Ocular Parallax for VR Grasp and AR Realism

Anastasios Nikolas Angelopoulos
EE367: Computational Imaging and Display
Stanford University
nikolasa at stanford dot edu

Robert Konrad
Mentor
Stanford
http://stanford.edu/~rkkonrad/

Abstract

Ocular Parallax (OP), a perceptual effect caused by a small difference between the nodal point of the eye and its center of rotation, allows humans to see depth with miniscule movements of the eye. But despite being well studied in the visual science community, OP has never been included in the rendering pipeline of virtual reality (VR) and augmented reality (AR) systems. To fill this literature gap, we conducted a series of psychophysical experiments meant to understand the effects of OP in VR and AR depth perception. We found that although humans are very sensitive to ocular parallax, it does not seem to improve depth perception or realism with the current state of eye-tracking.

1. Introduction and Motivation

Virtual and Augmented Reality (VR/AR) is an enabling technology for many socially impactful outcomes, including more accurate breast and laparoscopic surgery (*1, 2*) and safer manufacturing processes (*3*). Improving realism in VR and AR will improve the perceived physical accuracy of these systems, integrate them more seamlessly with the natural world, improve ease-of-use, and thereby, potentially improve effectiveness.

In order to increase realism, many advances are needed, including lighter head mounted devices (HMDs); fast, low-latency head and eye tracking; and high field-of-view stereoscopic displays. However, more fundamental than any of these technological advances is effective understanding and exploitation of the human visual system.

The most high-profile example of a human visual system effect which must be accounted for in VR and AR rendering is the coupling of vergence and accommodation. Vergence is the rotation of your eyes to triangulate a point. Accommodation is the changing of the lens shape in order to see a nearer or farther focal plane. Since VR and AR headsets have a fixed focal plane, vergence and accommodation become decoupled, causing the vergence-accommodation conflict. There is an ongoing race in the VR and AR communities to solve the vergence-accommodation conflict, and it has spurred fruitful research directions in rendering (*4, 5*). The ultimate goal is to increase comfort in VR and AR.

There is a second goal in modeling the human visual system in VR and AR: realism. Recent work in VR rendering methods aims to improve realism of virtual scenes by mimicking natural visual cues in processing, especially those related to depth. For example, Cholewiak et. Al. implemented a depth- and color- dependent blur and showed it increased perceptual realism (*6*). Also, Kellnhofer et. Al. improved depth dependent displays by introducing motion parallax in addition to binocular disparity (*7*). Six degree of freedom (6DOF) rendering is also a standard tool in VR headsets and introduces the effects of head parallax into rendering (*8*). These rendering techniques have sparked significant research directions in augmented and virtual reality (*5, 9, 10*). The effect we propose studying, ocular parallax, seeks to generate similar research directions by exploiting a quirk of the human visual system: movement of the eye causes relative motion of objects in a scene. We hope that by rendering in this technique to AR and VR systems we can achieve a boost in realism and also depth perception. Our ability to render in this ocular parallax is dependent upon gaze-contingent rendering methods, which rely on eye-tracking technology. Eye tracking has recently reached the point where it can accurately and quickly track the fixation point of users (as shown by Cholewiak et al. with their gaze contingent depth of field rendering method). Such gaze-contingent rendering methods are key to realism.

2. Related Work

Virtual and Augmented Reality Depth Cues

Humans see depth in many ways. According to Cutting and Vishton, there are nine major depth cues: binocular disparity, height in visual field, accommodation, vergence, motion perspective, aerial perspective, relative density, relative size, and occlusion (*11*). These depth cues are summarized in the figure below.

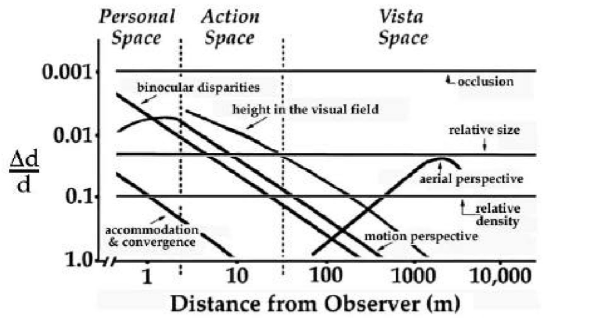

Figure 1: Just-discriminable depth thresholds as a function of the log of distance from the observer, from 0.5 to 5000 meters, for nine different sources of information about layout. Such plots were originated by Nagata (1981), and are extensively modified and elaborated here; they are plotted with analogy to contrast sensitivity functions. Our assumption is that more potent sources of information are associated with smaller depth-discrimination thresholds; and that these threshold functions reflect suprathreshold utility. These functions, in turn, delimit three types of space around the moving observer—personal space, action space, and vista space—each served by different sources of information, with different weights. This array of functions, however, is idealized; Figure 2 shows variations on the themes shown here.

