A neural model for perceptual organization of 3D surfaces

Brian Hu, Rüdiger von der Heydt, Ernst Niebur
Zanvyl Krieger Mind/Brain Institute
Johns Hopkins University
Baltimore, Maryland 21218

***Abstract*—Perceptual organization is the process by which the visual scene is structured into coherent units suitable for further processing and selection. Having been primarily studied in 2D, here we present a neural model for perceptual organization of 3D surfaces. We demonstrate that our model is able to reproduce several key psychophysical results, including the spread of visual attention across 3D surfaces.**

I. Introduction

Since we live in a complex 3D world, competent interaction with the surrounding 3D scene structure is indispensable to us and our machines. Access to information about surfaces present in the scene allows us to perform a wide variety of tasks, ranging from motor planning (*e.g.* reaching for a cup a certain distance away on the table) to spatial navigation (*e.g.* following directions in a new city environment).

In studying perceptual organization, researchers have traditionally relied on simple 2D stimuli such as oriented bars [1]. Results from these studies provide support for the importance of well-known Gestalt principles [2], [3], for instance that visual elements are grouped together in a way that begins to give meaning to the visual scene (*e.g.* figure *vs.* background). However, it is unclear how well the results from these relatively simple experiments generalize to the 3D objects and scenes regularly encountered in natural settings. Because we act in a 3D world, perceptual organization must also help to arrange 3D information in a way that can guide our actions.

Perceptual organization also provides a structure for selectively attending to groups of objects [4]. Supported by extensive psychophysical data, Nakayama, He, and Shimojo [5] proposed that surface representations play a key role in intermediate-level vision. For example, by selectively attending to a surface in 3D space, subjects can perform efficient search for a conjunction target [6]. In a separate cueing experiment, attention was shown to spread automatically across surfaces [7]. These abilities indicate powerful mechanisms for grouping objects into surfaces in 3D space, and suggest that structuring the world in terms of surfaces might be an ecologically important function. These results also have implications for the internal representation of surfaces, because they imply that the visual scene is processed in a way that preserves its 3D structure. This representation must also be able to bring together information from different sensory modalities (*e.g.* vision, audition, *etc.*), in order to form a common representation of the 3D environment that is useful for an agent's behavioral goals [8].

In addition to being studied by human psychophysical approaches, perceptual organization has been the subject of many studies in the visual system of non-human primates. Many neurons in early visual cortex encode the side to which an object border belongs, a phenomenon known as border ownership [9]. Selectivity for the side of ownership involves integrating global context information about the object. Several models have been proposed [10], [11] to describe how a neuron's border ownership selectivity can be modulated by visual input far away from its classical receptive field with the observed high specificity of object details. One view is that this contextual input is provided by feedback connections from "grouping cells" [11] which bias the activity of border ownership cells and thus generate their context-dependent responses. Mihalas et al. [12] have also shown that grouping cells can direct and sharpen a broad attentional spotlight to the lower-level features of a specific object. In the present study, we extend this grouping framework to 3D space to show how oriented 3D elements can be grouped into planar surfaces.

Currently, we know very little about how surfaces are represented in the brain, and how this representation is computed. Our model sheds light on a possible neural representation of 3D surfaces and relates this model to previous psychophysical results.

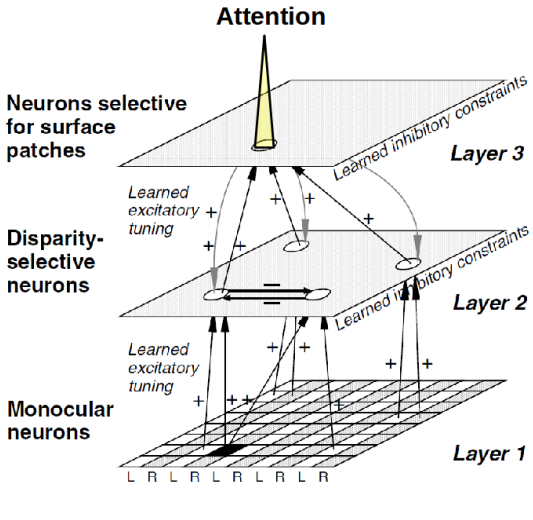

Fig. 1. Network structure (adapted from ref. [13]).

