A User Study towards Understanding Stereo Perception in Head-worn Augmented Reality Displays

Mark A. Livingston* Zhuming Ai† Jonathan W. Decker‡

Naval Research Laboratory

ABSTRACT

Properly perceived stereo display is often assumed to be vital in augmented reality (AR) displays used for close distances, echoing the general understanding from the perception literature. However, the accuracy of the perception of stereo in head-worn AR displays has not been studied greatly. We conducted a user study to elicit the precision of stereo perception in AR and its dependency on the size and contrast of the stimulus. We found a strong effect of contrast on the disparity users desired to make a virtual target verge at the distance of a real reference object. We also found that whether the target began behind or in front of the reference in a method of adjustments protocol made a significant difference. The mean disparity in the rendering that users preferred had a strong linear relationship with their IPD. We present our results and infer stereoacuity thresholds.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/Methodology; H.1.2 [Models and Principles]: User/Machine Systems—Human factors

1 INTRODUCTION

It has often been assumed that correct perception of stereo is a requirement in head-worn augmented reality (AR) displays used for applications in which virtual objects are presented at close working distances. This echoes the general understanding from depth perception literature that stereo is important in perceiving near-field depth [2]. However, the accuracy of of stereo depth perception in head-worn AR displays has not been studied extensively. Until recently, displays that allowed adjustable inter-pupillary distance (IPD) were limited to a few research systems and an occasional commercial offering. This made the issue difficult to explore beyond proper setting of the IPD used in rendering. The mismatch between real IPD and virtual IPD may be exploited [13], but for many researchers, the tacit assumption was that "proper" stereo would require and be satisfied by an adjustable IPD.

However, the question of the strength of stereo disparity to provide depth information versus the other depth cues available remains open in the minds of many researchers, and many aspects of stereo perception are thought to exhibit individual differences [13]. Certainly the range of normal human IPD seems wide relative to the absolute size; from the smallest "normal" value (approximately 53mm) to the largest normal value (approximately 73mm) is an increase of over 35%. Additionally, there are differences in the depth of the eyes from the nominal plane at which the eyepieces of a head-worn AR display may be situated. Finally, some people (estimated as high as 20% [13] and as low as 2-5% [1]) are stereo-blind, and thus receive no depth information from binocular disparity (though they often compensate well enough to obscure this).

*e-mail: mark.livingston@nrl.navy.mil

†e-mail:ai@itd.nrl.navy.mil

‡e-mail:jonathan.decker@nrl.navy.mil

IEEE International Symposium on Mixed and Augmented Reality 2009
Science and Technology Proceedings
19 -22 October, Orlando, Florida, USA

2 RELATED WORK

Incorrect IPD distorts distance perception; if the IPD is set at the population mean (65 mm), an observer with a larger IPD would perceive the object as farther away than it is; an observer with smaller IPD would perceive it closer [8]. Similarly, errors in the judged depth of nearby virtual objects have been measured as a function of changes in binocular vergence [3], with errors varying between subjects. Models have been built that characterize the various errors in stereo head-worn displays [11, 7]. These models – among many contributions – showed the importance of accounting for IPD.

Another concern about stereo vision is the conflict of binocular disparity, convergence, and accomodation cues; an inaccurate accomodative-convergence response is a candidate reason for observation of poor near-compensated stereoacuity [15]. Stereo presentation can help overcome accommodative demand [4]. Poor configuration can be a more acute concern in head-worn AR displays, and has been demonstrated to cause problems in head-worn VR displays [10]. Normal visual acuity is 1 arcmin, but hyperacuity for stereoscopic depth can reach 3 arcsec [5]. Also, it has been suggested that retinal image disparities can produce veridical depth perception without interpretation from a fixation point [9], albeit under reduced cue conditions.

