This suggests that nearby regions may interfere with one another's representation, but that interference is attenuated over time, perhaps as attention is more narrowly tuned to the region of interest [Moran & Desimone, 1985]. Whatever the mechanism, it seems likely that it is an iterative process.

Cores Alone Aren't Enough

A complete representation of even a simple region would also include a representation of its surface characteristics, and a detailed representation of the boundary at the smallest available scale. More complex objects would also require that the relations among the core-represented regions be encoded. One of these relations might be the one between corner cores and central cores described in connection with the stimulus shown in Fig. 11. While some spatial relations may be captured by re-calculation of cores after analysis for occlusions relations, other relations do not have corresponding cores and so would have to be represented in some other way. The core may be helpful in locating possible sites of attachment between parts, however, as the core will jog, increasing or decreasing in scale at the site of attachment of relatively large scale protrusions or indentations.

The problem of feature binding, e.g., tying a color, velocity, texture, etc. to a region, is a problem that has arisen in other models and is a current subject of both psychological investigation [Treisman & Gelade, 1980; Treisman, 1988] and considerable physiological research [e.g., Zeki, 1990]. The detailed representation of boundary characteristics is less well understood. The boundary of an object, such as that in Fig. 6, is surely represented in terms of some of its statistical properties. We don't know the location and scale of every zig and zag unless we attend to them specifically. Thus boundary representations may have some properties in common with the representation of background information and/or surface or texture characteristics. The core representation carries neither of these types of information. The core can be used, however, to associate such a statistically characterized boundary with its object.

Furthermore, because the core carries information about the boundary's location (at a lower resolution), it can be used to guide a more detailed analysis [Pizer, Murthy & Chen, 1994]. The core would, in fact, be a fairly useful guide for such an analysis because it depends on information from both sides of the object, and thus is less sensitive to random variations in the edge location.

Summary

The basis of the core model is the idea that a figure's boundaries are related to one another at a scale determined by the figure's width. This idea has considerable experimental support, is an economical scheme, and has the intrinsic property of zoom invariance. A key feature of the model is the process by which the boundaries are related. As the boundariness detectors vote for a position and scale, they allow the boundaries effectively to seek each other out, thereby contributing to the process of segregating figure from ground. This process operates directly on the gray-scale image. It divides complex objects into constituent components, the objective of some edge-based models, while simultaneously creating a representation of those components that captures essential shape properties and permits alternate interpretations, depending on each region's context. Particular cores of a region can be selectively suppressed by knowledge of the probable spatial configuration of the region's context. Thus the core representation is able to serve as a substrate for application of other shape information.

ACKNOWLEDGMENTS

Many colleagues have contributed significantly to the research reported here. We particularly want to thank Bryan Morse for discovering an efficient means of deriving medialness from boundariness, implementing the computational model, and providing examples of the model's output; David Eberly for his important contribution to defining ridges and developing metrics in scale space (with Bryan Morse), and for writing computer programs that implement these ideas. We thank Jannick Rolland and Derek Puff for their contributions to the early experimental work on the model. This research has benefited from interesting discussions with Jan Koenderink, Wim van de Grind, Daniel Fritsch, Ross Whitaker, James Coggins, Andrea van Doorn, Andre Noest, Bart ter Haar Romeny, Luc Florack and Charles Anderson. We thank Chad Petty, Shobha Murthy, and Alan Liu for their work on the computational aspects of the research. This research was sponsored by NIH grant P01 CA47982 to Stephen Pizer and AFOSR grant F49620-92-J-085 to Christina Burbeck.

REFERENCES

Beusmans, J.M.H., Hoffman, D.D. & Bennett, B.M. (1987). Description of solid shape and its inference from occluding contours. *Journal of the Optical Society of America*, *4*, 1155-1167.

Biederman, I. (1987). Recognition by components: a theory of human image understanding. *Psychology Review*, *94*(2), 115-147.

Blakemore, C. & Campbell, F.W. (1969). On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology, London*, *203*, 237-260.

Blum, H. (1967). A new model of global brain function. *Perspectives in Biology & Medicine*, *10*, 381-407.