978-1-4799-8428-2/15/$31.00 ©2015 IEEE

II. Methods

An overview of the network structure of our model is shown in Figure 1. We extend a neural model of visual stereomatching [13] that is conceptually similar to grouping models previously proposed for 2D stimuli [11], [12], [14]. The model contains three layers of neurons. The first layer consists of monocular cells, which respond to visual features (e.g. spots, edges, *etc.*) presented to either the left or right eye. The input to the model consists of pairs of stereo images, as would be seen by the left and right eyes. In our model, we set the image input to a value of unity wherever a stimulus is present and zero elsewhere.

The second layer consists of binocular cells, which receive excitatory input from monocular cells. These cells are tuned to a certain disparity based on a fixed spatial weighting between the left and right monocular cells, analogous to the disparity-selective binocular neurons in visual cortex of monkeys [15], [16] and cats [17], [18]. Lateral inhibition between cells representing different disparities along the same left- or right-eye line of sight reduces potential false matches [19].

The third and final layer consists of planar grouping cells which receive excitatory input from populations of disparity-selective cells. Receptive fields of the planar grouping cells are relatively broad and non-specific, resembling surface patches with a certain range of depth and orientation selectivity in 3D space. These cells may correspond to neurons found in parietal cortex, which have been shown to be selective for the tilt and slant of planar surfaces [20]. In our model, we used a total of 15 planar grouping cells, which was sufficient to "tile" the whole 3D visual scene. Planar grouping cells compete with each other through lateral inhibition, which helps to select the best possible interpretation of surfaces within the scene. Additionally, planar grouping cells send reciprocal feedback connections to the disparity-selective cells that define their surface, akin to the relationship between grouping cells and border ownership cells in models of 2D scenes [11], [12]. To avoid uncontrolled feedback excitation, feedback is multiplicative and only amplifies existing feedforward excitation. Selective attention is modeled as an additive input to those planar grouping neurons representing attended objects. This attentional modulation input is set to a value of 0.25 of the sensory input.

All model neurons are simulated as single compartment units with an activity that is modeled as a continuous variable (rate coding). These units are zero-threshold, linear neurons which receive excitatory and inhibitory current inputs. The activity of the units is determined by a set of coupled, first-order nonlinear ordinary differential equations, which can be solved in MATLAB (MathWorks) using standard numerical integration methods (Euler, Runge-Kutta, *etc.*)

III. Results

Figure 2 illustrates the experimental paradigm of He and Nakayama [7]. In Figure 2A, subjects had to search for the odd-colored target in the middle depth plane; planes are outlined by rectangles that were not visible to the observers.