3 USER STUDY

We designed a simple depth matching task for users to perform. We mounted the eyepiece from an nVisorST optical see-through display (1280×1024 at 60 Hz, for each eye) on a chin rest (Figure 1). We then approximately aligned the central view rays through the eyepieces with a monitor (at a distance of 1.2 meters) that would serve as the "real" portion of the environment. Users were asked to adjust the apparent depth of a virtual target square shown on the nVisorST to match the depth of a "real" reference square drawn on the monitor. Each square was drawn with a pair of filled triangles with four-sample antialiasing (Direct3D). We aligned the eyepieces such that the reference and target were separated by approximately one degree; this is more than the minimum separation necessary to obtain good stereo thresholds [14].

Users did not explicitly manipulate depth; instead, they manipulated binocular disparity between the left and right images; this had the effect of changing the apparent distance. The virtual target might appear to start either in front of or behind the correct distance, with a uniform random variable controlling the starting disparity. This random variable independently selected an offset distance of 5-10 pix and a positive or negative sign. This offset was then added to the disparity that yielded depth matching on average during a pilot study. The user pressed the up or down arrow to "move the target" further or closer "in depth" (i.e. to make the changes in the disparity that would induce such apparent changes in the 3D position of the target). We chose a step size of 0.15 pixels, which in the pilot study yielded a comfortable speed of motion when a key was held down. Users hit the space bar to record a response.

Figure 1: The experimental scene included the display eyepieces from our nVisorSR mounted on a chin rest (near right), a monitor to serve as the real world in the nVisorST's line of sight (far right), and two monitors to show the experimenter what image was being sent to the nVisorST (far left).

3.1 Subjects

Eleven subjects (nine male) completed the experiment with no complications; the subjects ranged in age from 22 to 44 (average of 31). As has become usual for AR perception experiments, we noted mild or moderate increases in eye strain, fatigue, or dizziness among seven of the eleven users. All subjects volunteered and were not compensated; they were drawn from the scientific and clerical staff of our laboratory. In pre-experiment procedures, we measured the user's IPD (Essilor Digital CRP) and set the IPD of the nVisorST to match this distance. We then screened the subject for stereo fusion with a simple test of nine targets, each of which consisted of a set of four circles. Users were asked to identify which one of the four circles was closer than the others; all users got at least seven correct and were thus judged to have passed this test. However, the efficacy of this procedure will be discussed in Section 3.3.

3.2 Variables and Hypotheses

Our goal was two-fold: to see if the AR display offered normal stereo acuity thresholds, and to see how such thresholds were affected by the contrast between the target and the background.

We varied the target size and contrast; the reference was always displayed with the same size and contrast as the target. To equalize the contrast, we selected a background used for all stimuli and adjusted the foreground brightness. We measured the brightness of the real background and the real reference through the AR display and the AR target with the selected background illumination [6], taken with a StellarNet EPP2000C color meter with a CR2 cosine receptor. As noted above, we randomized the start position for each trial, giving us an independent variable which was discretized for analysis into whether the target would have appeared to start closer or farther than the subject's mean apparent distance.

We selected four contrast levels and five target (and reference) sizes. These variables were completely crossed and randomly permuted. Each combination of these two parameters was repeated eight times, with a goal of having four trials begin with the target closer than the reference and four with it farther than the reference; this was not achieved perfectly since we did not know where each user would, on average, set the final disparity. It did result in a sufficient balance between the two conditions for analysis. This yielded $4 \times 5 \times 4 \times 2 = 160$ trials for each of our eight subjects with acceptable variance (defined below), for 1280 total trials in the analysis.

