Formation of Image Objects in Human Vision

TR91-031

July, 1991

Stephen M. Pizer
James M. Coggins
Christina A. Burbeck

Medical Image Display Group
Department of Computer Science
Department of Radiation Oncology
The University of North Carolina
Chapel Hill, NC 27599-3175

The research reported herein was carried out with the partial support of NIH grant number P01 CA47982.

To appear in Proc. 1991 Conference on Computer Assisted Radiology.

Formation of Image Objects in Human Vision

STEPHEN M. PIZER, JAMES M. COGGINS, CHRISTINA A. BURBECK

Medical Image Display Research Group
University of North Carolina
Chapel Hill, NC 27514 USA

The stage of human vision in which objects are formed apparently precedes the final determination of many object properties: object identity, brightness, size, and location. Analogously in computer vision, this stage must precede isolated display and measurement of the object. This paper brings together several ideas about how this grouping of space into objects may be done by the human visual system, for example when it is viewing medical images. The result of this synthesis is a new model of form perception.

Edges

An important train of thinking in the study of form perception has been that grouping is based on the local detection and tracking of edges. Hubel and Wiesel [1968] and others have found orientation-sensitive detectors in the visual cortex, detectors that are sensitive to object edges. Also, eye-tracking experiments [Gerrits, 1970] find that in saccadic eye movements, foveation on edges predominates. Pursuing this idea, Grossberg [1985] described a neural network model of vision in which local edges are connected into closed curves enclosing objects.

There are some difficulties with this idea. First, local measures of edge strength are sensitive to error, producing failures of edge tracking. Yet the human is able to jump large edge gaps formed by occlusion, blurring, or noise, perceiving edges when they are not present -- so called subjective edges -- to create perceptually complete objects from incomplete information. Second, it is hard to see how any network of neurons with a dominance of local connections and limited connections per neuron could find such perceptually obvious relations as that between opposite points on two sides of an object (see figure 1 for examples).

Medial Properties

The above difficulties are ameliorated with an encoding scheme responding to the opposite object edges simultaneously, sensing the object region

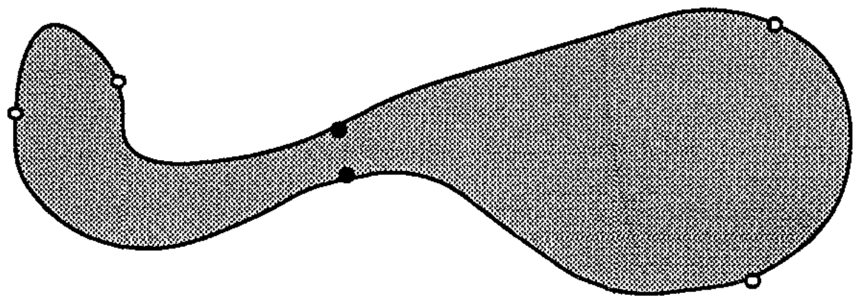

Figure 1: Visually related opposite points on an object

rather than its separate edges. Beginning with Blum [1967], many in the field of computer vision have been attracted by a scheme of this type in which an object is represented in terms of a medial axis or skeleton running down the center of the object, together with a width value at each point on the medial axis. Early research provided some limited psychological evidence that human vision operates in these terms [Psotka, 1978]. More recently Leyton [1984, 1987], a perceptual psychologist, argued that such central-axis-based operation characterizes vision. Leyton suggested further that the long known fact that corners and other object boundary locations of locally maximal curvature are perceptually important is related to the correspondence of these locations to endpoints of these central axes. It has also been noted [von der Heydt, 1984] that subjective edge perceptions derive especially strongly from extensions of edges from corners.

Related entities are the end-stopped cells found neurophysiologically by Hubel & Wiesel [1968] and others. They reported that these neurons are sensitive to the end of a bar but not to a part of a bar that crosses through the receptive field of the neuron. We suggest that this behavior may be described more generally as sensitivity to regions corresponding to medial axis endpoints.

The Blum medial axis is formally defined as the locus of centers of maximal disks in the object (see figure 2). As a result every axis point corresponds to two (or occasionally more) object boundary points where the maximal disk touches the boundary. These two boundary points appear to correspond to each other in a way consistent with the visual percept. The medial axis carries with it (in the radii of the disks) straightforward access to the angle of the object boundary at each of these two boundary points relative to the axis direction at the corresponding axis point. Moreover, the curvature of the axis and of the boundary pair relative to the axis is also straightforwardly accessible.

