Medial axis generation in a model of perceptual organization

Diego Ardila, Stefan Mihalas*, Rudiger von der Heydt, Ernst Niebur
Zanvyl Krieger Mind/Brain Institute
Johns Hopkins University
Baltimore, Maryland 21218
*Current Address: Allen Institute for Brain Science, Seattle WA 98103

Abstract—**Axial skeletons are promising intermediate representations of shape which are used in machine vision and higher level theories of vision to provide a concise and intuitive description of the shape. More recently, physiological correlates of axial skeletons have been reported. We show that a stable approximation of the medial axis can be generated based on an existing neurally plausible model of perceptual organization.**

I. INTRODUCTION

The shape skeleton is a sparse representation for shape. The notion was first introduced [1] as the Medial Axis Transform (MAT). The MAT is the set of the centers of all circles that meet the following two conditions:

1) Of all co-centric cirles that intersect the shape boundary, it is the smallest.
2) The circle intersects the boundary at 2 or more points.

Note that if (1) is met, and the point of intersection on the shape boundary has a defined curvature then the circle will be tangent to the shape boundary. In ref. [1] a ”grassfire” algorithm was used to calculate the MAT. Wave fronts propagate from the shape boundary, and points of collision, or ”shocks” are part of the medial axis. The analogy here is that if the fires were started in a grass field in the shape of the object boundary, then the medial axis would be the points where the fires meet.

The MAT is likely too unstable and too sensitive to noise for representation of shape in a biological nervous system. Any indentation in a shape boundary, however small, will lead to an additional ”rib” extending out from the medial axis to the tip of the indentation. A practically more desirable representation is the maximum a posteriori (MAP) skeleton proposed by ref. [2]. In this Bayesian formulation, branches are added to the skeleton only if the improved description of the shape outweighs the additional complexity. This leads to a more stable skeletonization than the MAT which more closely matches the intuitive notion of a shape skeleton.

The shape skeleton has been used extensively for object recognition, because it is invariant to several transformations of within-class objects [3]. It also plays a role in several theories of higher level vision [4], [5]. Psychophysical correlates of medial axis computation include increased contrast sensitivity along medial points in figures [6]. There is evidence that already cells in primary visual cortex, area V1, show increased sensitivity along medial points [7], [8]. Corresponding modulation of the mean firing rate of the neurons was observed approximately 100ms after stimulus onset.

It is therefore of interest to study how the shape skeleton can be computed and represented in the brain. A recent model [9] relies on the assumption of perfectly synchronized onset of border ownership selectivity (see ref. [10] and below for a definition of border ownership and its neural representation in visual cortex). This assumption is problematic because the visual system can perform figural binding even in cases when temporal asynchrony in the stimulus is large [11]. This same study also found that small changes in spatial presentation are more likely to cause disruptions in figural binding. We therefore set out to develop a neurally plausible model of shape skeleton generation that does not require high-precision temporal correlations.

The representation of perceptual objects in the visual system of primates has long been a subject of intense study. An important insight was obtained by neurophysiological recordings that showed that figure-ground assignment is at least partially encoded by “border-ownership selective” neurons in early (striate and extrastriate) visual cortex [10]. The study demonstrated that many neurons in these areas (indeed a majority in area V2) have a preferred side of figure, responding with a higher mean firing rate when the foreground object (figure) is on one side of its receptive field than on the other. The modulation starts about 70ms after figure onset (10-25ms after initial neuronals responses in these cortical areas). This short latency is incompatible with a modulation by (slow, unmyelinated) intra-areal connections. Instead, the authors argued that contextual input is provided by fast, white-matter connections from “grouping cells” (G cells) which bias border ownership cells and thus generate their context-dependent responses. A computational model based on this mechanism [12] could qualitatively explain the neurophysiological results. More recently, Mihalas et al [13] showed that this cell layer can also be used to sharpen a broad attentional spotlight to the lower-level and higher resolution features of a specific object. In the present study, we show that the same mechanisms also generate shape skeletons.

II. METHODS

An overview of the network structure of our model is shown in Figure 1. The input to the model is a grayscale image.