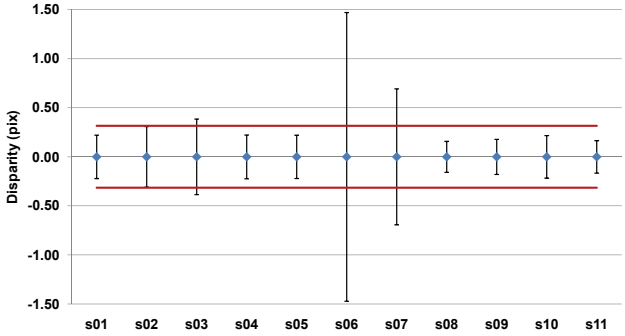

Figure 2: The variation of subjects from their individual means indicated that two of the subjects had significantly higher variation than the others; a third barely exceeded the threshold of twice the minimum variance exhibited. In this and all succeeding graphs, a positive disparity offset implies that the final disparity left the target closer than the average for that user. Also in this and all graphs, the error bars indicate the standard error.

We recorded the accepted disparity (for matching depth), the response time, and the number of changes in direction that the user made during the interactive adjustments. We hypothesized that lower contrast and smaller size would decrease the precision and increase the response time for the depth matches, while the starting position would have no effect. We hoped that the number of changes in direction might help indicate uncertainty of the users, but no significant effects were found.

3.3 Analysis

We began by inspecting the variation of each subject from his or her grand mean disparity. The purpose of this step was to determine whether any of the subjects were so inconsistent as to invalidate the data. We found that two of the subjects had extremely high variance relative to their individual means; we selected a threshold of twice the minimum variation (Figure 2). This resulted in the elimination of three subjects from further analysis. It is interesting to note that subject s07 asked to adjust the IPD after a few trials, complaining of diplopia (double vision). Subject s03 was the only subject to miss any of the targets in the stereo fusion screening questionnaire, getting the minimum (seven) correct to continue. This would appear to raise questions about the utility of the screening test we are using.

We analyzed the data from the remaining eight subjects. We conducted a repeated-measures analysis of variance (ANOVA) using Systat 11 for each of the dependent variables. In order to eliminate some individual differences, we replaced the dependent variable of disparity with "disparity offset from mean" for each user; we use the term disparity offset below to denote this measure. While there were numerous second- and third-order effects, we concentrate on the significant main effects found.

3.3.1 Main Effect: Contrast on Disparity Offset

As one would expect, contrast had a significant main effect on the final disparity chosen by users – $F(3,21)=8.286$, $p=0.001$ (Figure 3). Since we do not attempt to verify the "correct" disparity a user should have chosen, we note that the lower contrast (and brightness) caused users to set more disparity than their (respective) average, indicating a closer vergence point. This is the effect one would expect to see; greater brightness is generally interpreted as indicating closer proximity, and thus it makes sense that our users pulled dimmer targets closer than brighter targets. Thus users are compensating for the lower brightness (measured by contrast).

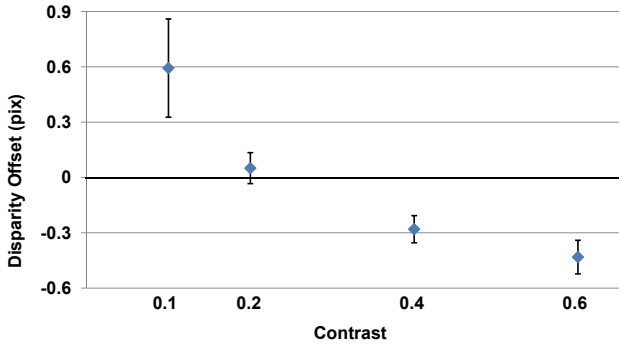

Figure 3: The disparity offset (from an individual subject's mean) over all users for each contrast (brightness) level shows that users felt the need to increase the disparity with decreasing contrast, bringing the vergence point closer to them. This would be consistent with the typical interpretation of greater brightness indicating that an object were closer to the user.

