Stimulating Nearly Correct Focus Cues in Stereo Displays

Kurt Akeley[1], Martin S. Banks[2], David M. Hoffman[2], and Anna R. Girshick[3]
[1]Microsoft Research, Mountain View, CA 94043, U.S.A.
[2]Vision Science, University of California Berkeley, Berkeley, CA 94720, U.S.A.
[3]Center for Neural Science, New York University, New York, NY, 10003, U.S.A.

Keywords : accommodation, focus, vergence, stereo

Abstract

We have developed new display techniques that allow presentation of nearly correct focus cues. Using these techniques, we find that stereo vision is faster and more accurate, and that viewers experience less discomfort, when focus cues are consistent with simulated depth.

1. Introduction

Typical stereo displays stimulate depth cues that indicate flatness at a fixed distance—the distance to the screen or to the focal distance of a head-mounted display (Fig. 1). Autostereoscopic volumetric displays [1] correct this problem by creating illumination at the correct depth, but they sacrifice key graphical properties such as hidden-surface elimination and view-dependent lighting. And their homogeneous 3-D organization requires huge numbers of voxels, making them impractical to build.

We have constructed a laboratory implementation of a *fixed-viewpoint* volumetric display. Such displays lose the desirable property of autostereoscopy—they require a separate display channel for each viewpoint, hence two head-mounted channels per stereo viewer—but they can be engineered to present all depth cues, including accommodation and blur cues, correctly to within a specified tolerance. And fixed-viewpoint volumetric displays require far less resolution in depth than in the spatial dimensions, so voxel count is moderate relative to autostereoscopic volumetric displays.

We describe our prototype implementation of a fixed-viewpoint volumetric display, and describe experiments using it that illustrate the practical importance of stimulating nearly correct focus cues. We also show that voxel depth blending, which is necesssary to avoid visible artifacts, has the additional benefit of maximizing retinal contrast when accommodation is to the desired fixation distance, even if that distance is between fronto-parallel planes of voxels.

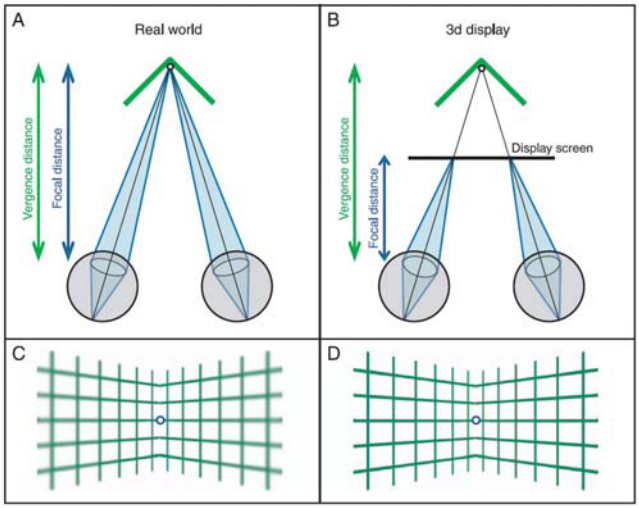

Fig. 1. Accommodation cues (A) and blur cues (C) in the real world indicate true geometric depth. In a typical stereo display accommodation cues (B) and blur cues (D) indicate flatness at a fixed distance.

2. Experiment

Typical human depth of focus (the amount of accommodation error that can be sustained without compromising the quality of the retinal image) is about 1/3 of a diopter. Our prototype fixed-viewpoint volumetric display, which is schematically depicted in Fig. 2, takes advantage of this low requirement for depth resolution by implementing very different resolutions in the spatial and depth dimensions: the minimum spatial resolution is 43 voxels/degree (800 x 400 voxels), while the depth resolution is 2/3 diopter (just three voxels deep). Using half-silvered mirrors, fronto-parallel planes of voxels at three distinct distances (31.1 cm, 39.4 cm, and 53.6 cm) are optically summed into a single

retinal image for each eye.

