A Neural Model for Grouping 3D Surfaces

Brian Hu and Ernst Niebur

Johns Hopkins University, Mind/Brain Institute

Contact: bhu6@jhmi.edu

OCNC 2014

Introduction

Both in 2D and in 3D, spatially separated figure elements can be grouped together. Grouping by planes in 3D improves the speed of visual search dramatically (He and Nakayama, 1995). Supported by extensive psychophysical data, Nakayama, He, and Shimojo (1995) proposed that surface representations rather than visual features play a key role in organizing intermediate-level vision in a variety of tasks, including texture perception, visual search, and motion perception (*Figure 1*). We hypothesize that the visual system organizes cluttered 3D scenes into planes shaped by the orientations of their elements. **How does the brain represent planes in depth so that we can selectively attend to objects in one plane of a cluttered scene?**

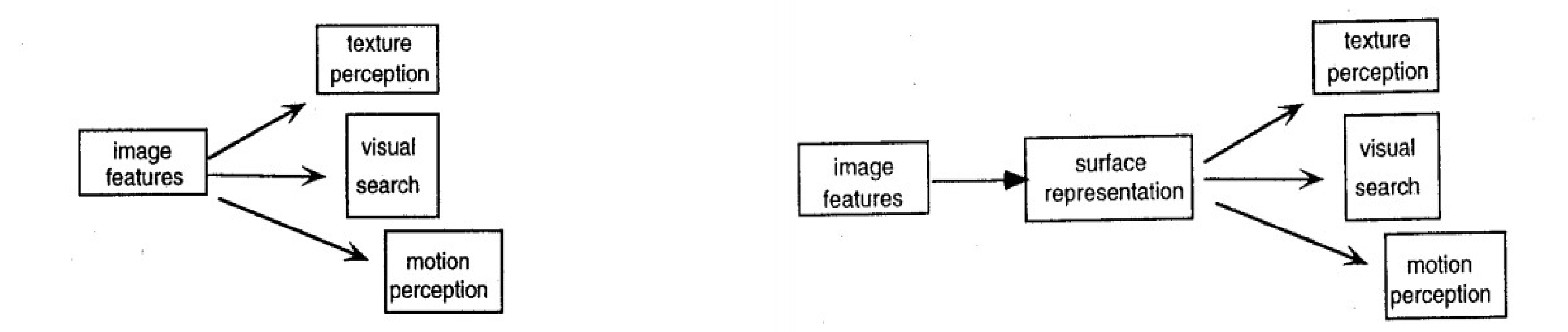


Fig. 1. *Left,* Traditional view where visual features form the basis for perceptual functions. *Right*, New view where surface representation precedes perceptual functions. Figure was adapted from Nakayama et al. (1995).

Methods

We extended a neural model of visual stereomatching (Marshall, Kalarickal, and Graves, 1996) to explain visual search within grouped 3D planes (*Figure 2*). The model equations describing the network are also given below.