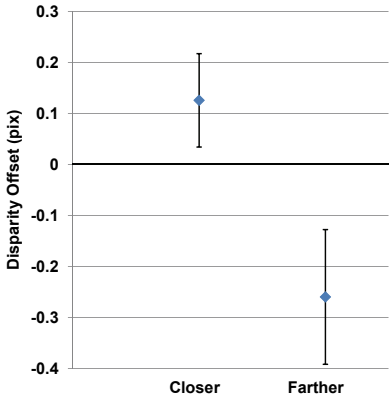

Figure 4: The disparity offset versus starting with a disparity that would indicate a closer rather than farther distance shows that users consistently stopped short of the amount of adjustment necessary, assuming that their grand mean was approximately correct.

3.3.2 Main Effect: Start Position on Disparity Offset

Using the discretized (binary) independent variable of whether the target started with a disparity that would yield a closer or farther apparent distance than the user's individual mean gave us a significant main effect. Users consistently stopped their adjustments too soon – $F(1,7)=6.089$, $p=0.043$. When the target started with a disparity that would make it appear too close (Figure 4), they left it slightly too close (positive disparity offset). When it started farther away than their average, they left it too far away. It would appear that as soon as users reached a threshold of satisfactory matching, they deemed the match to be accurate. This is typical with the method of adjustments and helps establish the stereoacuity threshold.

There were significant interactions of the start position with both contrast – $F(3,21)=9.636$, $p=0.000$ – and size – $F(4,28)=10.475$, $p=0.000$. Low contrast appears to have caused the tendency for users to leave a target too close; contrast levels of 0.2 and higher caused users to actually push a target from too close to slightly too far away. Strangely, a contrast of 0.2 and an initial disparity giving a too-distant apparent position caused users to pull the target too close; this differed from the general pattern as well (Figure 5).

3.3.3 Main Effect: Time

Contrast – $F(3,21)=7.168$, $p=0.002$ – and size – $F(4,28)=3.847$, $p=0.013$ – both had a significant main effect on response time.

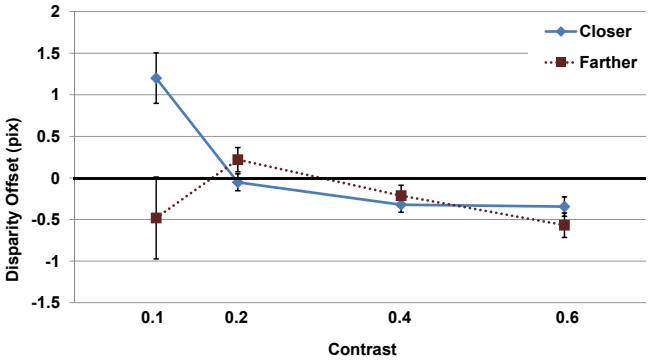

Figure 5: Separating the cases in which the target appeared to start too close from those in which it started too far shows that it was low contrast that caused the users to leave an object too close, whereas the effect of starting too far was more nearly consistent (though contrast of 0.2 differed from the other cases).

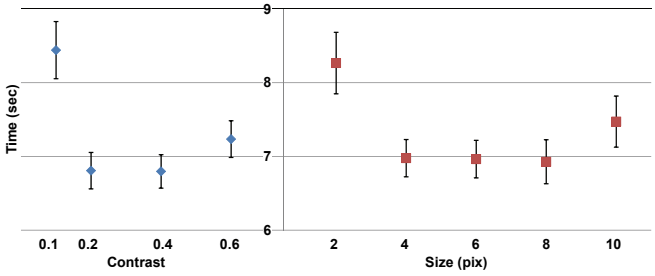

Figure 6: The main effect of contrast on time (left) shows that users were slowest with low contrast; the main effect of size on time (right) shows that users were slowest with the smallest size. Note that neither of these graphs are monotonic as one might expect, showing that the users were able to accomplish the task well in general.

Users were slowest with low contrast and small size, neither of which produce a great surprise. What is interesting in these graphs (Figure 6) is that neither are monotonic, which indicates that it was really only the smallest size and lowest contrast that presented difficulties for the users.

