53.2: A Binocular Stereoscopic Display System with Coupled Convergence and Accommodation Demands

Eli Peli, T. Reed Hedges, Jinshan Tang and Dan Landmann

The Schepens Eye Research Institute, Harvard Medical School, Boston, MA, USA

Abstract

Conflict between the accommodation and convergence demand in stereo display systems has been implicated as a cause of perceptual error (failure of size constancy), discomfort and visual system changes. A novel method is proposed for such displays, which, by utilizing a measured object of fixation can resolve the conflicts via computational approach. We implemented such a display system and used it to evaluate the feasibility of the approach.

1. Introduction

The main technique used currently in stereo display applications is dichoptic (separate image to each eye) presentation of binocular disparity (Figure 1). Such a stereo system may be implemented in a Head-Mounted Display (HMD) or on a desktop display using one of many possible binocular separation methods. The issue of viewer comfort with such a system is particularly important for situations that require many hours of continuous use. One of the problems most often cited is the de-coupling of convergence and accommodation demands [2][6]. We describe a novel solution to this problem, preliminary implementations and experiments.

In the real world shifting fixation from an object at one distance to another object at a different distance requires a change in both convergence and accommodation. A dichoptic stereo display requires de-coupling the natural relationship between convergence and accommodation responses that are coupled for real objects.

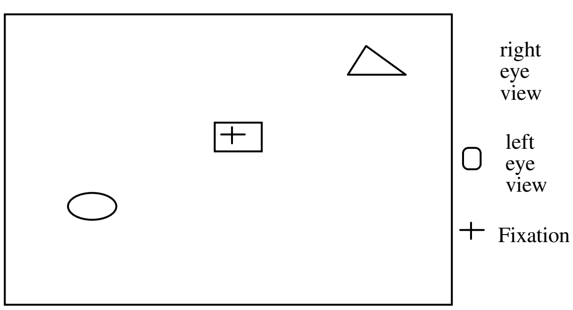

Figure 1. The typical dichoptic stereo display situation. The rectangle, fixated, is presented with zero disparity and therefore will be perceived to be on the screen causing no conflict between the accommodation and convergence demands. The triangle is presented with crossed disparity, and requires convergence in front of the screen while accommodation ought to remain at the screen, causing a conflict when fixated. Similarly the ellipsoid is presented with uncrossed disparity requiring convergence behind the screen.

This de-coupling has been implicated by many as the source of eye strain, fatigue, and visual system changes frequently noted with use of such systems [1][4][8]. However, Peli [7] directly tested the different effects of stereoscopic and non-stereoscopic imagery in the same HMD and found no significant differences.

A solution for the de-coupling of the convergence and accommodation by way of measuring eye position and adjusting the screen optical distance to match the object's simulated distance was proposed by Shiwa et al [9]. The same group also reported preliminary implementations [5]. By measuring the positions of both eyes using eye tracking devices they determined the fixated object and derived the object's simulated depth from the graphics data. They then changed the optical distance of the screen to match the simulated distance by moving a relay lens. Despite a significant level of effort and development they were careful not to report the ability to operate the complete system.

Although this concept may sound simple it is quite difficult to implement and may be practical only for an HMD system and not for desktop systems [9].

1.1 The New Stereo Display Method

Our method requires only computational changes for implementation. The proposed method uses eye movement monitoring to determine the object being fixated. However,

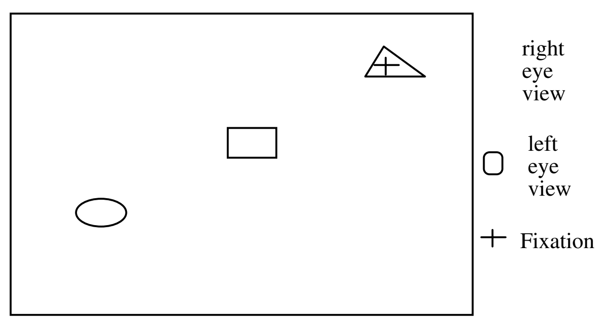

Figure 2. The layout from Figure 1 following a change of fixation to the triangle. The triangle is now presented with zero disparity. The other two objects are presented with uncrossed disparity of different magnitude and will be perceived to be farther than the triangle. Convergence demand for the fixated object is again equal to the accommodation demand. The relative depths of all objects are preserved. The absolute stereo depth is changed but all other monocular cues for absolute depth remain stable suggesting no change in depth at all.