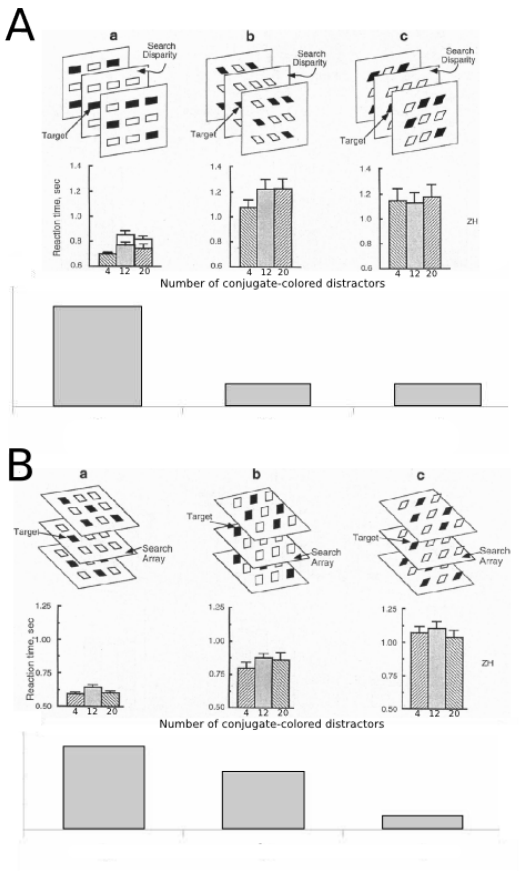

Fig. 2. Psychophysical and model results (adapted from ref [7]). For each trial type (A or B), the top row shows the different stimuli, the middle row shows representative reaction times, and the bottom row shows the degree of attentional modulation of disparity-selective cells on the attended plane. Increase in activity is assumed to be inversely proportion to reaction times.

The target was unique in this search plane but visually identical distractors were present in other depth planes. In A-a, objects are aligned with the search plane while in A-b and A-c, they are slanted out of the plane. The middle row in A shows measured reaction times for three different numbers of distractors. They are significantly shorter when the objects are coplanar with the search plane (A-a) compared to when they are not aligned with the plane (A-b and A-c).

In our model, we assume that the visual system can selectively direct attention towards a specific surface by providing additional excitatory input to the grouping cell that represents this surface. As shown in Figure 1, activity from the grouping cell selectively feeds back to all objects on that surface. In the case of Figure 2A-a, the grouping cell corresponding to the middle fronto-parallel plane receives this attentional input. Activation of the disparity-selective cells in the search plane, shown in Figure 2A-a, bottom row, is thus high. Among the objects in the search plane, the target has a unique color, which results in efficient search and the target being identified immediately. Reaction times are difficult to simulate in detail, therefore the increase in mean activity of disparity-selective cells on the attended plane due to attentional modulation is plotted instead, which is assumed to be inversely proportional

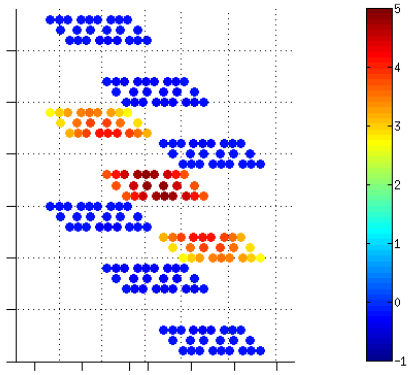

Fig. 3. Spread of attention across surfaces. When attention is directed to the center slanted plane (as in the experiment in Figure 2B-a), attention enhances the activity of all cells along the surface (red), while suppressing the activity of cells belonging to other surfaces (blue).

to reaction times. The high activation level, bottom row, translates thus into short reaction times, middle row.

In contrast, in Figure 2A-b, the search plane is no longer a well-formed surface but contains objects that are slanted backwards. Directing attention to the middle fronto-parallel plane then has little effect on the disparity-selective cells in the search plane. Search therefore cannot occur entirely within a single plane of coplanar, grouped elements and becomes inefficient, with much higher reaction times (middle row). These long reaction times are reflected in the low activation level of the disparity-selective cells (bottom row). Figure 2A-c shows the analog result for figure elements that are slanted forward rather than backwards, as in A-b. Again, reaction times are long and population activity is low.

The result is not restricted to fronto-parallel planes, as similar reaction time results were also found for slanted planes in Figure 2B. When subjects are instructed to search in a plane that is coplanar with the orientation of the figure elements (middle plane in B-a), search is fast (B-a, middle row). In contrast, when the figure elements do not align with the search plane (top row in B-b and B-c), reaction time is increased (middle row). Again, under the assumption that reaction times are inversely related to reaction times, the model reproduces human behavior (bottom row).

Attention to grouping neurons is then able to select sets of objects organized in planes. Reaction times were fastest when the search array was a well-formed surface defined by locally coplanar elements. When search array elements were slanted away from this surface, reaction times increased. These results suggest that attention is linked to and spreads across perceived surfaces, which organize the visual scene (Figure 3).