3.3.4 Regression of Disparity and IPD

As noted above, there is a relationship between the IPD and the apparent depth that comes from binocular disparity. Since we have a fixed depth and allow adjustment of the disparity, we find a well-fit linear regression of the mean disparity for each individual user against his or her measured IPD (Figure 7, Pearson correlation $R=0.9666$, $F\text{-Statistic}=85.4780$, $p=0.0001$). This would indicate that users were indeed able to use the IPD in the AR display software and hardware as they would in a normal binocular viewing situation (i.e. one without intervening optics or AR).

4 Discussion

Many of the observations in the previous section serve to confirm the ecological validity of our task; users behaved as if they were able to match the real and virtual objects in depth. The primary task remaining is thus to compute the stereoacuity exhibited by our users with the virtual targets in this experiment. Table 1 shows the mean disparities for each value of contrast and size of the target (and reference). With the exception of the lowest level of contrast and smallest size, users generally reached a threshold of approximately two arcminutes before they could no longer tell that the

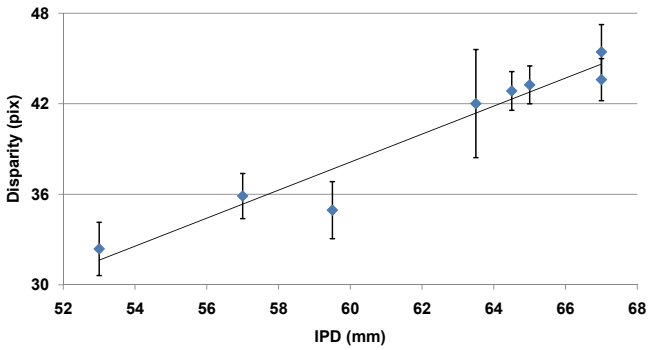

Figure 7: Regression of the mean disparity for each subject versus each subject's IPD shows an excellent linear fit, indicating that users were able to convincingly verge the real and virtual objects.

Contrast	Mean Abs Disparity	Size	Mean Abs Disparity
0.1	3.953 arcmin	2	3.877 arcmin
0.2	2.182 arcmin	4	2.441 arcmin
0.4	1.941 arcmin	6	2.084 arcmin
0.6	2.133 arcmin	8	1.960 arcmin
		10	2.434 arcmin

Table 1: Mean disparities for the levels of contrast and target size (in pixels) show that users generally achieved a disparity threshold of approximately two arcminutes. This is the difference in depth that they no longer recognized as being separate depths. Averages were taken of absolute values, to avoid artificially reducing the error by averaging positive errors (target too close) with negative errors.

target and reference were at different depths.

The two exceptional cases are consistent with the general literature on stereo thresholds and contrast. Stereo thresholds are a power law function of contrast [12]. We see the same result in Figure 3 for the case of a head-worn AR display. The numbers in Table 1 don't quite as well reflect this pattern due to lacking a monotonic function. The implications of this are a subject for future investigation.

5 Conclusions

Many of our results are interesting "only" in the sense that they are verification that the stereo perception through the head-worn AR display is indeed behaving like stereo perception should in everyday vision. While we do not have sufficient power to make equality claims, this is encouraging that no complete failures in the AR perception were revealed in our study.

We do find that the stereo threshold users were able to achieve with the AR display is much greater than the thresholds that have been achieved by experiments with normal vision, by a factor of almost 60 compared to the thresholds of a few arcseconds under ideal conditions. There are a number of reasons that could cause such a difference. The cue of binocular disparity does not work in isolation from other near-field cues that are inconsistent in a head-worn AR display, notably accomodation and vergence. One potentially interesting next step is to test the utility of AR displays with adjustable focus mechanisms. The transparent nature of the graphics in the optical see-through display is perhaps another reason for the poor stereo thresholds, although we did equalize the apparent brightness of the real and virtual through the color meter readings.