 ISSN/0001-0966X/01/3201-1296-$1.00+.00

instead of changing the accommodation demand to match the simulated distance we change the disparity or convergence demand of all objects to bring the fixated object to zero disparity and thus remove the convergence accommodation de-coupling conflict (Figure 2). We initially tested this approach without eye movement monitoring, but will soon integrate that aspect into further experiments.

2. Experimental Evaluation

Two experiments were performed to determine if users of our display method noticed artifacts as a result of the disparity change. During the experiments, a simple 3D scene and target were presented to the subject. As the target jumped or bounced from one point to another in a prerecorded set, the point of regard (POR) was determined to be the point on the 3D scene geometry directly behind the target, and this point's distance from the viewpoint was used to change the disparity of all points in the scene until the POR reached zero disparity (Figure 3).

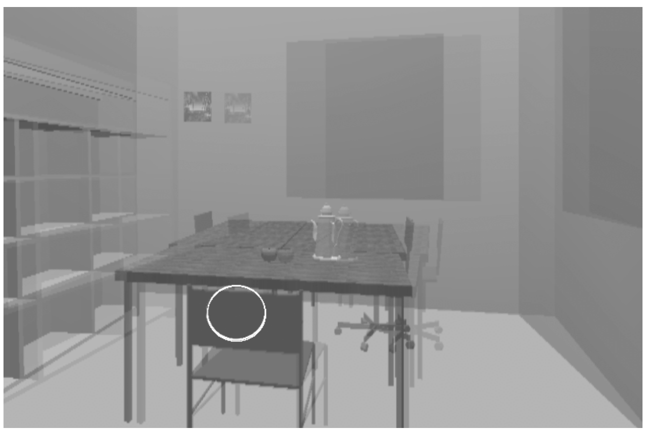

Figure 3. Illustration of the 3D scene as seen without the LCD shutters to show the screen disparities (In this figure the left- and right-eye views have been overlaid, and disparity can be seen as a double image). At the position of the target the disparity is zero. As the target moves to each location in turn the disparity at these locations are turned to zero shortly after the transition. A commercial software library called World Tool Kit (WTK) (Sense8/EAI) was used to render the 3D geometry on a desktop CRT. CrystalEyes LCD shutter glasses (Stereographics) were used to produce a stereoscopic effect by a Field-Sequential method at 120 Hz (first experiment) or 100 Hz (second experiment). The scene was presented in color.

The subject was instructed to follow the target, a small white circle, which had no depth and no disparity. Each time the circle jumped to its next location, the "convergence distance", or distance from the virtual viewpoint at which every point in the 3D scene converged between each eye's image, was gradually increased or decreased until it eventually reached the distance from the viewpoint to the point of regard. As this convergence change occurred, every point at the convergence distance approached zero disparity, and simultaneously the disparity of every other point in the scene automatically changed to retain its depth relative to the convergence distance. The change in convergence distance per second is the "convergence rate". No size change was caused by this convergence change.

2.1 First Experiment: Static Scene

2.1.1 Method

In the first experiment, two methods of convergence changing were tried and compared. In the "2-eye" convergence method, WTK's "Asymmetric" projection model was used, in which the eye views are equally rotated or skewed horizontally, inward or outward to give the new convergence point zero disparity. The "1-eye" convergence changing was done by just moving the right-eye view laterally. We used WTK's "Symmetric" projection model, and adjusted the virtual eye separation proportional to the desired convergence distance, creating zero-disparity at that distance.