As explained in Cutting and Vishton's diagram above, each of these depth cues operates in a different "space" – personal space, action space, or vista space. In VR and AR, we are mostly about personal and action space. Binocular disparity is accomplished through traditional stereo rendering in VR and AR. Stereo rendering also fixes convergence, as users fuse two images with similar objects as long as they are within Panum's fusional area (*12*). Accommodation is traditionally fixed at the focal plane, although with the recent development of varifocal displays, one can create systems allowing users to verge and accommodate correctly (*13*). Ocular parallax is a motion perspective cue, so it is extremely strong in the personal and action space – thus it is worth studying in the context of VR and AR, where the fixed focal plane is generally about two meters from the user (corresponding to the highest sensitivity of motion perspective cues).

Biological Background

Ocular parallax has been known in the biology community for over a century. In 1854, Brewster calls ocular parallax "the change in the apparent position of the drawings of solid bodies" (*14*). Since the discovery of this effect, some work has been done to characterize it biologically. For example, Bingham did a series of experiments showing that ocular parallax is visible within the visual acuity limit of human vision (*15*). He also made some critical estimations of human eye structure, such as nodal point location, which we used later on in analysis.

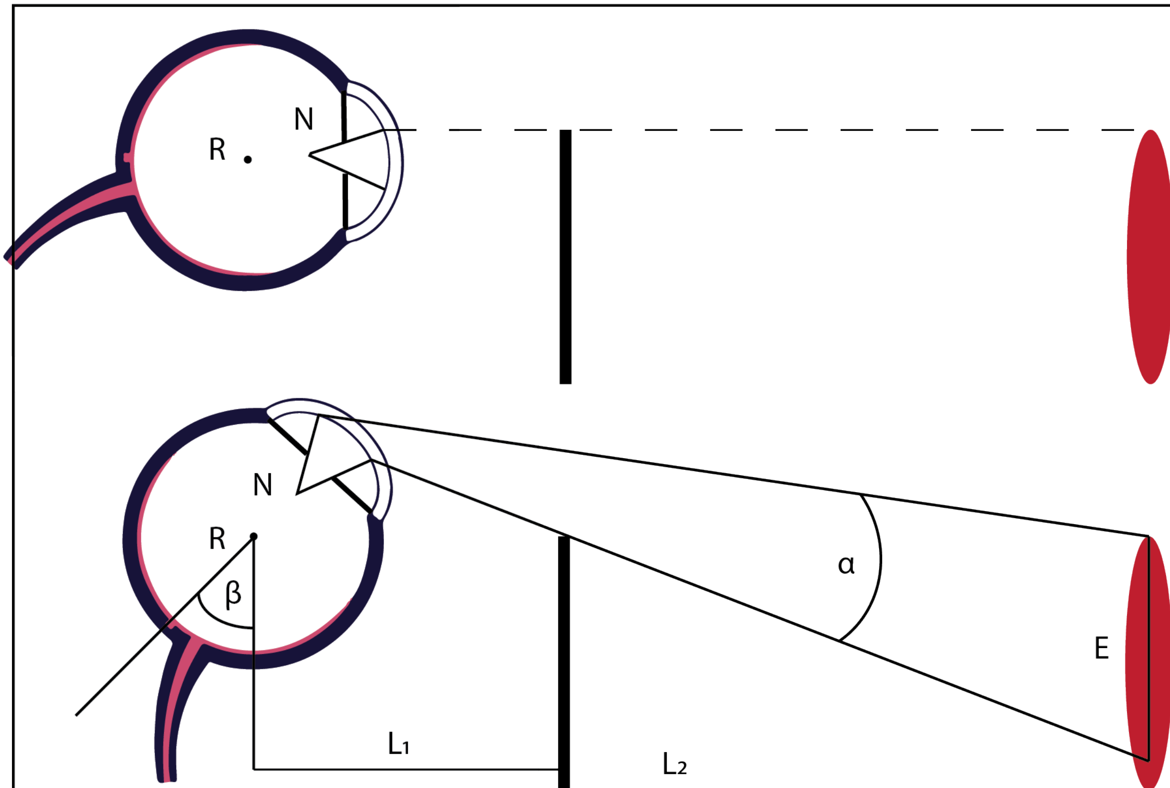

Fig. 2: Ocular parallax explanation. This figure depicts the effect of ocular parallax, caused by a small (few millimeter) difference between the center of rotation and projection of the human eye. As seen above, during central fixation, the red target is occluded. When the eye rotates, however, the nodal point (assumed center of projection) translates, leading to a parallax effect, and allowing the user to "peek around" the occluding ojbect. Consequently, some of the red is exposed.

