

THE ROLE OF V1 IN SHAPE REPRESENTATION

Tai Sing Lee,¹ David Mumford,² Song Chun Zhu,² and Victor A. F. Lamme³

¹Center for the Neural Basis of Cognition
and Department of Computer Science
Carnegie-Mellon University
Pittsburgh, PA

²Department of Applied Mathematics
Brown University
Providence, RI

³Department of Medical Physics
University of Amsterdam
The Netherlands Ophthalmic Research Institute
1100 AC Amsterdam, The Netherlands

ABSTRACT

The analysis of higher level global visual attributes such as geometric shapes is thought to be a function of specialist modules in the extrastriate cortices. However, V1 is unique because only in this area can one find cells tuned to orientation and other local features with the spatial precision needed for representing high resolution and geometric aspects of an image. Geometric computation that demands such resolution may need to involve V1 via feedback. V1 neurons are known to exhibit very different types of responses in the short latency (40-80

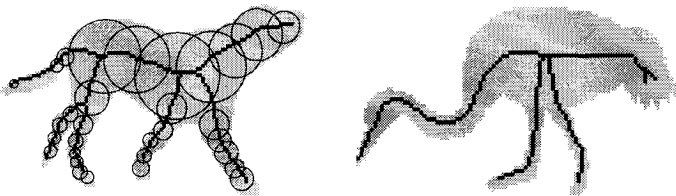
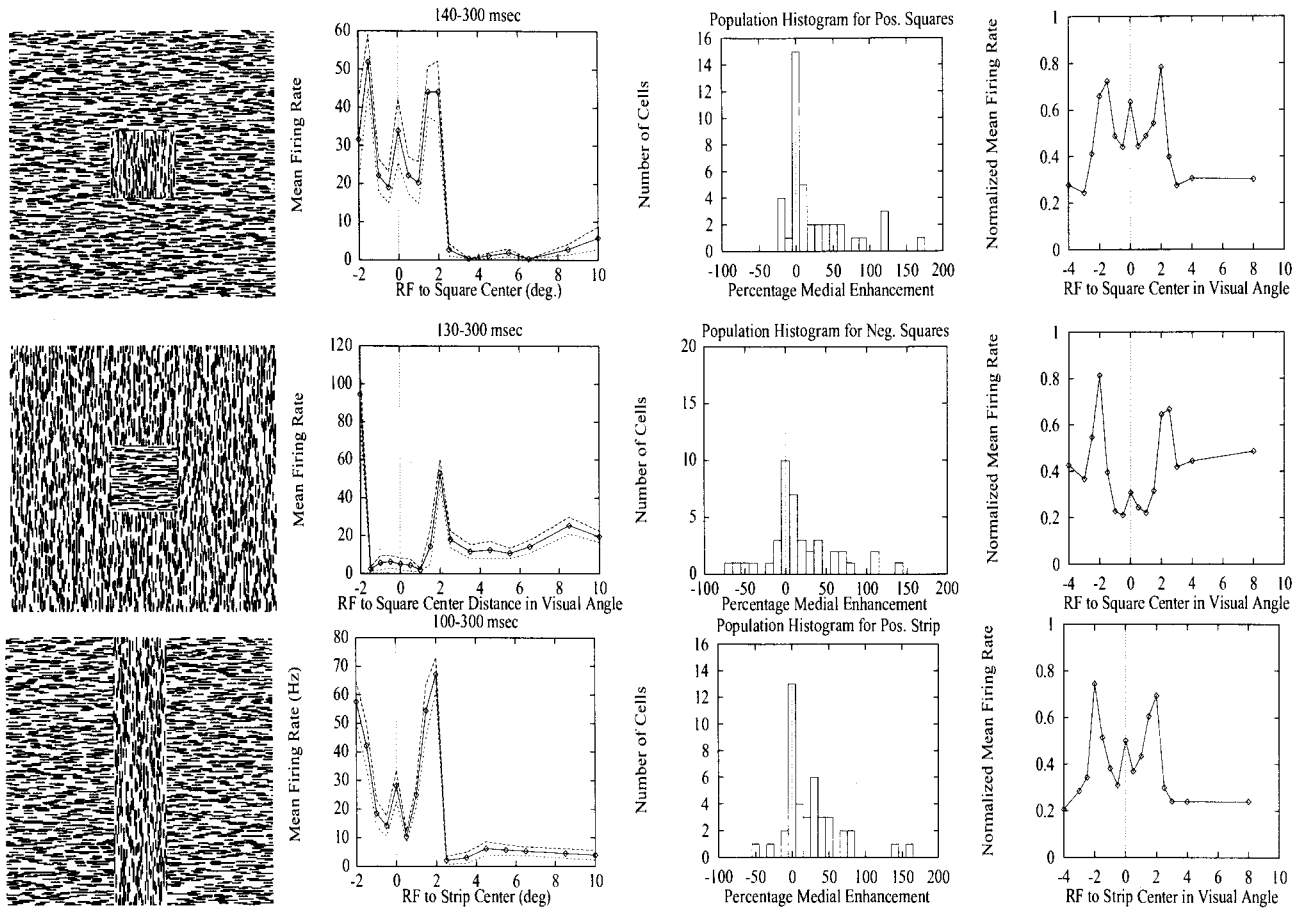