Blum, H. & Nagel, R.N. (1978). Shape description using weighted symmetric axis features. *Pattern Recognition*, *10*, 167-180.

Blum, H. (1973). Biological shape and visual science: part I. *Journal of Theoretical Biology*, *38*, 205-287.

Brady, M. (1983). Criteria for representations of shape. *Human and Machine Vision*, Beck, J., Hope, B. and Rosenfeld, A., eds. Academic Press, New York.

Burbeck, C.A., & Yap, Y.L. (1990). Spatiotemporal limitations in bisection and separation discrimination. *Vision Research*, *30*, 1573-1586.

Burbeck, C.A. (1991). Encoding spatial relations. *Vision and Visual Dysfunction*, *14*, Chapter 2. Watt, R. J., ed. MacMillan, London.

Burbeck, C.A. & Hadden, S. (1993). Scaled position integration areas: accounting for Weber's law for separation. *Journal of the Optical Society of America*, *10*, 5-15.

Burbeck, C.A. & Pizer, S.M. (1994). Linking object boundaries at scale: bisection with extended objects. University of North Carolina, Computer Sciences Technical Report TR - 94-041.

Burt, P.J. & Adelson, E.H. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, *31*, 532-540.

Daugman, J.G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, *20*, 847-856.

DeValois R.L. & DeValois, K.K. (1980). Spatial vision. *Annual Review of Psychology*, *31*, 309-341.

Eberly, D., Gardner, R.B., Morse, B.S., Pizer, S.M. & Scharlach, C. (1994). Ridges for image analysis. *Journal of Mathematical Imaging & Vision*, *4*:, 351-371.

Eberly, D. (1994). Geometric methods for analysis of ridges in n-dimensional images. Ph.D. dissertation, Department of Computer Science, University of North Carolina.

Enns, J.T. & Rensink, R.A. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review, 98,* 335-351.

Fritsch, D.S. (1993). Registration of radiotherapy images using multiscale medial descriptions of image structure. Ph.D. dissertation, Department of Biomedical Engineering, University of North Carolina.

Fritsch, D.S., Pizer, S.M., Morse, B.S., Eberly, D.H. & Liu, A. (1994). The multi-scale medial axis and its applications in image registration. *Pattern Recognition Letters*, 15, 445-452.

Frome, F.S. (1972). A psychophysical study of shape alignment. Technical Report TR-198, University of Maryland, Computer Science Center.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. *Perception and Psychophysics*, *38*, 141-171.

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception and Psychophysics*, *55*, 48-121.

ter Haar Romeny, B.M., Florack, L.M.J., Salden, A.H. & Viergever, M.A. (1993). Higher order differential structure of images. *Information Processing in Medical Imaging* (IPMI 1993), Barrett, H.H. & Gmitro, A.F., eds. *Lecture Notes in Computer Science*, *687*, 77-93. Springer-Verlag.

Hoffman, D.D., & Richards, W.A. (1984). Parts of recognition. *Cognition*, *18,* 65-96.

Hubel, A.H. & Wiesel, T.N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, *195*, 215-243.

Kellman, P.J. & Shipley, T.F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology, 23,* 141-221.

Kelly, D.H. & Burbeck, C.A. (1984). Critical problems in spatial vision. *CRC Critical Reviews in Biomedical Engineering*, *10*, 125-177.

Kimia, B.B., Tannenbaum, A.R. & Zucker, S.W. (1992). The shape triangle: parts, protrusions, and bends. McGill University Research Center for Intelligent Machines, Montreal, Canada. Technical report, TR-92-15.

Kimia, B.B., Tannenbaum, A.R. & Zucker, S.W. (1994). Shapes, shocks, and deformations I: The components of shape and the reaction-diffusion space. To appear in *International Journal of Computer Vision*.

Koenderink, J.J., & van Doorn, A.J. (1982). Invariant features of contrast detection: an explanation in terms of self-similar detector arrays. *Journal of the Optical Society of America*, 72, 83-87.

Koenderink, J.J. (1984). The structure of images. *Biological Cybernetics*, *50*, 363-370.