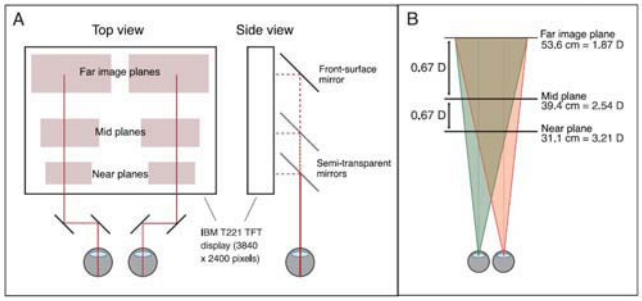

Fig. 2. Schematic of the prototype fixed-viewpoint volumetric display. Semi-transparent mirrors are used to sum images at three distinct distances for each eye. Other mirror pairs implement periscopes that separate the lines of sight so that the left-eye and right-eye images do not overlap on the single LCD display screen, and to allow adjustment for subject-specific inter-ocular distances.

A fronto-parallel object at the distance of one of the three image planes is rendered by illuminating voxels on that image plane only. But objects that span the distance between two image planes are correctly rendered using depth blending, that is, by assigning image intensity to voxels in each plane in inverse proportion to the distance to that plane (Fig. 3).

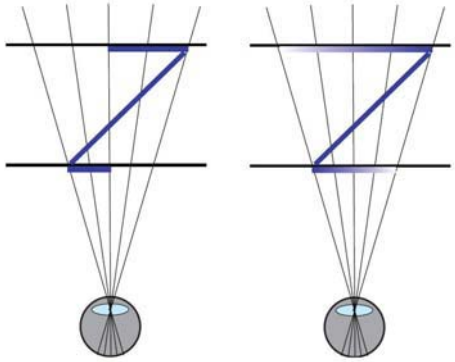

Fig. 3. Depth blending. The object (slanted blue line) spans the distance between two image planes (horizontal black lines). With no depth blending, image intensity is assigned to the nearer plane (left). With depth blending enabled, image intensity is inversely proportional to dioptric line-of-sight distance between the object and the image plane (right).

Experiments were performed on young subjects who had good vision and were unaware of the mechanism of the display or the purpose of the experiment. Subjects viewed the display while clenching a bite-bar between their teeth, fixing the locations of their eyes relative to the display. Careful calibration ensured that image-plane to image-plane alignment errors remained below one arcmin.

Many different experiments have been run during the five years that the prototype display has been operational. Here we report on two: measuring subject image-fusing performance in both cues-consistent cases (where focal distance is equal to fixation distance) and cues-inconsistent cases (where focal distance differs from fixation distance), and measuring subject-reported discomfort in both cues-consistent and cues-inconsistent cases.

Time-to-fuse. Three subjects viewed a fronto-parallel, periodically corrugated surface rendered with randomly distributed points. The surface was tilted (rotated about the line of sight) either 15 degrees to the left or 15 degrees to the right. When properly fused, it was easy for the subjects to see the variations in depth and determine the orientation of the surface. But when viewed with only one eye, or while not fused, the dots appeared as a flat, random distribution, and it was not possible to determine the orientation. We measured the time required for subjects to fuse the image. Each trial was chosen at random from several staircase test conditions, each with a different pairing of vergence distance and focal length.

Discomfort. Eleven subjects fixated stereograms at different vergence distances and focal distances for a total session period of 45 minutes. Each subject completed two sessions: one with only consistent vergence and focal distances, and one with only inconsistent vergence and focal distances. Session orders were randomized among the subjects. After each session, the subject completed a questionnaire regarding his or her discomfort.

3. Results and discussion

Time-to-fuse. The time required to fuse the corrugated surface is lower when converging eye motion is required than when diverging eye motion is required (Figure 4). More importantly, for either motion, substantially more time is required to fuse the corrugated surface when there is a large difference between the simulated (vergence) distance and the focal distance. And the time required generally increases as the difference in vergence and focal distances is increased.

Discomfort. Subjects reported significantly worse discomfort after the cues-inconsistent sessions than after the cues-consistent sessions (Figure 5). They also preferred the cues-consistent session over the cues-inconsistent one. This study offers the most compelling evidence to date that some of the

discomfort associated with viewing stereo displays can be attributed to the unnatural relationship between vergence and focal distance.