Figure 1. The skeletons of biological forms can be computed by tracing the locus of the centers of the set of maximally inscribing disks inside the figures¹⁰. Hence, the skeletal or the medial axis transform actually include information about the location of the skeleton and the diameter of the inscribed disk. This representation leads directly to methods for calculating salient shape properties of a figure such as relationship between boundaries, aspect ratio, parts and symmetries⁹⁻¹¹. In contrast to contour which is inherently local, the medial axis transform is a robust semi-global descriptor for region.



milliseconds post-stimulus time window) versus longer latencies (80-200 milliseconds)¹⁻⁵. Here we report that the later part of V1 neurons' responses is sensitive to geometric attributes of globally defined shapes. We propose a computational interpretation of these results.

The outline of an object defined by luminance contrast or by discontinuity in surface qualities such as texture, color, disparity and optical flow can easily be recognized by *local filters*. But the computation of shape requires *global methods*: one of these is to label all the 'pixels' that belong to the same surface, in what Ullman⁶ referred to as 'coloring'. Coloring a region takes time: Paradiso and Nakayama⁷ showed that the percept of a white disk forms in stages over 50-100 milliseconds, propagating in from the edges and that it can be interrupted by masks of different shapes. Recent neurophysiological data suggested two possible mechanisms could be used to 'color' a region. One is to link activity of already active cells, as Gray and Singer's data⁸ suggested, using synchrony of spikes to represent the presence of a single large percept. The other is to dedicate a cell or part of a cell's activity to signalling that specific elementary regions are part of a single figural surface that are not cut up by boundaries, as suggested by Lamme's finding that the response of V1 neurons are enhanced when their receptive fields are located within the figure than when they are outside the figure².

Only when a figure is 'popped out', its surface 'colored' and its boundaries detected, can one begin to compute properties of its shape. These properties may suggest the identity of the visible object or they may need to be modified if the object is partly occluded or in shadow. Therefore, we propose that *the processes of scene segmentation, shape computation and object recognition must happen both concurrently and interactively in feedforward/feedback loops* that involve the whole hierarchical circuit in the visual system at the same time. In the presence of feedback, any perceptual reasoning such as shape computation that requires high resolution and spatial precision may be expected to show up in the single cell recording in V1.

What would be the strategies employed by the visual system to encode figural shapes? Shapes in nature are complex and do not have easy geometry. Biological shapes, in particular, with flexible joints, vary drastically in movement and different postures. Blum⁹, in the 70s, observed that with view point changes and motion, the edge-based description can change dramatically while the region-based description such as the skeleton remain relatively stable.

Figure 2. The square and the strip figures were defined by texture contrast. Both *positive* and *negative* stimuli were tested: a *positive* stimulus in which the figure was composed of the texture of the preferred orientation of the cell and a *negative* stimulus in which the figure's texture was of the orthogonal orientation. The width of the figure was selected to be 4-6 times that of the classical receptive field. Typically, the width of the figure was 4° at about 3-4° eccentricity, where the RF were about 0.7-1.2° in diameter. The monkeys fixated within a 0.3° fixation window while the stimulus was presented on the screen for 350 msec. The texture defined figure was presented in a randomized series of sampling positions relative to the cell's receptive field (see reference 2 and 14 for details in method). The responses of a vertically oriented V1 neuron to the different part of the figures 100-300 msec after stimulus onset were shown in column 2. The abscissa is the distance in visual angle from the RF center to the center of the figure. The solid lines in these graphs indicate the mean firing rate within the 100-300 msec time window, and the dashed lines depict the envelopes of the standard errors. The dots on the solid lines are the data points. To examine the prevalence of these central peaks, we analyzed a population of 41 vertically oriented cells in response to the positive and negative square figures along a horizontal cross-section in the middle of the figure. Column 3: The histograms of the percentage response enhancement at the center of the figure relative to other interior points within the figure showed that there was a general bias toward extra enhancement at the center. A percentage medial enhancement was computed from $(R_0 - R_a)/R_a$, where R_0 is the response at the center (i.e. position=0) and $R_a = [\min(R_1, R_2) + \min(R_{-1}, R_{-2})]/2$ is the averaged valley response at the adjacent positions. 13 out of these 41 cells showed at least 30 percent extra enhancement at the center of the figure. Column 4: The average normalized response of these 13 cells, each normalized by its maximum firing rate, revealed response peaks at the center of both the positive and negative squares and strips (fig 2 column 4).