978-1-4673-3140-1/12/$31.00 ©2012 IEEE

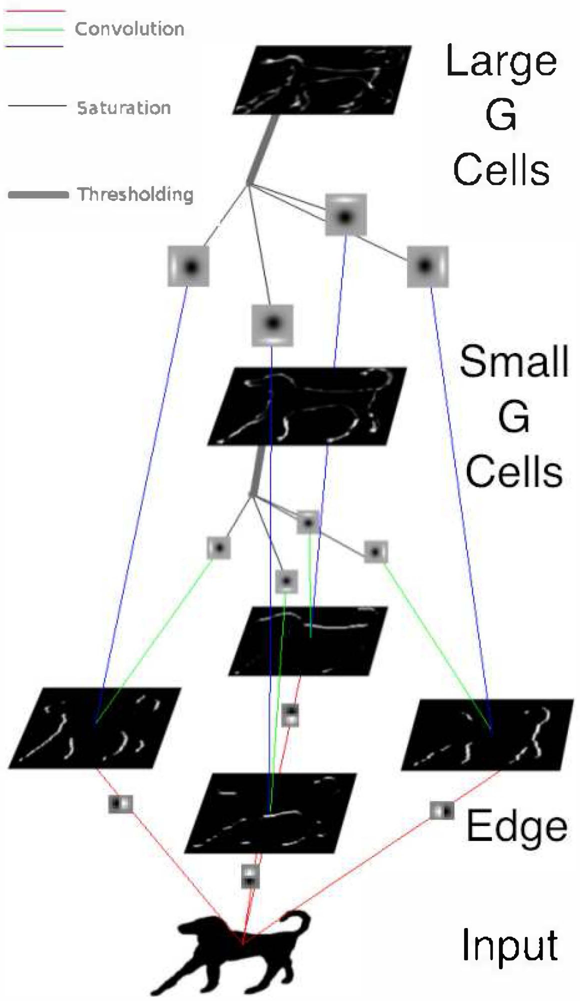

Fig. 1. Network Structure

Neurons are organized in two-dimenstional arrays which interact through the application of convolution masks. Sets of arrays with similar convolution masks can be considered as layers of one type of cell. The first layer consists of edge cells which respond to oriented edges in their receptive fields (independent of context), with each array being characterized by its orientation. An edge cell receives input from a filter that consists of the difference between two, offset, two-dimensional Gaussian functions, a shape similar to a Gabor filter. There are 16 orientations of edge cells in our model, with each of the arrays receiving input via the same mask rotated at different angles. Generation of border ownership selectivity by feedback from grouping cell input has been characterized in previous work [12], [13] and is not the focus of the present study. Therefore, for simplicity, we assign border-ownership explicitly, making input to the grouping cells strictly feed-forward and thus increasing its computational performance. This is done by using figures that are always darker than the background, and using edge cells that respond to one contrast only, instead of pooling several different contrasts for each orientation.

The next layer consists of grouping cells, with each array characterized by its radius, R. Each grouping cell receives excitatory input from all 16 orientations of edge cells. If the masks are summed over all orientations, the result resembles a circle with radius R (see also ref. [12]). This circle is composed of 16 "arcs" modeled as Gaussian functions, akin to observed response functions in posterior inferotemporal cortex [14]. Each arc is the mask used to calculate the input from one orientation-selective edge cell. More precisely, the mask used to calculate the input from an edge cell array with orientation θ to a grouping cell array with radius R is an elongated Gaussian with a major axis (larger length constant in the denominator of the exponential) orthogonal to θ, and its mean offset from the center of the mask by a vector of length R and orientation $\theta + \pi/2$. The input from each orientation saturates at a low value, and separately from other orientations. A Hill equation with $N = 1$ is used. Neurally, this nonlineariy is easily implemented when inputs from similar orientations make synapses on the same dendrite. After this first layer of nonlinearities, another nonlinearity in the form of a Hill equation with $N = 3$ is applied. This 2-layer processing scheme results in the grouping cells responding only if they receive input from 2 different orientation channels.

Grouping cells also receive inhibitory input from this same orientation via a mask that is on the opposite side of the circle. More specifically, input is from a Gaussian orthogonal to θ which is offset from the center of the mask by a vector of length R and orientation $\theta + \pi$. Another way to picture this inhibition is that it is the same as the excitation, but coming from edge cells of the opposite polarity, since rotating the edge cell mask by π is equivalent to changing its sign.

Grouping cells of different radii interact via inhibition of two kinds: cocentric and cotangential inhibition. Cocentric inhibition is exerted by grouping cells with small receptive field (referred to as "small" grouping cells in the following) on those with larger receptive fields ("large" grouping cells). The mask used is a two dimensional Gaussian. Smaller grouping cells thus inhibit larger grouping cells that have the same center. Cotangential inhibition acts only from large to small grouping cells. The mask used for inhibition from radius R_1 to R_2 is the value of a normal distribution with mean $R_1 - R_2$ evaluated at each point's distance from the center of the mask. Maximal inhibition is therefore received by grouping cells which share a tangent with a larger grouping cell. The nonlinearities in the grouping cells and the co-tangential connection pattern help ensure that the grouping cell is active only if it intersects the boundary at 2 or more points, and the co-centric inhibition ensures that it is the smallest of all circles to do so.