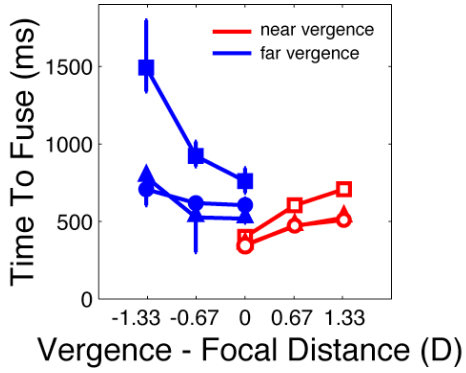

Fig. 4. Results of the time-to-fuse experiment. Stimulus time required to fuse, and therefore perceive, a corrugation in depth is plotted as a function of the difference between vergence and focal distances in diopters. Red depicts trials during which the initial fixation was metrically farther than the final fixation distance—the eyes had to converge to fuse the scene. Blue depicts trials during which the initial fixation was metrically nearer than the final fixation distance—the eyes had to diverge to fuse the scene. Error bars are 95% confidence intervals.

Image contrast and accommodation. In the time-to-fuse experiment described above, simulated distance is always chosen to exactly match one of the three image-plane distances. In each of these special cases voxels on only one image plane are lighted—the plane whose focal distance exactly matches the simulated distance. Therefore accommodation, which is driven in part by maximization of retinal contrast, is optimized when accommodation and vergence distances are equal—the cues-consistent case. We were not surprised to learn that time-to-fuse is minimized in this case, when there is no motivation to decouple accommodation distance from vergence distance. But is time-to-fuse performance hindered when simulated distances that do not match the distance to one of the image planes are tested?

In the general case, when simulated distance is at a point that is between two image planes, depth blending distributes image energy to these two image planes in inverse proportion to the distance from the point to the corresponding plane (Fig. 3). Thus accommodation is not to light coming from the simulated distance, but rather to the sum of light coming from two differing distances—the distances of the two image planes. We might expect that retinal contrast would be optimized by accommodating to one of the two light sources, perhaps the one with greater intensity. And indeed this is correct, but only for high spatial frequencies—those above a threshold value that is determined by a combination of factors, including pupil diameter and the dioptric spacing between the image planes.

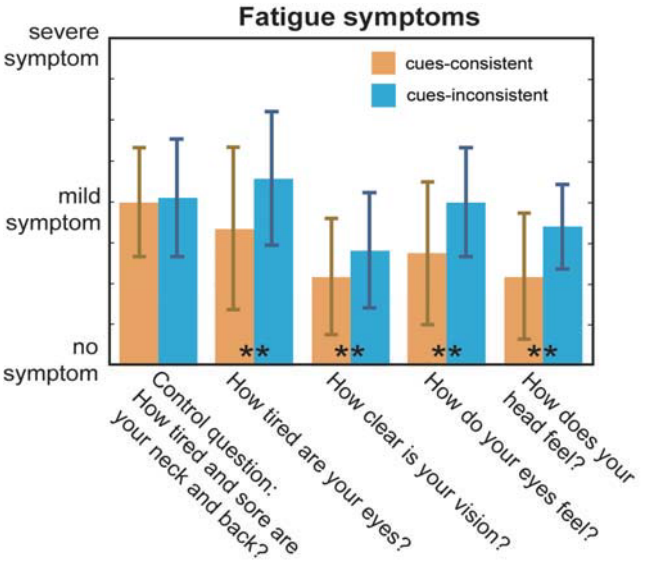

Fig. 5. Results of the visual-discomfort experiment. Orange and blue bars represent the data from the cues-consistent and cues-inconsistent sessions, respectively. Error bars are the standard deviation of reported symptoms from the 17 sets of observations (11 subjects, 6 tested twice). ** indicates $p < 0.025$.

At spatial frequencies below this threshold, retinal contrast is maximized by accommodating to, or very near to, the simulated distance. This surprising result was predicted in Akeley's dissertation [3], where it was demonstrated with a simplified ray-tracing approximation. As reported in Hoffman, Girshick, Akeley, and Banks [2], this result has been verified by analyzing wavefront measurement data using capture and analysis tools provided by Austin Roorda, UC Berkeley.

Fig. 6 illustrates the effect, by plotting retinal contrast as a function of simulated object distance and varying accommodation distance, for three different light sources: the real world, a conventional single-image-plane display, and the prototype 3-image-plane display. All contrasts were computed using wavefront data taken from one of the author's eyes with a pupil diameter of 4.5 mm. Real-world retinal contrast is maximized when accommodation distance exactly matches actual object distance (top row,

diagonal red bars). Single-image-plane retinal contrast is maximized when accommodation distance matches the distance to the display screen, regardless of the simulated distance (middle row, horizontal red bars). But, for the three-image-plane display, simulating distances between the near and far image planes, the contrast of a five cycle-per-degree spatial signal is maximized when accommodation distance equals simulated distance (bottom-left plot, diagonal red region). Whereas the retinal contrast of a 12 cpd spatial signal, which is high enough to be outside the range of spatial signals that drive accommodation, is optimized at discrete image-plane distances (bottom-right plot).