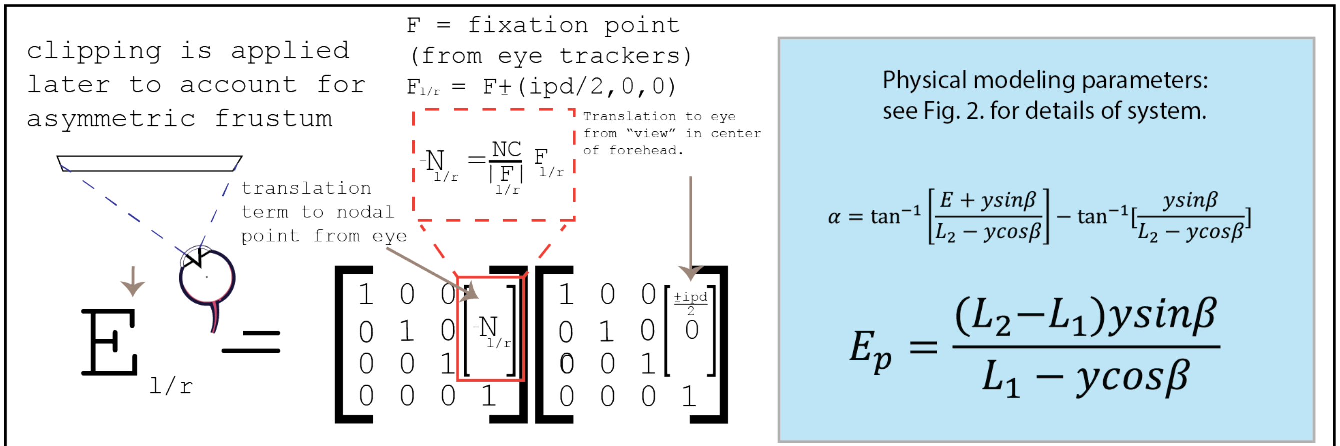

Fig. 3: Rendering method for ocular parallax. Modifying the projection matrix of the traditional graphics pipeline with a simple linear transformation based on physical modeling of the nodal point location yields a simple, effective method.

Hadani et al. also presented a deterministic mathematical model showing that it is possible to extract the movement field from optical flow using a set of differential equations (*16*). Finally, Mapp and Ono studied a special case of ocular parallax, where targets near the limit of the nasal visual field disappear when monocular gaze is directed towards them (*17*). This biological work has fully characterized ocular parallax in the human visual system, with the only exception being a lack of knowledge about the location of the nodal point with respect to the center of rotation. However, to the knowledge of the authors, all contemporary virtual and augmented reality systems make the faulty assumption that eye movement does not create parallax. As we show in Fig. 2., ocular parallax exists based on a difference between the nodal point and the center of rotation of the eye; current VR/AR manufacturers have assumed it is the same. The goal of our study is to understand what effect this assumption has and to correct it.

3. Methods

Summary

In order to ascertain the effects of ocular parallax on depth perception and realism, we conducted four user studies: a threshold study to assay sensitivity, a blind reaching task to assay depth perception, a VR preference study to assay VR realism, and an AR preference study to assay AR realism. The focus of this paper will be on the blind reaching task and the AR preference study.