Koenderink, J.J. (1990a). The brain as a geometry engine. *Psychological Research, 52*, 122-127.

Leyton, M. (1992). *Symmetry, Causality, Mind*, The MIT Press, Cambridge, Massachusetts.

Marr, D. & Nishihara, H.K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London, B200*, 269-294.
Moran, J. & Desimone, R. (1985). Selective attention gates visual processing in the extastriate cortex. *Science*, *229*, 782-784.
Morgan, M. J., Ross, J. & Hayes, A. (1991). The relative importance of local phase and local amplitude in patchwise image reconstruction. *Biological Cybernetics, 65,* 113-119.
Morse, B.S., Pizer, S.M. & Liu, A. (1993). Multiscale medial analysis of medical images. *Information Processing in Medical Imaging* (IPMI 1993), Barrett, H.H. & Gmitro, A.F., eds. *Lecture Notes in Computer Science*, *687*, 112-131. Springer-Verlag. (To appear in Image and Vision Computing, July 1994.)
Morse, B., Pizer, S.M. & Fritsch, D. (1994). Robust object representation through object-relevant use of scale. *Medical Imaging 1994: Image Processing*, SPIE, 2167, 104-115.
Nakayama, K. & Silverman, G.H. (1988). The aperture problem — II. spatial integration of velocity information along contours. *Vision Research* , 28, 747-753.
Nothdurft, H.C., Gallant, J.L. & Van Essen, D.C. (1992). Neural responses to texture borders in macaque area V1. *Society of Neurosciences Abstract*, *18*, 1274.
Nothdurft, H.C. & Li, C.Y. (1985). Texture discrimination: representation of orientation and luminance differences in cells of the cat striate cortex. *Vision Research*, *25*, 99-113.
Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology, Human Perception and Performance*, *16*, 332-350.
Pizer, S.M., Burbeck, C.A., Coggins, J.M., Fritsch, D.S. & Morse, B.S. (1994). Object shape before boundary shape: scale-space medial axes. Presented at *Shape in Picture* (NATO Advanced Research Workshop). To appear in *Journal of Mathematical Imaging and Vision.*
Pizer, S.M., Murthy, S. & Chen, D.T. (1994). Core-based boundary claiming. *Medical Imaging 1994: Image Processing*, SPIE, 2167, 151-159.
Psotka, J, (1978). Perceptual processes that may create stick figures and balance. *Journal of Psychology: Human Perception & Performance*, *4*, 101-111.
Sáry, G., Vogels, R. & Orban, G.A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science*, *260*, 995-997.
Shimojo, S., Silverman, G. H., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research* , *29*, 619-626.
Shipley, T.F. & Kellman, P.J. (1992). Strength of visual interpolation depends on the ratio of physically-specified to total edge length. *Perception & Psychophysics,* 52, 97-106.
Treisman, A. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology 12*, 97-136.
Treisman, A. (1988). Features and objects: the fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology, A40*, 201-237.
Van Essen, D.C., DeYoe, E.A., Olavarria, J., Knierim, F., Sagi, D., Fox, J.M. & Julesz, B. (1989). Neural responses to static and moving texture patterns in visual cortex of the macaque monkey. *Neural Mechanisms of Visual Perception*, Lam, D.M.K., & Gilbert, C., eds. Portfolio Publishing, Woodland, Texas. 135-153.
Watson, A.B. (1987). The cortex transform: rapid computation of simulated neural images. *Computer Vision, Graphics, Image Processing*, *39*, 311-327.
Watt, R.J. (1988). *Visual Processing: Computational, Psychophysical, and Cognitive Research.*, Lawrence Erlbaum Associates, London.
Watt, R. & Morgan, M.J. (1985). A theory of the primitive spatial code in human vision. *Vision Research, 25*, 1661-1674.
Wolfe, H.K. (1923). On the estimation of the middle of lines. *American Journal of Psychology*, *34*, 313-358.
Zeki, S. (1990). Parallelism and functional specialization in human visual cortex. *Cold Spring Harbor Symposia on Quantitative Biology, 55*, 651-661.