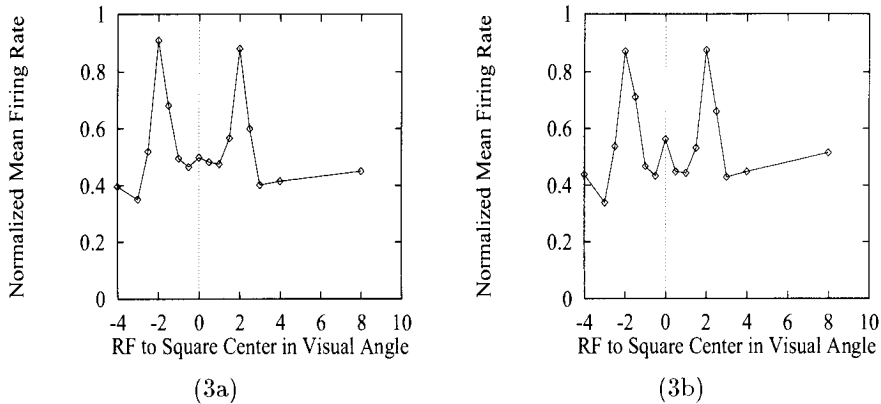


Figure 3. (a) The population average of the net normalized responses of the 37 vertically oriented neurons to square figures showed a considerable enhancement at the boundary and 25 percent enhancement within the figure, with a hint of extra central enhancement. This was computed by summing the response of each individual V1 neurons to *positive* and *negative* squares at the corresponding positions and then normalized the resulting net response by its maximum. (b) The population average of the net normalized responses of the 13 neurons that showed the extra central enhancement. Note that the average response of the rest of the population, apart from the enhancement at the boundary, was generally flat within the figure.

He therefore proposed to describe the ‘infinite’ geometrical variability using the skeleton and a small finite set of shape primitives. These ideas have been further developed and beautifully utilized by Zhu and Yuille¹⁰ in their *flexible object recognition and modeling system* (FORM).

Recent psychophysical evidence suggested that the skeleton transform might play a role in the biological vision^{12–13}. In particular, Kovacs and Julesz¹² studied the contrast-sensitivity threshold to the perception of Gabor filter patches on a gray background with a surround of many randomly oriented similar patches. They arranged some of the Gabor patches so as to form closed contours of various shapes. This caused a marked increase in sensitivity at the points of the medial axis of their closed contours. In neurophysiological studies, Lee *et al.*^{4,14} had found the figure-ground enhancement had certain spatial structure: the responses were particularly accentuated at the figural boundaries for most V1 neurons, and at the medial axis of the figure for a subset of the neurons, as shown in fig 2.

Since the responses within the positive figure was partly due to the tuning properties of cells, we can compute a texture-independent net response by summing together the responses of each cell to the positive and the negative figures, and then normalizing it by its observed maximum firing rate. The computed net response (shown in fig 3a) showed that for this population of cells, on the average, the cells responded 25 % better when they were inside the figure than when they were outside the figure, with only a hint of extra enhancement at the figural center. Some individual cells showed very significant response peaks at the center of the figures (fig 2, col 2). The population histograms of the percentage response enhancement at the center of the figure relative to other interior points within the figure showed that there was a general bias toward extra enhancement at the center. The average net response of the subset of neurons (N=13/43 cells) that were positive for medial axis (as described in fig 2) showed that the figure-ground enhancement was concentrated at the center for this subset of neurons (fig 3b).

For texture strips, the figure-ground enhancement was significantly less. The central enhancement histogram still showed a bias toward positive extra enhancement at the center.

We tested the responses of the cells to strips of different widths and found that collectively, at each eccentricity, there were cells sensitive to the center of strips of different widths. However, individual neurons only showed central peaks for a narrow range of widths (fig 4).