Rendering Ocular Parallax

The rendering algorithm for implementing ocular parallax is described in Fig. 3. Essentially, while stereo rendering applies a transformation which shifts from the center of the forehead to each eye individually, we implemented an extra transformation from the eye to the nodal point. This transformation is calculated using the interpupillary distance (IPD) of the subject, the distance between the nodal point and the center of rotation, and the fixation point derived from tracking the eye. The details of this rendering method are not the focus of this writeup as the rendering method was primarily the work of Robert Konrad – look to the recently submitted SIGGRAPH manuscript for details on the view frustum clipping and other math (Konrad et al.).

Head Mounted Devices and Eye Tracking

For the VR threshold and preference studies, a custom VR headset was built comprising an HTC Vive headset with attached Pupil Labs eye trackers. The eye trackers sit in the bottom of the headset, outside the field of view, and perform all necessary analysis to determine the user's fixation point within one degree of visual field. The calibration of this system is included with the purchase of the eye trackers.

For the AR preference study, a Microsoft Hololens was used. Pupil Labs eye trackers for the Hololens were attached to the HMD and calibrated per-subject before use.

Population

The population had normal best-corrected vision. Each study enrolled between nine and fifteen participants. For binocular testing, subjects had to pass a RanDot stereo test which has been shown to correspond with real world stereo acuity (*18*). Subjects were also asked if they could fuse the stereo images, and if they could not, they were excluded. All experiments were performed with oversight by the Stanford Institutional Review Board, performed after informed consent by subjects, and cooperated with the guidelines in the Declaration of Helsinki.

Threshold Study and VR Preference Study

The sensitivity of humans to ocular parallax was estimated with a monocular detection and discrimination threshold study. Two stimuli were shown in the center of the VR headset field of view (FOV) such that the front stimulus occluded the back stimulus with central fixation. A fixation target was rendered to rotate in the periphery subject to smooth pursuit constraints. Multiple distances between the stimuli were tested to understand at what level people can see ocular parallax. After calibrating the eye trackers, subjects were asked to fixate on the rotating target and performed a two-alternative forced choice test between two conditions: ocular parallax on and ocular parallax off. They were asked to choose the interval which “exhibited more relative motion between the front and back target.” User input was handled with a keyboard. Recalibration was performed after each individual trial. Each subject completed a total of 225 trials randomly and uniformly distributed among the viewing conditions.

For the VR preference study, users wore the custom VR setup and were shown three different realistic scenes with three rendering modes: ocular parallax on, ocular parallax off, and reverse ocular parallax (looking to the left reveals objects on the right and vice versa). Each scene had a large depth of field. The closest object in each scene was at 3D, 2D, and 1D respectively. Users ranked the three rendering modes uniquely for each scene. This was done in monocular and binocular modes.

More detail on these protocols is contained in the SIGGRAPH submission.

Blind Reaching Task

Because verbal estimates of depth are an unreliable measure, we performed a blind reaching task instead (*19*). Blind reaching is a validated measure of depth perception (*20*).

After inclusion and baseline testing, subjects wore the custom VR head mounted device setup with eye tracking and completed the calibration procedure. Then, they were shown a virtual living room, shown in Fig. 6, comprising multiple real-world objects at many different depths within action space. Subjects held a hand tracker. Head tracking was turned on. Users were first asked to familiarize themselves with the room. During this phase, the hand tracker was rendered into the scene so subjects would have a reference point for depth. Subjects were given two practice trials. Then, they entered the experimental phase. Subjects were asked to view a grasp target (a virtual pencil) at a particular distance. They were asked to use their eyes to look at different parts of the room in the process of estimating its distance. Then, after a minimum of five seconds, subjects were allowed to turn off the display by pressing the HTC Vive trigger and perform the blind reaching task. Without moving their torso, subjects placed the hand tracker at the estimated location of the grasp target, then pressed the trigger. All tracking data was recorded during the blind reach phase of the experiment. Data was normalized for the maximum arm reach of the subject after experimentation. The final reach position was calculated as the average of the last ten samples of the participant’s reach data. Data was normalized for head movement.

The conditions tested in this study are: rendering mode (ocular parallax enabled, ocular parallax disabled) and reach target distances as a percentage of maximum reach (50%, 58%, 67%, 75%, 82%, and 90%). Each combination of conditions was repeated 5 times for a total of 60 trials. Experiments took less than half an hour.