III. Results

Using the original figures from the study by Feldman and Singh [2], we tested the ability of the algorithm to generate the medial axis. Medial axes very similar to the algorithm used by [2] were recovered, for examples see Figure 2. There was slight activity along minor ribs that the Feldman and Singh

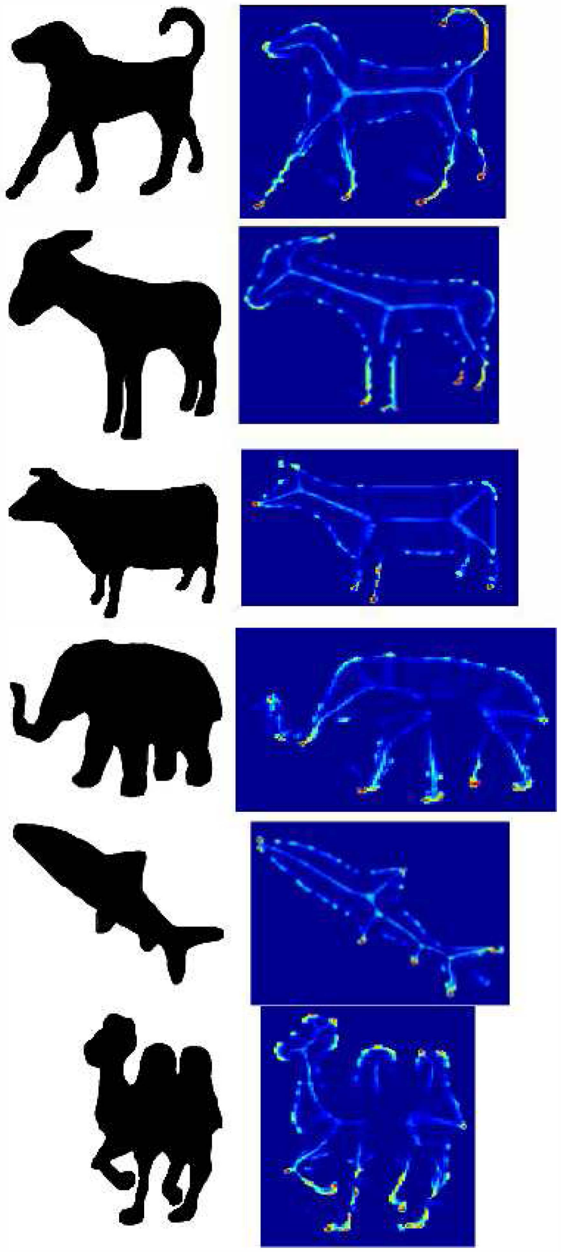

Fig. 2. Model Results. Example figures (from ref. [2]) are shown left. On the right is the corresponding activity of G cell populations in our model, summed over all scales. Activity along the figure contours is that G cells at the smallest scales.

algorithm avoided completely which was in most cases substantially lower than that on the main medial axis. For greater stability and invariance with respect to minor modifications of the figure borders, it is important that not every single point of the medial axis be recovered, and the "fuzziness" inherent to the Gaussian connection patterns ensures that small changes in the border do not create very large changes in the medial axis. The model has the most difficulty detecting the medial axis at points where the circle corresponding to that point of the medial axis is tangent to a point of high curvature. At these points, edge activity is strong mostly for orientations different from the orientation that would be tangent to the circle corresponding to the point on the medial axis in question. Since the grouping cells are activated by edges tangent to the circle, it is difficult for these corners to properly activate the grouping cells. An example can be seen in the elephant silhouette (4th picture from the top in Figure 2) where the grouping cell activity is weak in the center of the body. The reason is the location and scale of these grouping cells requires input from horizontally oriented edge cells from both top and bottom (and, ideally, from vertically oriented edge cells from left and right) but strong input is only available from the top. At the bottom, the orientation of the legs is mainly vertical and therefore only very few edge cells with horizontally oriented filters are activated in this area. However, corners will always generate slight activity at the orientation tangent to the circle, so it is still possible to detect the medial axis in these cases.

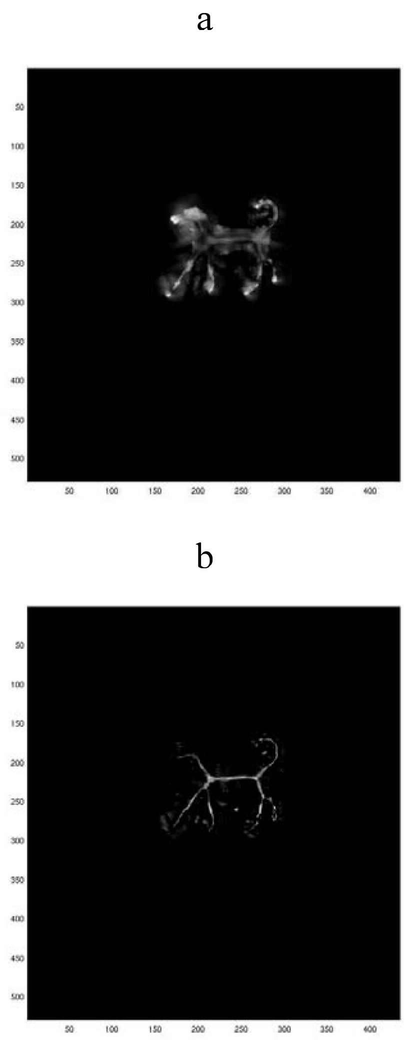

Fig. 3. Dynamics of G cell activity. (a) Initial, (b), Final summed G cell response to top picture in Fig 2.

