Approximate Spatial Layout Processing in Early Vision

Michael Hucka [1] **and Stephen Kaplan** [1,2]
[1] Department of Electrical Engineering and Computer Science
[2] Department of Psychology
The University of Michigan
Ann Arbor, MI 48109
{hucka,skap}@umich.edu

Imagine yourself running through rough terrain, perhaps fleeing a predator, or perhaps chasing after prey. Your visual system does not have time to scrutinize the countless trees, rocks, and other objects you pass by. What you need most is enough spatial information to avoid obstacles, to orient yourself, to pick a path. In this situation, even a rough sketch of the spatial layout of the environment can provide crucial information.

Without time or opportunity to perform a more careful analysis, an initial estimate of the layout of visible structures may be all the visual system *can* extract. However, this kind of basic spatial information, while devoid of details about shape and other features useful for object recognition, would often be sufficient to fulfill the requirements above. It could also serve as a stepping-stone to more extensive spatial and object analysis in less constrained situations.

Humans and other natural systems are remarkably adept at extracting spatial organization from vision. Yet the form of this information, and the neural information-processing mechanisms used to obtain it, remain poorly understood. To better understand this capability, it would be useful to know how the visual system can make an initial estimate of the spatial layout of a visual input.

The hypothesis we are exploring is that the system automatically and preattentively extracts the approximate locations, sizes and spatial orientations of major elements in the visual input, thereby obtaining a rough sketch of the spatial layout. We are investigating the mechanisms by which the visual system can extract this from monocular views of natural scenes.

Texture is one well-known source of information that the visual system can exploit (Gibson, 1950). Patterns of texture can be obtained even from brief glimpses, during which a scene will appear static and motion cues are unavailable, and at distances and visual angles at which the effectiveness of stereopsis is limited. Visible texture and texture gradients, arising from markings on surfaces or the spatial arrangements of objects, are useful sources of information both for segregating different regions in a scene and for estimating spatial properties such as surface orientation (Watt, 1995).

Texture and texture gradients can be characterized in terms of the local spatial-frequency content at different points in an image (Bajcsy & Lieberman, 1976). It is widely believed that neurons in the primary visual cortex are responsive to spatial-frequency content. The properties of complex cells in particular would make them highly useful as starting points for texture-based analysis (De Valois & De Valois, 1990). These neurons, together with other neural circuits beyond the primary visual cortex, could serve as part of the mechanisms for *both* performing a rough segmentation of the scene, *and* estimating the general spatial orientations (slant and tilt) of segmented regions. But the question of how both processes can be combined into a single system has rarely been addressed (Krumm & Shafer, 1994). Most existing models of texture-based segmentation assume that textured regions within the visual input are free of systematic distortions due to surface slant; conversely, most models of texture-based shape estimation assume inputs consisting of a single surface.

For spatial layout analysis, the system should also be able to estimate the locations and sizes of the different regions. There is empirical evidence that the visual system computes the locations of simple figures automatically and preattentively. This location information appears to take the form of the centers-of-mass of the regions (Morgan, Hole & Glennerster, 1990). There is also evidence suggesting that the visual system automatically computes the general sizes of visual stimuli (Findlay, Brogan & Wenban-Smith, 1993).

Taken together, the approximate locations, sizes and spatial orientations of major elements in a scene would provide an agent with a rough, initial sketch of the spatial layout. One of the goals of our research is to develop a biologically reasonable model and simulation of this processing.

References

Bajcsy, R., & Lieberman, L. (1976). Texture Gradient as a Depth Cue. *Computer Graphics and Image Processing*, 5, 52–67.

De Valois, R.L, & De Valois, K.K. (1990). *Spatial Vision*. Oxford University Press.

Findlay, J.M., Brogan, B., & Wenban-Smith, M.G. (1993). The Spatial Signal for Saccadic Eye Movements Emphasizes Visual Boundaries, *Perception & Psychophysics*, 53(6), 633–641.

Gibson, J.J. (1950). The Perception of Visual Surfaces. *The American Journal of Psychology*, 58(3), 367–384.

Krumm, J., & Shafer, S.A. (1994). Segmenting Textured 3D Surfaces Using the Space/Frequency Representation, *Spatial Vision*, 8(2), 281–308.

Morgan, M.J., Hole, G.J., & Glennerster, A. (1990). Biases and Sensitivities in Geometrical Illusions. *Vision Research*, 30(11), 1793–1810.

Watt, R.J. (1995). Some Speculations on the Role of Texture Processing in Visual Perception. In *Early Vision and Beyond* (pp. 59–67). Cambridge, MA: MIT Press.