For AR applications in which detailed work is done at close distances, the correct perception of stereo can be of great benefit. Our study helps establish the limits of this cue. Comparing values in Table 1 to the display resolution (2.25 min/pix) and the step size (0.15 pix=0.3375 arcmin) reveals that our subjects were in most cases achieving at least the performance that would be expected from the resolution of the display. One might suspect that anti-aliasing did not assist users. If so, one would expect no size or contrast to yield a threshold under 2.25 arcmin. Some conditions did yield such a result, and the lowest disparities users could achieve for the most advantageous contrast and size in Table 1 are statistically better than 2.25 arcmin. However, some combinations of size and contrast limited stereoacuity, and the highest disparities are significantly worse than 2.25 arcmin. Further investigation is required to determine the relative strength of the factors leading to better stereoacuity than the display resolution yet limiting it from normal human performance, but it appears that our methodology would elucidate such a ranking given sufficient data.

Acknowledgements

The authors thank John Philbeck, the anonymous subjects, and the anonymous reviewers. Sponsored by DARPA under the ULTRA-Vis program for work as a performing member of the Lockheed-Martin team, contract #FA8650-09-C-7908. Approved for Public Release, Distribution Unlimited. The views, opinions, and/or findings contained in this article/presentation are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Department of Defense.

References

[1] V. Bruce, P. R. Green, and M. A. Georgeson. *Visual Perception: Physiology, Psychology, and Ecology*. Psychology Press, 3rd edition, 1996.

[2] J. Cutting. Reconceiving perceptual space. In *Perceiving Pictures: An Interdisciplinary Approach to Pictorial Space*, pages 215–238. MIT Press, 2003.

[3] S. R. Ellis, U. J. Bucher, and B. M. Menges. The relationship of binocular convergence and errors in judged distance to virtual objects. In *Proceedings of the International Federation of Automatic Control*, pages 297–301, June 1995.

[4] S. R. Ellis and B. M. Menges. Localization of virtual objects in the near visual field. *Human Factors*, 40(3):415–431, Sept. 1988.

[5] M. Fahle and T. Troscianko. Computation of texture and stereoscopic depth in humans. Technical Report 1208, Massachusetts Institute of Technology Artificial Intelligence Laboratory, Oct. 1989.

[6] M. A. Livingston, J. H. Barrow, and C. M. Sibley. Quantification of contrast sensitivity and color perception using augmented reality displays. In *IEEE Virtual Reality*, Mar. 2009.

[7] J. W. McCandless and S. R. Ellis. The effect of eye position on the projected stimulus distance in a binocular head-mounted display. In *Stereoscopic Displays and Applications XI, Proceedings of SPIE Vol. 3957a*, Jan. 2000.

[8] P. Min and H. Jense. Interactive stereoscopy optimization for head-mounted displays. In *SPIE Stereoscopic Displays and VR Systems*, Feb. 1994.

[9] M. Mon-Williams, J. R. Tresilian, and A. Roberts. Vergence provides veridical depth perception from horizontal retinal image disparities. *Experimental Brain Research*, 133:407–413, 2000.

[10] M. Mon-Williams and J. P. Wann. Binocular virtual reality displays: When problems do and don't occur. *Human Factors*, 40(1):42–49, Mar. 1998.

[11] W. Robinett and J. Rolland. A computational model for the stereoscopic optics of a head-mounted display. *Presence: Teleoperators and Virtual Environments*, 1(1), Winter 1991.

[12] A. M. Rohaly and H. R. Wilson. The effects of contrast on perceived depth and depth discrimination. *Vision Research*, 39:9–18, 1999.

[13] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, second edition, 2004.

[14] G. Westheimer and S. P. McKee. Stereogram design for testing local stereopsis. *Investigative Ophthalmology & Visual Science*, 19:802–809, 1980.

[15] B. P. H. Wong, R. L. Woods, and E. Peli. Stereoacuity at distance and near. *Optometry and Vision Science*, 79(12):771–778, Dec. 2002.