Bitmapped Computer Graphics as a Tool in the Psychophysics of Depth Perception

Heinrich H. Bülthoff[1] and Hanspeter A. Mallot[1,2]

1. Center for Biological Information Processing, MIT E25-201, Cambridge, MA 02139, USA
2. Joh. Gutenberg-Universität, Institut für Zoologie III, Saarstr. 21, D-6500 Mainz

The human visual system works much more reliably for complex natural images than for simple synthetic ones. In natural scenes, it is able to analyze complex shapes correctly under a wide variety of viewing conditions, whereas false or ambiguous solutions (illusions) occur frequently for simple line drawings such as the Necker cube. Similar observations can be made for vision modules other than shape and depth perception as well, e.g. color, stereo and motion. Many illusions occur mainly when only one or a few cues are available but are rare in complex natural situations because the interactions of different cues can avoid false interpretations. Until recently, these interactions have been quite difficult to adress with psychophysical methods. Bitmapped computer graphics systems provide a new tool for the convenient control of different cues in complex synthetic images. Shading, for example, can be computed for arbitrary objects, and ray-tracing, texture mapping, and animation techniques allow to compute synthetic images of three-dimensional scenes which are hard to distinguish from natural images.

Depth information can be derived from a number of different cues, such as stereo, shading, texture and motion. Possible types of interaction include accumulation, veto, cooperation, disambiguation and hierarchy. To distinguish between these interactions experimentally, we studied the depth perceived from computer generated images (smooth- or flat-shaded ellipsoids of revolution with different elongation along the viewing axis) containing different combinations of depth cues. The cues could be either consistent or contradictory. Perceived depth was measured by interactively adjusting a stereo depth probe to the surface of the ellipsoid. In a second experiment, perceived shape was measured by matching a fixed image to one whose shape could be changed interactively by the subject. Depth perception is correct when stereo disparity information can be derived from the relative locations of intensity edges in stereo images. If edges are missing, as in a smooth-shaded sphere, stereo depth information can still be derived from the intensities themselves. If shading is the only information available, the perceived depth may be as low as 30% of the correct depth and is almost independent of the elongation. We draw the following conclusions:

(1) Perceived depth is greater when more information is available (accumulation). It increases with the following sequence of cues: shading only; shading and intensity-based stereo; shading, intensity- and edge-based stereo.
(2) Information from shading and texture is accumulated in a non-linear way (cooperativity).
(3) Since the perceived depth of non-disparate flat-shaded surfaces is zero, we may conclude that edge-based stereo overrides shading (veto).
(4) Intensity-based stereo is an important cue when edges are missing or sparse (surface interpolation).
(5) In an ambiguous situation, where intensity-based stereo suggested a flat disc while a small disparate edge cue appeared in the center in front of that disk, subjects reported to see two surfaces. A 'solid' one with about half the depth of the edge cue, and a 'subjective surface' with a glass-like appearance unto which the edge cue appeared to be drawn.

These results are discussed in terms of recently developed models for integration of differen modules in computer vision (coupled Markov-Random-Fields).