Returning to the question of time-to-fuse results for non-image-plane simulated distances, additional experiments reported in Hoffman, Girshick, Akeley, and Banks [2] demonstrate that there is no performance penalty for depth-blended illumination for inter-plane simulated distances.

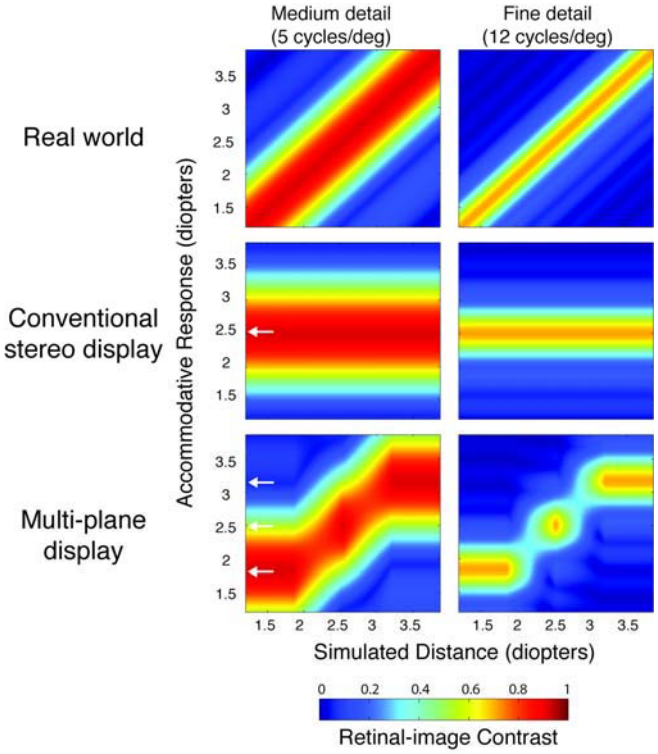

Fig. 6. Retinal-image contrasts for different display techniques. Contrast is a function of both simulated distance (horizontal axis) and the accommodation distance of the eye (vertical axis), both in diopters. The resulting retinal-image contrast of an object of contrast 1.0 is indicated by the colors. White arrows in the middle and bottom rows represent the distances to the image planes.

4. Summary

Our three-image-plane fixed-position volumetric display prototype has proven to be a reliable research tool. We have used it to demonstrate that inconsistencies between simulated distance and focal distance significantly reduce subject performance in tasks that require stereo fusion of complex scenes. We have further demonstrated that such inconsistencies alone significantly increase subject discomfort, a result that had been anticipated but not unambiguously shown. Finally, analysis of multi-image-plane image generation using measured wavefront data show that, for reasonable image-plane separations and typical pupil diameters, retinal contrast of spatial frequencies that drive accommodation is maximized when accommodation is to, or very nearly to, the simulated distance, rather than to the distance of an adjacent image plane.

Ongoing work includes engineering variations, such as the use of dynamic optics elements, to make fixed-viewpoint volumetric displays more practical, as well as development of a display apparatus with the ability to measure accommodation, vergence, and pupil size during multi-plane viewing. Other researchers are investigating alternate approaches to stimulating nearly correct focus cues, such as scanned-voxel displays [4].

5. References

1. G. E. Favalora, J. Napoli, D. M. Hall, R. K. Dorval, M. G. Giovinco, M. J. Richmond, and W. S. Chun, 100-million-voxel volumetric display. *In Proceedings of SPIE,* Vol. **4712**, p.300-312 (2002).
2. K. Akeley, Achieving near-correct focus cues using multiple image planes, PhD dissertation, *Stanford University,* appendixes C and D (2004).
3. D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, Vergence-accommodation conflicts hinder visual performance and cause visual fatigue, *Journal of Vision,* **8**(3):33, p.1, http://journalofvision.org/8/3/33/, (2008).
4. B. T. Schowengerdt and E. J. Seibel, Scanned voxel displays, *Information Display,* **24**(7), p.26 (2008).