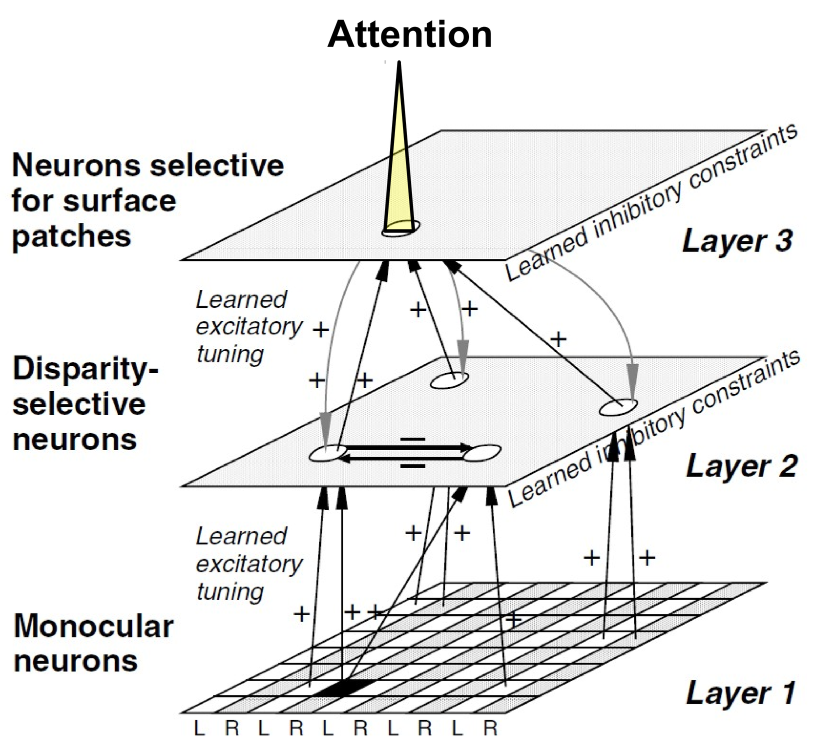


Fig. 2. Layer 1 neurons serve as input and are arranged in left/right ocular dominance columns. Layer 2 neurons are tuned for stimuli at a particular position and disparity. Layer 3 neurons are grouping cells selective for planes in depth. Selective attention acts at the level of grouping cells. Figure was adapted from Marshall et al. (1996).

The network is described as a system of ordinary differential equations, with the dynamics of each layer 2 and layer 3 neuron,

$$\tau f'(t) = -f + [\Sigma W]_+ \tag{1}$$

where f represents the neuron's activity and τ its time constant (= 10^{-2} s), W is the neuron's inputs, and $[\,]_+$ means rectification.

The feedforward weight from a layer 1 to layer 2 neuron, W_{12}, is

$$W_{12} = \exp\left(-\left(\frac{(X_1 \pm D - X_2)^2}{\sigma_x^2} + \frac{(Y_1 - Y_2)^2}{\sigma_y^2}\right)\right) \tag{2}$$

where D is the disparity of the layer 2 neuron.

The feedforward weight from a layer 2 to layer 3 neuron, W_{23}, is

$$e_x = cos(S)cos(T)(X_2 - X_3) + cos(S)sin(T)(Y_2 - Y_3) - sin(S)(D_2 - D_3) \tag{3}$$

$$e_y = -sin(T)(X_2 - X_3) + cos(T)(Y_2 - Y_3) \tag{4}$$

$$e_z = sin(S)cos(T)(X_2 - X_3) + sin(S)sin(T)(Y_2 - Y_3) + cos(S)(D_2 - D_3) \tag{5}$$

$$W_{23} = \exp\left(-\left(\frac{e_x^2}{\sigma_x^2} + \frac{e_y^2}{\sigma_y^2} + \frac{e_z^2}{\sigma_z^2}\right)\right) \tag{6}$$

where S and T are the slant and tilt, respectively, of the layer 3 neuron.

The lateral inhibition weight between two neurons in either layer 2 or layer 3, W_{ij}, is proportional to the amount of overlap in their feedforward inputs,

$$W_{ij} = \sum_{k \epsilon layer(l-1)} min(W_{ki}, W_{kj}) \tag{7}$$

The feedback weight from a layer 3 to layer 2 neuron, W_{32}, is reciprocal and multiplicative in that it only modulates the feedforward input,

$$W_{32} = W_{23}' \tag{8}$$

Selective attention is applied as an additive input to specific layer 3 neurons with a value of 0.25 of the sensory input.

Results

Subjects had to find the odd-colored target within a cued search array (He and Nakayama, 1995). 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 greatly increased. These results suggest that attention is linked to and spreads across perceived surfaces, which organize the visual scene. We used our model to reproduce these results (*Figure 3* and *Figure 4*).