AR Preference Study

This study was designed to understand whether ocular parallax improves realism in a highly controlled mixed reality setting, with virtual and real objects in the same scene. The study was a monocular three forced choice preference experiment. Subjects were asked to rank three conditions uniquely: ocular parallax on, ocular parallax off, and ocular parallax reversed. Rendering mode order was randomized. The number of repeats was three. Subject angle and xyz position were fixed with a bite bar. Bite bars are clinically validated to fix people’s movement within half a degree and a millimeter of movement (*21*).

Subjects were shown four stimuli: two real and two virtual. The real stimuli were laser cut circles placed in optical alignment with the user’s pupil such that if the user fixated in the center of the front circle (colored with a gray gaussian noise pattern), the rear circle (red) was not visible. The virtual stimuli were visible at the same time, rendered above the physical stimuli, depicting the exact same circles in AR. A digital fixation target like the one in the threshold study was shown, rotating within smooth pursuit constraints. Subjects fixated on the target and were asked to pay attention to the stimuli with their peripheral vision.

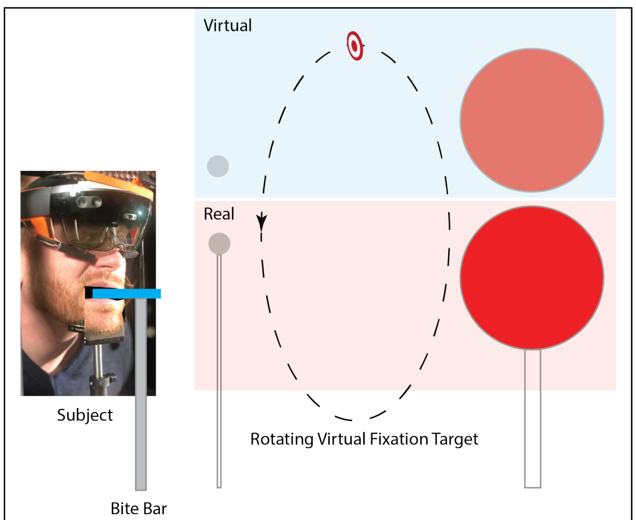

Fig 4: AR study setup. Subject wears a hololens with Pupil Labs eye trackers and bites onto a bite bar. Fixating on the rotating target, the subject ranks in which of three rendering modes (OP on, OP off, and OP reversed) the virtual looks more similar to the real stimulus.

Two iterations of this study were completed. In the preliminary version, eye tracking was used, and in the final version, eye tracking was not used, and ocular parallax was rendered assuming subjects fixed on the target. Both results are the same and are reported below.

We recorded each subjects' ranking along with the rendering mode and order. Data analysis was performed only after all subjects were enrolled and tested. No subjects were excluded. The experiment is depicted in Fig. 4.

4. Analysis, Evaluation, and Comparison to Previous Work

All protocols used in these experiments refer to cited standard psychometric tests that have undergone validation in the psychophysics community in previous work. We have endeavored to make our system extremely reproducible. Prior work does not exist in the field of VR/AR ocular parallax, but we have used validated methods for our tests.

Threshold Study and VR Study

Statistical analysis on threshold data for both detection and discrimination was done using a standard Bayesian estimation of the psychometric function (*22*).

Data from the VR study were analyzed with a Friedman test.

VR Depth Perception Study

The final position of the tracker was calculated based on a ten sample average of the sampled positional data and normalized for head position. Linear models were fit to both sets of reaches.

AR Study

Data from the AR study were analyzed with a Friedman test.

5. Results

Threshold Study

Results from the threshold study are disclosed in supplemental figure S1. The most compelling result of the threshold study shows that people can detect ocular parallax at a threshold of roughly 0.36D. This is an order of magnitude lower than the visual acuity limit in the periphery, probably because OP is a motion cue which traverses multiple photoreceptors during eye movement. Still, humans seem to be very sensitive to the effects of OP. More analysis is included in Konrad et. al.

VR Preference Study

The Friedman test showed that people expressed no preference in any of the three scenes for any of the rendering modes. Figure 5 contains results for each condition.