In another experiment (Lee et al 1997), we observed that the central enhancement at the center of uniformly colored disks, e.g. a black disk in a white background or vice versa. This suggests that the central enhancement effect is of a more general nature. It is not unique to texture figures nor can it be simply an effect of intracortical lateral inhibition or disinhibition. These data suggest that V1 might play a role in the computation of figure-ground (figural enhancement) and the medial axis (the central peaks). Abstractly, the medial axis could be produced by a conjunction of three features: two bounding border segments with a homogeneous surface in between (fig 5). One mechanism for computing the medial axis of a figure was the grass fire algorithm originally proposed by Blum⁹, which mimics a *grass fire* propagating in from the boundary of the figure; the points of collision form the skeleton of the figure. However, the grass fire spreads in both direction (in and out of the figure). To ensure only the medial axis of the figure is computed, only the borders belonging to the surface of the figure (intrinsic border) is allowed to ignite the fire within the surface it belongs to. von de Heydt et al's finding¹⁶ of border belongingness signal in V2 and Zipser's⁵ finding on amodal surface completion effect lend further support to these ideas.

These data, together with other works^{1-5,16-17}, inspire a possible reinterpretation of the classical paradigm: V1 is not just the first stage of visual processing, but might be considered to be a *high-resolution buffer* that participates in high as well as low level visual algorithms. Any computation that requires high acuity, resolution, and spatial specificity such as finding object boundaries and medial axes would refer back to V1 for image details and spatial registration. Individual cells in V1 are therefore involved in the computation and representation of many kinds of global perceptual structures rather than being a dedicated detector of a particular local feature. It is possible that different types of information coexist or are multiplexed in the spike trains of individual neurons through synchronized firings with different neuronal assemblies. For V1 to play the role we are sketching, it is essential that feedback from extrastriate areas play a major role in driving V1. Consistent with this hypothesis, psychophysical¹⁸ and PET/MRI studies¹⁹ have shown that V1 activity can be caused or influenced by mental imagery. Recent neurophysiological evidence²⁰ also suggests some context-sensitive surround effects can be eliminated by deactivating V2. Furthermore, the observation that these higher order effects were observed at long latencies suggested they might be a consequence of the continuous interaction between V1 and the extrastriate cortices.

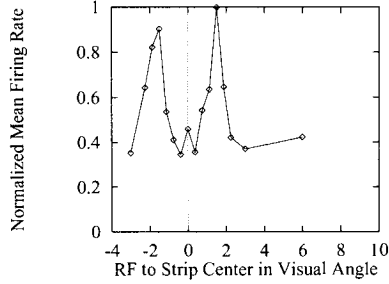
ACKNOWLEDGMENTS

We are grateful to P.H. Schiller, W. Slocum, V. Lamme, K. Zipser, D. Pollen A. Tolias, J. Mazer, T. Moore for insightful discussion and technical assistance. This research is supported by a McDonnell-Pew grant to T.S. Lee, a NSF grant DMS-93-21266 to D. Mumford, and a NIH grant EY00676 to P.H. Schiller.

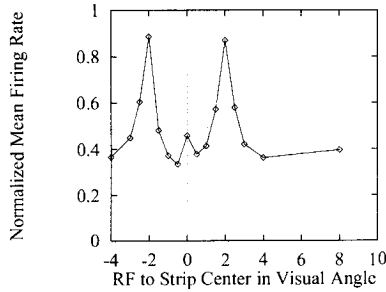
REFERENCES

- [1] Knierim, J.J. & Van Essen D.C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophysiology*, 67, 961-980.

Width = 3 deg ($N = 3/11$)



Width = 4 deg ($N = 16/36$)



Width = 6 deg ($N = 2/10$)

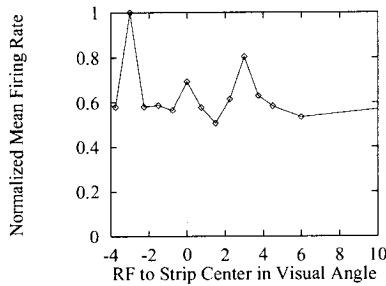


Figure 4. Individual neurons at V1 were sensitive to center of the strips of a narrow range of width. Collectively, the neurons were sensitive to the medial axis of strips of different widths. Shown here are the average normalized net responses of neurons which show central peaks for strips of 3 different widths. 3 out of 11 cells tested for 3° width strip, 16 out of 36 cells tested for 4° width strip, and 2 out of 10 cells tested for 6° width strip show 30 percent extra enhancement at the medial axis of the strip. Because of the difficulty in maintaining a neuron for more than a thousand trials, we usually could test the neurons' response to strips of only one or two widths.

