DEPTH PERCEPTION WITH MULTIPLE DEPTH CUES

Hanspeter A. Mallot and Heinrich H. Bülthoff
Inst. für Zoologie III, Joh. Gutenberg-Universität, 6500 Mainz, FRG, and
Center for Biological Information Processing, MIT E25-201,
Cambridge MA 02139, USA

The retrieval of depth or range data from two-dimensional images is an intensively studied problem in the computational theory of vision. Psychophysically interesting solutions do exist for synthetic images with single depth cues whereas the interpretation of complex natural images is only in its infancy. Two important questions are (1) what depth cues should actually be relied on and (2) how can information from different cues be integrated.

We addressed these questions psychophysically with complex yet well controlled images generated with a computer graphics system. The images showed end-on views of polyhedral and smooth ellipsoids, i.e., images with and without intensity discontinuities or edges (as defined by zero-crossings in the DOG-filtered images). Shading was computed for parallel illumination from the viewing direction and different reflectance functions. Edges that could be used as primitives for stereo matching were present in the images of the polyhedral objects but not in those of the smooth ellipsoids. Stereo pairs were interlaced on a CRT-monitor, the interlace signal of the screen triggered shutter glases such that only the appropriate half-image of the stereo pair was visible to each eye. A map of perceived depth was measured by adjusting a small stereo depth probe interactively to the perceived surface at 45 regularly spaced points in the image (3 subjects, at least 5 sessions per image). Since binocular viewing was required for the stereo probe, test patterns without stereo cues had to be presented as a pair of identical images. In some experiments, this was used to generate contradictory cue combinations.

Shape recognition is perfect for the disparate flat-shaded ellipsoids where sufficient edge information is available. If the edge information is removed as in the smooth-shaded ellipsoids, the human visual system can still make use of the disparate intensities in the two half-images of the stereo pair (*intensity-based stereo*). Depth perception is slightly poorer, leading to reduced depth values rather than to an increased scatter in the measurement. The perceived depth is significantly reduced if the disparity information is removed completely, as in the case of binocularly viewed identical smooth-shaded ellipsoids (*shape-from-shading*). The performance of the human visual system for this task is poor. (It can, however, be improved by additional pictorial cues such as contour or texture.) If sufficient edge information is available, whether disparate or not, it overrides the other depth cues (shading and intensity-based stereo) completely. For sparse edge information this veto effect is restricted to a certain vicinity of the edge.

We conclude that intensity-based stereo is used by the human visual system, in spite of some disadvantages pointed out by computational theory. However, the visual system does not rely on intensity-based stereo in contradictory situations where disparate edges are combined with non-disparate shading. Second, intensity-based stereo is a distributed mechanism. Additional data suggest that the disparity-information used for surface interpolation is not localized at sparse features such as the intensity-peak alone. Finally, the accumulation of perceived depth through integration of multiple depth cues may be interpreted as a trade-off between the available depth information and a "flatness-assumption". If integration was implemented as joint regularization, one might expect this result from the presence of a smoothness constraint.