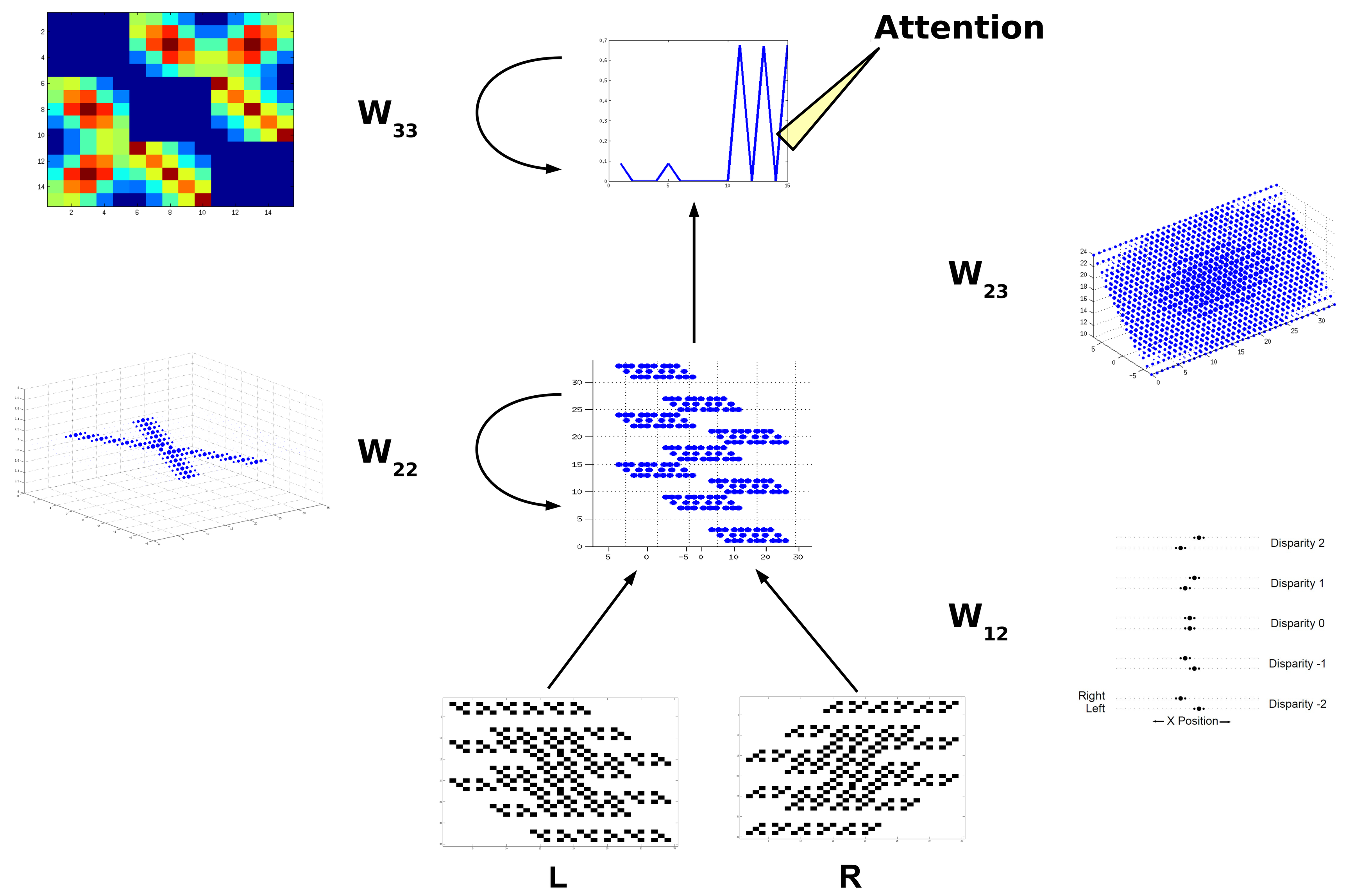
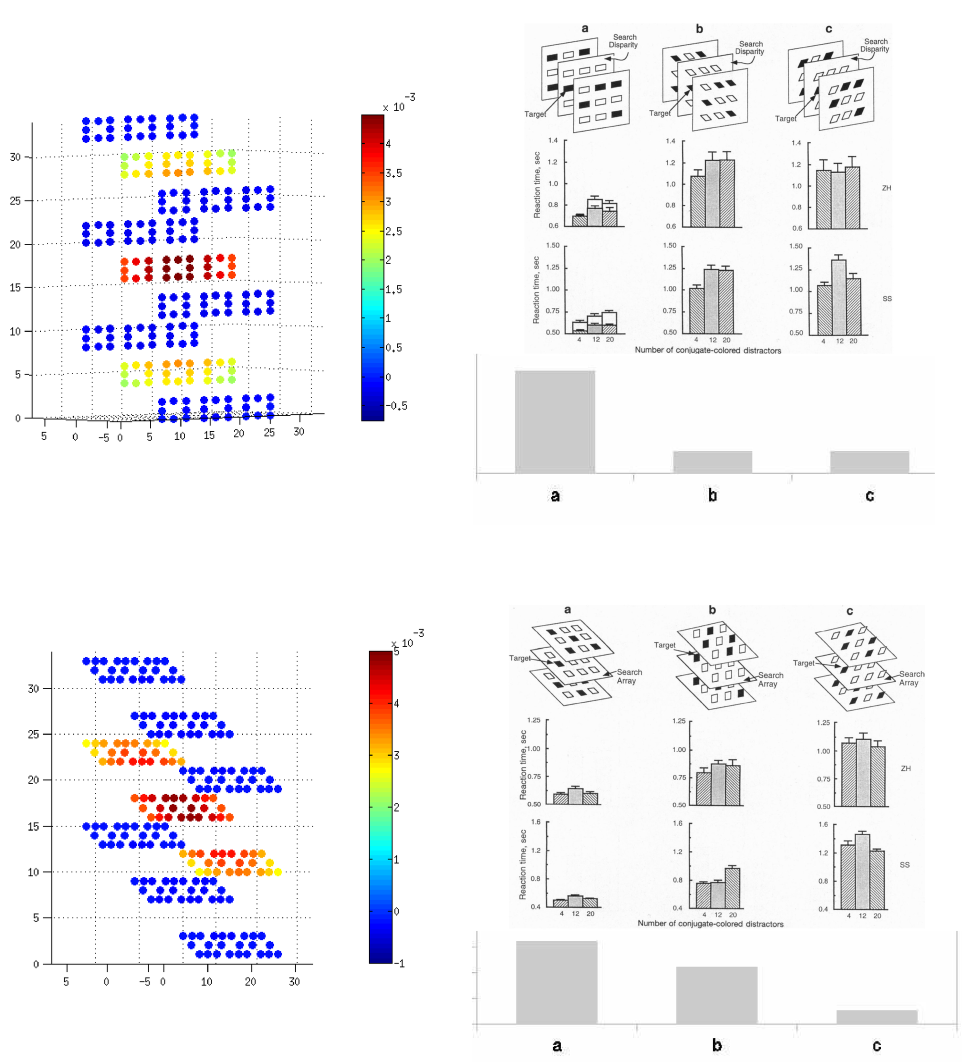


Fig. 3. The grouping model response to an input stereogram.



Surface Modulation Index (SMI)

$$SMI = S_{att} - S_{unatt}$$

Fig. 4. When selective attention is applied to the grouping cell corresponding to the search array, there is a higher surface modulation index when the search array elements are coplanar and on the same surface, corresponding to lower reaction times in the visual search task. Figure was adapted from He and Nakayama (1995).

Conclusions

- A neural model for grouping 3D surfaces reproduces basic psychophysical results from a visual search task that requires 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

References

1. He and Nakayama, 1995a. Proceedings of the National Academy of Sciences 92, 11155-11159.

2. Nakayama et al., 1995b. "Visual surface representation: a critical link between lower-level and higher-level vision." In Kosslyn, S.M. and Osherson, D.N. Vision. In Invitation to Cognitive Science. M.I.T. Press, p. 1-70.

3. Marshall et al., 1996. Network: Computation in Neural Systems 7, 635–669.