Six subjects were tested**,** all were adults, aged from 20 to 35. Two subjects had left dominant eye. Experiments compared the sensitivity for motion detection among different convergence rates and in two different convergence methods: 1-eye convergence and 2-eye convergence. For each method, three different parallax levels (parallax here means the distance between the two eyes' virtual views), and four different convergence levels were used. Thus 24 (2 methods X 3 parallaxes X 4 convergence rates) conditions were completed by each subject. The subject was seated 57 centimeters away from the screen, and was asked to track the jumping target with his eyes, following it as soon as possible. At the same time, the subject was asked to pay attention to any unusual changes or movements in the image. The subject was asked to report what changes he noticed and give a score that represented the amount of the change or motion, ranging from 1 to 5, where 1 meant a change could not be detected, 2 meant the change was slightly detectable, 3 meant the change was somewhat detectable, 4 meant the change was easily detectable and 5 meant the change was very easily detectable.

The parallax levels used in the first experiment were 2.2 cm., 5.5 cm., 8.8 cm. Eight points in the scene were selected to be fixated by the subjects. The disparity difference between two consecutive points were 0.20 deg., -0.063 deg, -0.26 deg., 0.05 deg., 0.02 deg., -0.42 deg., and 0.12 deg. when the parallax was 2.2 cm. (The disparity difference between two consecutive points in the other two parallax levels were multiplied by 2.5 and 4 respectively; minus sign here means that the latter point was closer to the virtual eyes than the previous point). Four convergence rates were used for each parallax: 11cm/sec, 88 cm/sec, 440 cm/sec, 44000 cm/sec. When the parallax was 2.2 cm, for the slowest rate of change, the corresponding rate of disparity change between points were 0.09deg/sec, 0.04 deg/sec, 0.05 deg/sec, 0.08 deg/sec, 0.07 deg/sec, 0.07deg/sec, 0.12 deg/sec, 0.11 deg/sec respectively. The corresponding disparity convergence rates of the eight objects for the other three levels can be obtained by multiplying by 8, 40 or 4000, respectively.

2.1.2 Results

The subject's response (1-5) was used to calculate receiver operating curves (ROCs)[1], plotting the probability curves of motion detection in the given conditions against each other. The probability curve in a given condition versus the standard condition was obtained as follows: For the first data point on the curve, the fraction of the subjects giving the rating 1 in the given condition was plotted against the fraction of the subjects giving the rating of 1 in the standard condition. For other points on the curve, the corresponding cumulative fractions were calculated. The data were fitted by binormal model [3]. The area under the ROC (A_z) was taken as a measure of motion detection and the level of correlation was used in determining the significance of the difference between two areas.

Two ROC analyses were performed: The first analysis compared the sensitivity of motion detection between two convergence methods. While the symmetric condition was slightly better, the difference was not significant. We therefore collapsed the convergence rate data over these 2 conditions. The second ROC analysis compared the sensitivity of motion detection among different convergence rates. The receiver operating curves are shown in Figure 4. We set the fastest convergence rate condition as the standard condition and the other convergence conditions were compared with it.

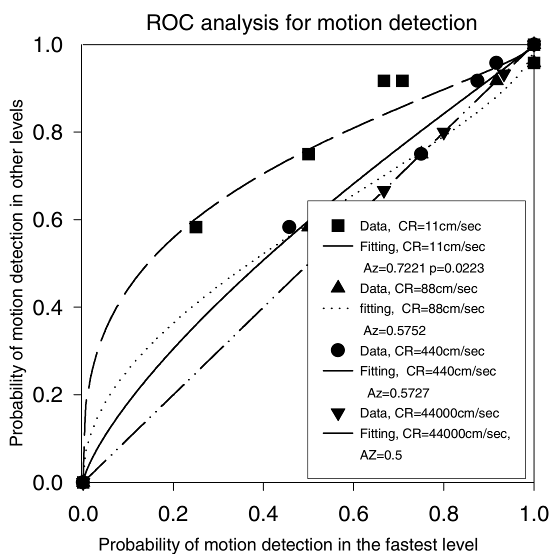

Figure 4. Comparison between different convergence rate levels. CR represents convergence rate. Where *p* is not given, the difference is not significant. Only the slowest convergence rate gave a significant change in motion detection compared to fast convergence rate.