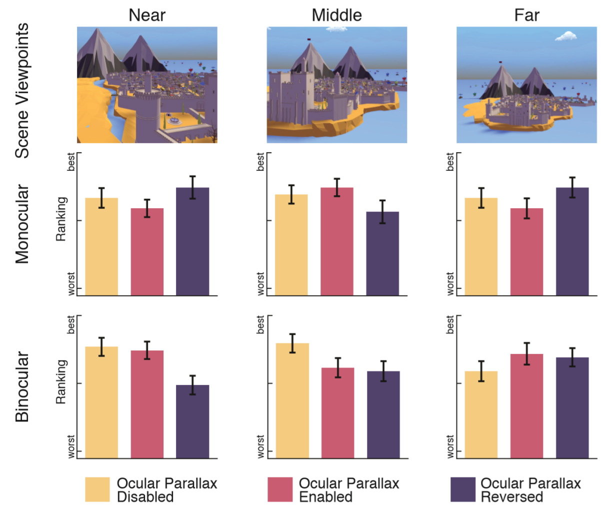

Fig. 5: Subjective depth perception. We asked users to rank conventional rendering, ocular parallax rendering, and reverse ocular parallax rendering by how effective these modes are in conveying the 3D structure of the scene. Neither the monocular nor the binocular conditions show a significant effect for any of the scenes (top row).

VR Depth Perception Study

The linear models for ocular parallax enabled and disabled had slopes 1.141 and 1.128 and intercepts 0.167 and 0.165 respectively. The multiple regression analysis did not show a significant difference. Fig. 6 shows the regression.

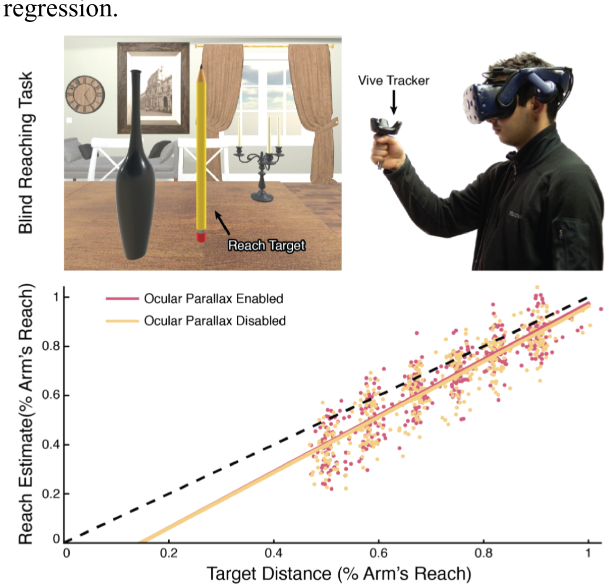

Fig. 6: Egocentric depth perception. This study investigates whether the additional depth cues in ocular parallax aids egocentric depth perception. The viewing environment including the reach target, a pencil, are shown in the top left. Users viewed the target and then, with the display turned off, reached to it as seen in the top right. Participants' reaches in the ocular parallax enabled and disabled modes are found in the bottom plot.

AR Study

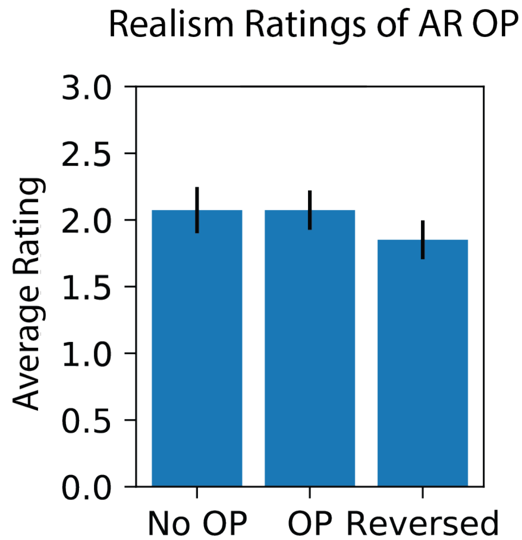

Fig. 7: AR Realism Results. Subjects expressed no preference for any rendering method within the context of our test. Results were same with the eye tracking enabled and disabled.

As in Fig. 7, there is no significant preference for any rendering mode in AR by Friedman test. The p-value for the Friedman chi square is p=0.64.