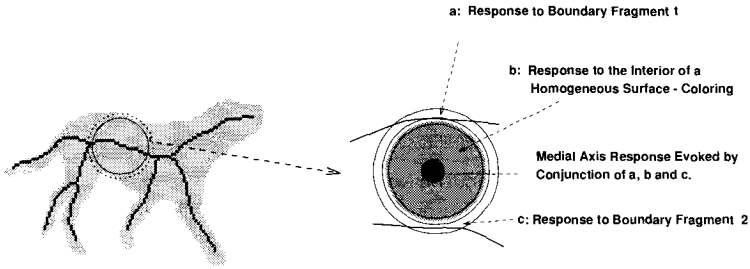


Figure 5. The figure illustrates how a cell may be constructed so that it fires when located on the medial axis of an object. The conjunction of 3 properties has to be present: at least 2 distinct boundary points on a disk of a certain radius and the homogeneity of surface qualities within an inscribing disk. Such a response is highly nonlinear, but can be robustly computed⁹⁻¹³.

- [2] Lamme, V.A.F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neuroscience*, *10*, 649-669.
- [3] Zipser, K. Lee, T.S. Lamme, V.A.F. & Schiller, P. (1994) The spatial extent of contextual modulation in macaque V1. *Soc of Neuro science Abstract*, **20** 608.7.
- [4] Lee, T.S., Mumford, D. & Schiller, P.H. (1995). Neuronal correlates of boundary and medial axis representation in primate striate cortex. *Invest. Opth. Vis. Sci*, *36*, 477.
- [5] Zipser, K. Lamme, V.A.F. & Schiller, P.H. (1996) Contextual modulation in primary visual cortex. *J. Neuroscience* **16**, *22*, 7376-7389.
- [6] Ullman, S. (1984). Visual routines. *Cognition*, **18**, 97-159.
- [7] Paradiso, M.A. & Nakayama, K., (1991) Brightness perception and filling-in *Vision Research*, **31**, 1221-1236.
- [8] Gray, C.M. & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Nat. Acad. Sci. USA*, *86*, 1698-1702.
- [9] Blum, H. (1973). Biological shape and visual science. *J. Theoretical Biology*, *38*, 205-287.
- [10] Zhu, S.C. & Yuille, A.L. (1995) FORMS: a flexible object recognition and modelling system. *Int. J. of Comp. Vis.* 187-212.
- [11] Ogniewicz, R. (1994) Skeleton-space: a multiscale shape description combining region and boundary information. *Proc. Conf. On Computer Vision and Pattern Recognition* 746-751.
- [12] Kovacs, I. & Julesz, B. (1994). Perceptual sensitivity maps within globally defined visual shapes. *Nature*, *370*, 644.
- [13] Burbeck, C.A. & Pizer, S.M. (1995). Object representation by cores: identifying and representing primitive spatial regions. *Vision Research*, *35*, *13*, 1917-1930.
- [14] Lee, T.S. (1996). Neurophysiological evidence for image segmentation and medial axis computation in primate V1. *Computational Neuroscience*, Bower, J. Eds., 373-378, Academic Press.
- [15] Li, C.Y. & Li, W. (1994). Extensive integration field beyond the classical receptive field of cat's striate cortical neurons - classification and tuning properties. *Vision Research*, *34*, 2337-2355.
- [16] Zhou, H. Friedman, H. von de Heydt, R. (1996) Edge assignment in cells of monkey area V2. *Invest. Ophthalmol. Vis. Sci.* **37**, 904.
- [17] Gilbert, C.D. Das, A. Ito, M. Kapadia, M. & Westheimer, G. (1996). Spatial integration and cortical dynamics. *Proc. Nat. Acad. Sci. USA*. *93*, 615-622.
- [18] Ishai, A. & Sagi, D. (1995). Common Mechanisms of Visual Imagery and Perception. *Science*, *268*, 1772-1774.
- [19] Kosslyn, S., Thompson, W.L., Kim, I.J. & Alpert, N.M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, *378*, 496-498.
- [20] James, A.C., Hupe, J.M., Lomber, S.L., Payne, B., Girard, P. & Bullier, J., (1995). Feedback connections contribute to center surround interactions in neurons of monkey area V1 and V2. *Soc. Neuroscience Abstract*, **21**, 359.10.