The first experiment was carried out with a graphics card limiting the spatial resolution to 800 x 600 pixels (which was about 40 x 30 dpi on our monitor of 20 in. diagonal). At this resolution the movements of many objects, especially those with small or sharp components resulted in a flickering appearance caused by errors. This artifact most likely interfered with our results and increased motion detection.

[1] The software used in the ROC analysis is available at (http://www-radiology.uchicago.edu/krl/toppage11.htm).

2.2 Second Experiment: Moving Scene

Since it appears that static images can be viewed without detectable motion only with very slow change of disparity, we performed a small second experiment to test the effect with moving images.

2.2.1 Method

To add overall motion to the scene, the virtual viewpoint was moved about the scene and re-oriented to bring different areas of the scene into view at different angles, resembling camera movements familiar from television and movies. Simultaneously, the target jumped from point to point. The test consisted of 20 joined segments in each of which the virtual "camera" was moving or was still (as in the first experiment), and also the convergence was updated or was not updated for each target jump. Each segment contained about 4 target jumps.

Five adults between the ages of 20 and 60 were tested, at a distance of 57 cm from the screen. For this test the parallax value was 4.4 cm and the convergence rate was 66 deg/sec at 57 cm. distance. This experiment also used a thin white circle rather than the white filled disc of the first experiment for a target. Both targets had no disparity. The subjects were given a short demonstration of the task and were shown example convergence changes before testing. During the test, they were asked to indicate how easily they could detect the convergence motion artifacts on a continuous scale ranging from "not detectable" through "very easily detectable".

2.2.2. Results

The second experiment was performed at a resolution of 1152 x 864 pixels (about 57 x 43 dpi). At this resolution the aliasing artifact was reduced significantly, though not eliminated.

Five naive subjects participated in the second experiment. The data for four could be properly fit by the ROC curve (see Figure 5 for an example). The fifth subject's data was inconsistent and fell mostly below the midline indicating detection of more motion in the standard display when no convergence motions are induced. This data was rejected.

For all other subjects the detection of the motion artifact caused by our method was more noticeable for the static condition without camera movements than during camera motion. In both cases with and without camera motion the artifact was not very apparent as indicated by proximity of the ROC curve to the diagonal line (Az = 0.5). The average area under the ROCs (Az) was 0.62 and 0.56 for the no camera motion and camera motion conditions, respectively. In most cases the ROC fit seems to underestimate the improvement in artifact reduction as seen in Figure 5.

The motion artifact detection in the no camera motion condition is lower than that exhibited in the previous experiment, average Az 0.65 vs 0.72 found there. This reduction is most likely the result of higher resolution that removed the aliasing artifact associated with image movements that was noted with the lower resolution graphic card used in the first experiment.

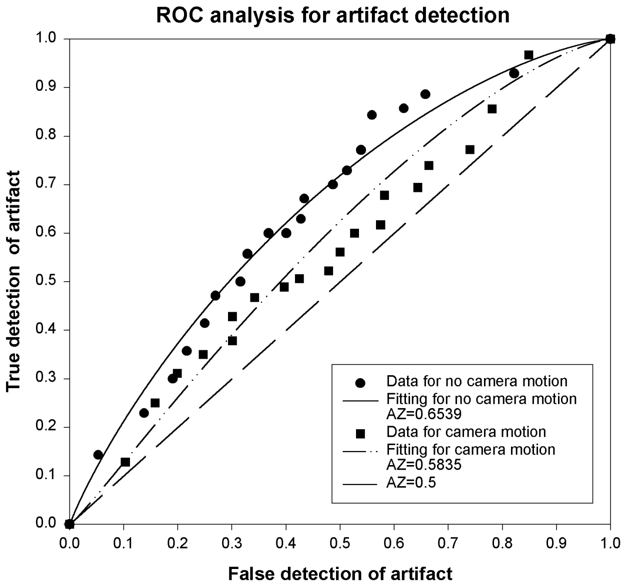

Figure 5. Comparison between camera motion and no camera motion, from one typical subject's data.