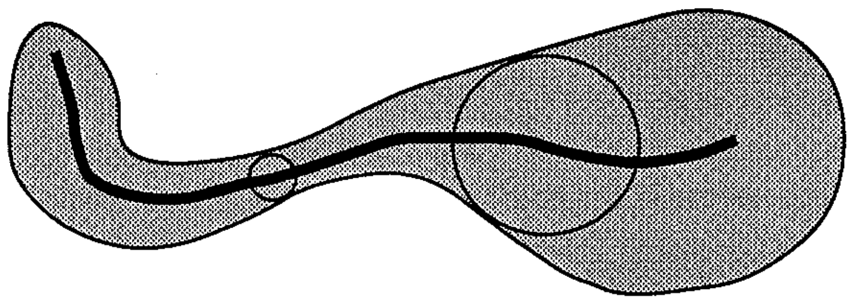

Figure 2: The medial axis for an object

Multiscale Geometry Detectors

Many investigators have suggested that visual grouping must be based on sets of detectors that sense a regional rather than curvilinear (e.g., edge) property, with each detector (neuron) sensing the same property but at different spatial scales. Each neuron calculates a weighted sum of intensities about the point. The spatial weighting function is called the receptive field of the neuron (see figure 3). When applied at all points in the visual field, the spatial weighting function can be thought of as a filter kernel. Different spatial scales correspond to different widths of the receptive field.

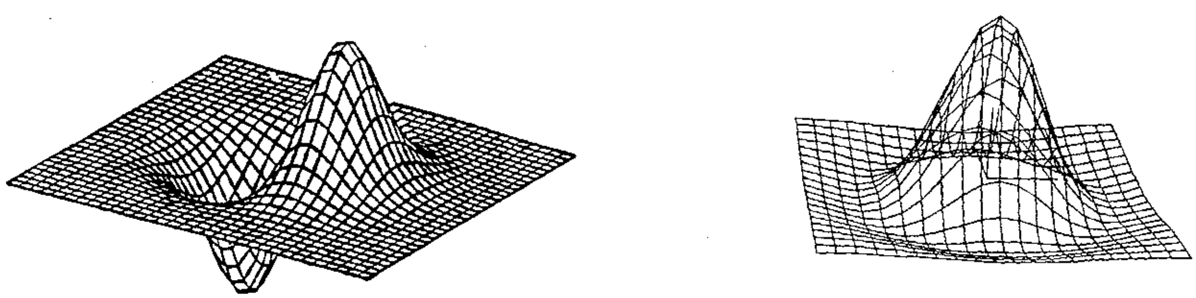

Figure 3: Some neuron receptive fields found in the visual system

Such a multiscale arrangement is suggested by the ability of the visual system to sense an object independent of its size in the visual field and to focus on various levels of detail. Neurophysiological research confirms this property [Young, 1987]. A spatial multiscale operation was first suggested by Campbell & Robson [1968], who proposed that detectors were selective to a limited range of spatial frequencies, spatial frequency corresponding to scale in this case. This idea was implemented quantitatively assuming receptive fields that were sinusoids in a Gaussian envelope (Gabor functions) [Daugman, 1980; Watson, 1987]. However, the most persuasive case for the form of receptive fields was derived recently by Koenderink [1990]. He argues mathematically that any visual system, like ours, that can ignore size change and see at various scales, with detail decreasing as the spatial scale of the neurons increases, must have multiscale receptive fields which are linear combinations of derivatives of a Gaussian. These receptive fields or combinations of them can be thought of as measuring geometrical properties such as "edgeness", "cornerness", and "t-junctionness", in many cases with an orientation. Many of these receptive field types have been found neurophysiologically [Young, 1987].

A particularly interesting set of these receptive fields is a polar set in which each receptive field (relative to its center) is the product of a radial function given approximately by a derivative of a Gaussian and an azimuthal function of the form $\sin(k\theta)$ or $\cos(k\theta)$ for some integer k (see figure 4). All of the receptive fields in figure 4, except those in the central column, come in a sin, cos pair measuring their geometric property at two orientations. This pair of outputs can be taken as the two coordinate values of a spatial vector whose magnitude gives the intensity of some local geometric property and whose direction gives orientational information with regard to that property. The family of receptive fields in this model captures information which the human visual system measures using receptive fields at multiple orientations rather than just two.