IV. Conclusion

Using a simple model of perceptual organization in 3D, we are able to reproduce psychophysical results from a visual search task that required allocation of selective attention to surfaces within the scene. The same grouping cells which organize the scene into planes also act as "handles" for top-down selective attention, enhancing the activity of coplanar elements belonging to the plane. Competition between grouping cells results in surface enhancement of the plane corresponding to the attended grouping cell, and suppression of other planes within the scene. Our proposed surface representation aids visual processing by providing a critical link between low-level visual features and high-level object representations.

Acknowledgment

This work is supported by the Office of Naval Research grant N000141010278, the National Institutes of Health grant R01EY016281-02, and the Visual Neuroscience Training Program fellowship (T32EY07143).

References

[1] S. E. Palmer, "Perceptual organization in vision," *Stevens' handbook of experimental psychology*, 2002.

[2] K. Koffka, *Principles of Gestalt psychology*. New York: Harcourt-Brace, 1935.

[3] M. Wertheimer, "Untersuchungen zur Lehre von der Gestalt II," *Psychol. Forsch.*, vol. 4, pp. 301–350, 1923.

[4] A. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, pp. 97–136, 1980, pMID: 7351125.

[5] K. Nakayama, Z. J. He, and S. Shimojo, "Visual surface representation: a critical link between lower-level and higher-level vision," in *Visual Cognition: An Invitation to Cognitive Science*, 2nd ed., S. Kosslyn and D. Osherson, Eds. The MIT Press, 1995, vol. 2, ch. 1, pp. 1–70.

[6] K. Nakayama and G. H. Silverman, "Serial and parallel processing of visual feature conjunctions," *Nature*, vol. 320, pp. 264–265, 1986, pMID: 3960106.

[7] Z. J. He and K. Nakayama, "Visual attention to surfaces in three-dimensional space," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 9, no. 24, pp. 11 155–11 159, 1995, pMID: 7479956.

[8] M. S. Lewicki, B. A. Olshausen, A. Surlykke, and C. F. Moss, "Scene analysis in the natural environment," *Frontiers in psychology*, vol. 5, 2014.

[9] H. Zhou, H. S. Friedman, and R. von der Heydt, "Coding of border ownership in monkey visual cortex," *J. Neurosci.*, vol. 20, no. 17, pp. 6594–6611, 2000, pMID: 10964965.

[10] L. Zhaoping, "Border ownership from intracortical interactions in visual area V2," *Neuron*, vol. 47, pp. 143–153, 2005, pMID: 15996554.

[11] E. Craft, H. Schütze, E. Niebur, and R. von der Heydt, "A neural model of figure-ground organization," *Journal of Neurophysiology*, vol. 97, no. 6, pp. 4310–26, 2007, pMID: 17442769.

[12] 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. 7583–8, 2011, pMC3088583.

[13] J. A. Marshall, G. J. Kalarickal, and E. B. Graves, "Neural model of visual stereomatching: slant, transparency and clouds," *Network: Computation in Neural Systems*, vol. 7, no. 4, pp. 635–669, 1996.

[14] A. F. Russell, S. Mihalas, R. von der Heydt, E. Niebur, and R. Etienne-Cummings, "A model of proto-object based saliency," *Vision Research*, vol. 94, pp. 1–15, 2014.

[15] G. F. Poggio and B. Fischer, "Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey," *J. Neurophysiol.*, vol. 40, pp. 1392–1405, Nov 1977, pMID: 411898.

[16] G. Poggio and T. Poggio, "The analysis of stereopsis," *Ann. Rev. Neurosci.*, vol. 7, pp. 379–412, 1984.

[17] P. Bishop and J. Pettigrew, "Neural mechanisms of binocular vision," *Vision research*, vol. 26, no. 9, pp. 1587–1600, 1986.

[18] I. Ohzawa, G. C. DeAngelis, and R. D. Freeman, "Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors." *Science*, vol. 249, pp. 1037–1041, 1990.

[19] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, 1976, pMID: 968482.

[20] A. Rosenberg, N. J. Cowan, and D. E. Angelaki, "The visual representation of 3d object orientation in parietal cortex," *The Journal of Neuroscience*, vol. 33, no. 49, pp. 19 352–19 361, 2013.