The dynamics of the model are summarized in the example shown in Figure 3. The initial response in Figure 3a, is broad, non-specific and "blob-like." Qualitatively speaking, this type of response may be useful for calculating saliency and assigning border ownership. There is still a preference for the shape skeleton due to the non-linearities in the model, which prefer G cells tangent to two points on the figure boundary. Through time, the co-tangential and co-centric inhibition sharpen the G cell response to the shape skeleton seen in Figure 3b, a response useful for object recognition.

IV. Conclusion

We show that mechanisms of perceptual organization and top-down attention [12], [13] naturally generate activity patterns that correspond to a shape skeleton. In addition, we find that the initial dynamics of the model are more akin to a proto-object representation that could be used to calculate saliency. As time goes on, inhibitory connections sharpen the shape skeleton. The dynamics also suggest that G cells will show a more proto-object like representation when lateral inhibition is weaker, and a more shape-skeleton like representation when lateral inhibition is stronger. We suggest that these mechanisms are useful for object recognition in the primate visual system.

Acknowledgment

This work is supported by the Office of Naval Research grant N000141010278 and by the National Institutes of Health Grants R01EY016281-02 and R01EY02966. It is also supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of the Interior (DOI) contract number D10PC20022. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained hereon are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI, or the U.S. Government.

References

[1] H. Blum, "Biological shape and visual science (part i)," *Journal of Theoretical Biology*, vol. 38, no. 2, pp. 205 – 287, 1973.

[2] J. Feldman and M. Singh, "Bayesian estimation of the shape skeleton," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, p. 18014, 2006.

[3] K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker, "Shock graphs and shape matching," *International Journal of Computer Vision*, vol. 35, no. 1, pp. 13–32, Nov. 1999.

[4] I. Biederman, "Human image understanding: Recent research and a theory," *Computer Vision, Graphics, and Image Processing*, vol. 32, no. 1, pp. 29 – 73, 1985.

[5] D. Marr and H. K. Nishihara, "Representation and recognition of the spatial organization of three-dimensional shapes," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 200, no. 1140, pp. 269–294, Feb. 1978.

[6] I. Kovacs and B. Julesz, "Perceptual sensitivity maps within globally defined visual shapes," *Nature*, vol. 370, pp. 644–646, 1994.

[7] T. S. Lee, "Neurophysiological evidence for image segmentation and medial axis computation in primate V1," in *Computational Neuroscience*, J. Bower, Ed. Academic Press, 1995, pp. 373–8.

[8] T. S. Lee, D. Mumford, R. Romero, and V. A. Lamme, "The role of the primary visual cortex in higher level vision," *Vision Research*, vol. 38, no. 1516, pp. 2429 – 2454, 1998.

[9] Y. Hatori and K. Sakai, "Representation of medial axis from synchronous firing of border-ownership selective cells," in *Neural Information Processing*, ser. Lecture Notes in Computer Science, M. Ishikawa, K. Doya, H. Miyamoto, and T. Yamakawa, Eds. Springer Berlin / Heidelberg, 2008, vol. 4984, pp. 18–26.

[10] H. Zhou, H. S. Friedman, and R. V. D. Heydt, "Coding of border ownership in monkey visual cortex," *Journal of Neuroscience*, vol. 20, pp. 6594–6611, 2000.

[11] M. Fahle and C. Koch, "Spatial displacement, but not temporal asynchrony, destroys figural binding," *Vision Research*, vol. 35, no. 4, pp. 491 – 494, 1995.

[12] E. Craft, H. Schtze, E. Niebur, and R. von der Heydt, "A neural model of figureground organization," *Journal of Neurophysiology*, vol. 97, pp. 4310–4326, 2007.

[13] S. Mihalas, Y. Dong, R. von der Heydt, and E. Niebur, "Mechanisms of perceptual organization provide auto-zoom and auto-localization for attention to objects," *Proceedings of the National Academy of Sciences*, vol. 108, no. 18, pp. 7583–7588, Apr. 2011.

[14] S. Brincat and C. Connor, "Underlying principles of visual shape selectivity in posterior inferotemporal cortex," *Nature Neuroscience*, vol. 7, pp. 880–886, 2004.