The Multiscale Medial Model

Collectively the above ideas have led us to the development of a new model for visual grouping. This model appears reasonable at the neural level as a model of human visual processing, but it is at present untested psychophysically. It produces a group of medial axes by multiscale, regional, geometric measurements. It is based on a set of neurons each of

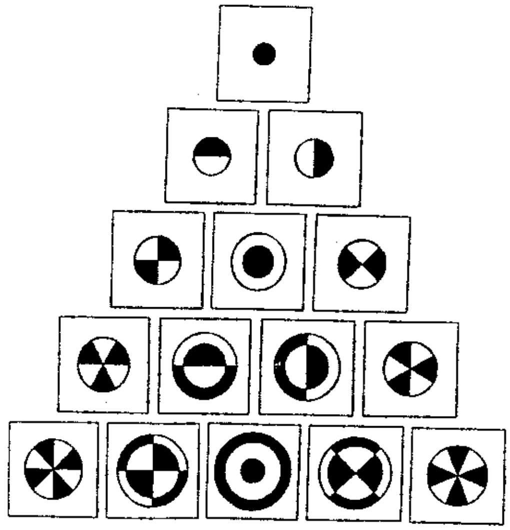

Figure 4: Koenderink polar receptive fields shown by their sign (positive = black) [Koenderink, 1990].

which is identified by a location, a scale, and a geometric property, that can be thought to capture edgeness or cornerness, for example. Specifically, each neuron has a particular receptive field from Koenderink's polar set, at the scale in question. These neurons, dense in space and scale, compute their geometric property by applying their receptive field at their location. Pairs of these neurons produce a magnitude and a direction at their scale. For most of these neuron pairs, the magnitude is high when the receptive fields engage both edges of an object. This occurs when the neuron is equidistant from the two edges because the receptive fields have a circular shape. The direction given by the neuron pair is the axis direction. The neurons cooperate along their directions and compete across scale and position, so that a well-defined set of axis positions is defined. More specifically, neurons excite other neurons that are of similar scale and are along or near their direction in either sense. They inhibit neurons of different scales at the same position and those at the same scale and nearby positons but along different directions. End-stopped cell pairs produce a single-sense direction on the concave side of an object boundary and a magnitude proportional to curvature or cornerness. They excite along the axis in only the single direction sense and inhibit maximally in the opposite direction. Thus they start and end axes.

The result of this operation is a set of traces in scale space (*x,y, scale*) which are ridges of neural response (see figure 5). The *x,y* positions of these ridges form a medial axis for an object, and their scales specify its width at each axis point. Just as with the Blum medial axis, width (scale) angles and curvatures (boundary orientation and curvature relative to the axis) are straightforwardly available. Also, the excitatory and inhibitory connections should produce subjective edges in the appropriate way. Note that this operation applies to grey scale objects with fuzzy edges as well as those with sharp edges.

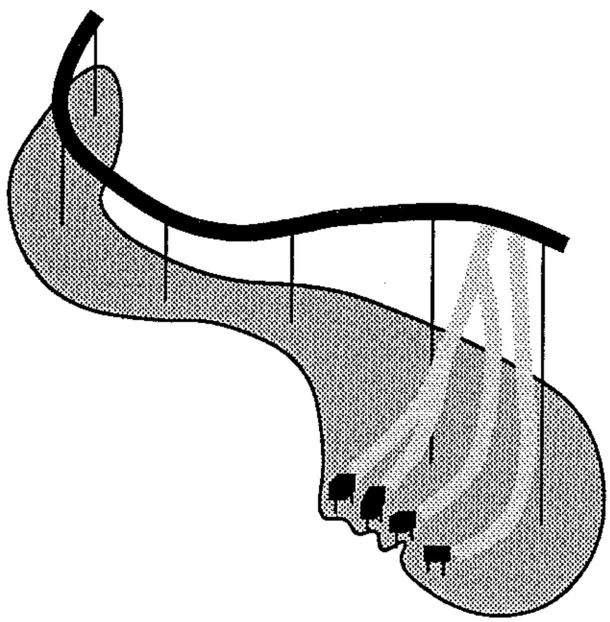

Figure 5: Scale space medial axis traces for an object. Dotted traces are less strong than solid traces.