3. Further Developments

The system described here will be completed by the addition of continuous eye movement monitoring to determine actual point of regard, and an overall scene motion or camera motion.

An eye tracking system from I-Scan Inc. will be used to record eye position. This system can record point-of-regard data through the LCD shutter glasses. Only a single (dominant) eye position is recorded to provide subject POR that will be used to drive the stereo system disparity change. Results of the eye tracker calibration verification (Figure 6) show that the I-Scan device is less accurate in the vertical direction. The measurement error was random in direction and very small

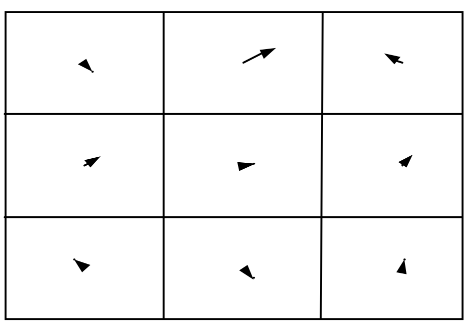

Figure 6. Verification of eye position calibration. 27 targets were randomly presented to a subject, and the mean difference between target position and eye tracker data is drawn as vectors in 9 sectors of the screen.

4. Conclusion

A novel method for presenting stereo images was proposed and implemented. With this solution, plus eye movement measurements to determine point of regard, a stereo display system in either HMD or desktop format would provide comfortable viewing with no perceptual conflicts and be free of binocular eyestrain. Preliminary experiments show that the system can be implemented and provide comfortable stereo viewing with minimally noticeable artifacts.

5. Acknowledgements

Supported in part by NASA grant NLC2-1039.

6. References

[1] Howarth P.A. Empirical studies of accommodation, convergence, and HMD use. In Proceedings of the Hoso-Bunka Foundation Symposium: The human factors in 3-D imaging, Tokyo, Japan, 1996

[2] Ellis SR, Hiruma N and Fukuda T (1993) Accommodation response to binocular stereoscopic TV images and their viewing conditions. *SMPTE J* 102; 1137-1144.

[3] Metz C, Wang P, and Kronman HB (1983) A new approach for testing the significance of differences between ROC curves measured from correlated data. In Proceedings of the VIII Conference on Information Processing in Medical Imaging, Deconinck F, editor. The Hague: Martinus Nijhoff, 431-445.

[4] Mon-Williams, M., J.P. Wann, and S. Rushton (1993) Binocular vision in a virtual world: visual deficits following the wearing of a head-mounted display. *Ophthalmic and Physiological Optics* 13 (4): 387-391.

[5] Omura, K, S Shiwa, and F Kishino (1996) 3-D Display with accommodative compensation (3DDAC) employing real-time gaze direction. *SID'96 Digest* 889-892.

[6] Peli E. (1995) Real vision and virtual reality (invited paper). *Optics and Photonics News*. July ; 28-34.

[7] Peli E. (1998) The visual effects of head-mounted-display (HMD) are not distinguishable from those of desk-top computer display. *Vision Res*. 38:2053-2066.

[8] Rushton, S., M. Mon-Williams, and J.P. Wann (1994) Binocular vision in a bi-ocular world: new-generation head-mounted displays avoid causing visual deficit. *Displays* 15: 255-260.

[9] Shiwa, S., K. Omura, and T. Miyasato. Development of three-dimensional display with accommodative compensation. In Proceedings of The Third International Display Workshops, Kobe, Japan, SID 1996.