The above model also produces scale space ridges that correspond to smaller boundary detail (see figure 5). The scale of the main axis allows the identification of a boundary region; smaller scale responses in that region establish the detailed structure of that boundary.

Besides bringing together two important theoretical approaches to understanding shape perception, this model has the advantage of naturally incorporating size constancy and orientation independence. It also suggests that shape will be preserved across small changes in local orientation produced by warping or bending, thus corresponding to the human percept. Detailed psychophysical testing of the basic properties of this model are currently underway. In addition, we anticipate extending this model to deal with objects within objects, objects defined by texture, line boundaries, or even texture boundaries rather then intensity change. We also anticipate modeling spatial interactions between groups such as the apparent stretching of space between nearby groups that we have observed psychophysically.

Implications for Medical Imaging

If this model is supported by experiment, it will have many uses in medical imaging:

1) Its performance will predict human performance, so its success in a grouping-related task can serve as an image quality measure. This property will in time allow the development of image acquisition, processing, and display methods that optimize this image quality.

2) The object-subobject relationships it defines can be computed for any image, to produce a quasi-hierarchy that can be used in interactive computer systems for the fast definition of objects in images [Pizer, 1989]. These defined objects can in turn serve 3D display and object measurement.

3) The groupings defined by this approach can be used to define object inclusion likelihoods that in turn can be used to produce automatic measurements of object volumes, e.g. tumor volumes, or other object properties such as integrated metabolic function.

4) The position and scale co-ordinates along the medial ridges and the outputs of various receptive fields there can be used as a basis for matching of structures in tasks involving registration between objects in rather different images of the same anatomy, such as a simulation and a portal image in radiation oncology.

Work in all of these directions is proceeding in our laboratory.

Acknowledgements

The research reported here was partially supported by NIH grant #P01 CA47982, NASA contract #NAS5-30428, and AFOSR grant #91-00-58. We are indebted to Michael King for derivations, and to Daniel Fritsch and Bryan Morse for computer implementations. We thank Jan Koenderink for permission to use figure 4 and Carolyn Din and Graham Gash for help in manuscript preparation.

References

Blum, H, A new model of global brain function, *Perspectives in Biol. & Med.* **10**:381-407, 1967

Campbell, FW, JG Robson, Application of Fourier analysis to the visibility of gratings, *J Physiol* **197**: 551-566, 1968

Daugman, JG, Two-dimensional spectral analysis of cortical receptive field profiles, *Vis Res* **20**:847-856, 1980

Gerrits, HJM, AJH Vendrik, Simultaneous contrast, filling-in process and information processing in man's visual system, *Exp. Brain Res* **11**:411-430, 1970.

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

von der Heydt, R, E Peterhans, G Baumgartner, Illusory contours and coritical neuron responses, *Science* **224**:1260-1262, 1984

Hubel AH, TN Wiesel, Receptive fields and functional architecture of monkey striate cortex, *J Physiol* **195**:215-243, 1968

Koenderink, JJ, The brain a geometry engine, *Psychol Research*, 1990.

Leyton, M, Perceptual organization as nested control, *Biol Cyb* **51**:141-153, 1984

Leyton, M, Symmetry-curvature duality, *Comp Vis, Graphics, Im Proc* **38**:327-341, 1987

Pizer, SM, JM Gauch, JM Coggins, RE Fredericksen, TJ Cullip, VL Interrante, Multiscale geometric image descriptions for interactive object definition, *Mustererkennung 1989, Informatik-Fachberichte* **219**: 229-239, Springer-Verlag, 1989

Psotka, J, Perceptual processes that may create stick figures and balance, *J Exp Psychol: Human Percep. & Perform* **4**:101-111, 1978

Watson AB, The cortex transform: rapid computation of simulated neural images, *Comp Vis ,Graphics, Im Proc* **39**: 311-327, 1987

Young, RA, The Gaussian derivative model for spatial vision: I. Retinal mechanisms, *Spatial Vision* **2**:273-293, 1987