6. Discussion, Limitations, Future Work

Discussion

The primary contribution of this work is the evaluation of ocular parallax as a depth cue in VR and its effect on realism in AR. We conducted the set of user experiments and showed a compelling result that people have a very low perception threshold (0.36 diopter difference between optically aligned objects) to ocular parallax. We also estimate discrimination thresholds and confirm that the just noticeable difference between two objects at different relative depths in front of a background stimulus is directly proportional to their absolute depth from the background. Though this result confirms that ocular parallax is an important visual cue in VR, our experiment in VR show that it is not important for depth perception.

Our AR experiment produced a surprising negative result. We constructed the experiment with the gray circle occluding the red circle, so that when ocular parallax was turned off in the virtual rendering, there would be none of the red circle visible. In the physical system, however, red should always have been visible to the user. We hypothesized that this dissonance would cause users to pick the OP on or OP reversed modes, simply in order to match the colors in the real system. However, they did not. This may be because the physical alignment process involved was imprecise even with the use of a bite bar. So, subjects may have perceived themselves as correcting for errors in alignment by ignoring any red they saw in the physical system (because they were "supposed to be in alignment").

The subjective experiences of users in the study corroborate negative results in our depth perception and AR/VR realism experiments. After testing, subjects usually experienced no strong preference and wondered what they were "supposed to see." It could be that a series of unavoidable confounding factors in our experiments call this result into question. These factors are discussed in Limitations. In our future work section, we describe a series of experiments that can be done as eye-tracking technology improves.

Limitations

We suspect that latency and jitter caused by eye-tracking significantly worsened people's perception of realism when evaluating the ocular parallax enabled and reversed modes in the preference studies. Subjects in our studies commonly expressed after testing that they noticed and disliked the jitter effect, ranking it low in the experiment. In order to mitigate this in the AR realism study, we tried to turn off ocular parallax and assume the subject was fixated on the gaze target – however, if the subject "cheated" and looked in the center to see ocular parallax being rendered, it would look very unphysical, and they would rank it lower. We suspect that multiple subjects may have done this. In any case, there is no way around the requirement for faster, lower-latency eye tracking to avoid jitter.

Future Work

Two follow up studies should be conducted. Firstly, as eye tracking improves, another realism study should be conducted to see if ocular parallax improves realism in the absence of jitter. Secondly, a longer study should be done to see if ocular parallax improves visual comfort over the range of 2-3 hours in a realistic scene. Even though ocular parallax does not drive vergence or accommodation, it may still have effects in the oculovestibular system that we do not fully understand.

7. Conclusion

We have shown that ocular parallax is a significant visual cue in AR and VR that should be explored further for subtle but important optimizations based on the human visual system. In addition to improving the hardware of VR/AR systems, this work suggests that further basic inquiries into human perception may benefit the development of head mounted VR and AR systems.

References

1. Botden, Sanne MBI, and Jack J. Jakimowicz. "What is going on in augmented reality simulation in laparoscopic surgery?" Surgical endoscopy 23.8 (2009): 1693.
2. Sato, Yoshinobu, et al. "Image guidance of breast cancer surgery using 3-D ultrasound images and augmented reality visualization." IEEE Transactions on Medical Imaging 17.5 (1998): 681-693.
3. Caudell, Thomas P., and David W. Mizell. "Augmented reality: An application of heads-up display technology to manual manufacturing processes." System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on. Vol. 2. IEEE, 1992.
4. D. M. Hoffman, A. R. Girshick, K. Akeley, M. S. J. J. o. v. Banks, Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. **8**, 33-33 (2008).
5. G. J. I. t. o. v. Kramida, c. graphics, Resolving the vergence-accommodation conflict in head-mounted displays. **22**, 1912-1931 (2016).
6. Steven A. Cholewiak, Gordon D. Love, Pratul P. Srinivasan, Ren Ng, and Martin S. Banks. 2017. Chromablur: Rendering Chromatic Eye Aberration Improves Accommodation and Realism. ACM Trans. Graph. (SIGGRAPH Asia) 36, 6 (2017), 210:1–210:12.
7. Petr Kellnhofer, Piotr Didyk, Tobias Ritschel, Belen Masia, Karol Myszkowski, and Hans-Peter Seidel. 2016. Motion Parallax in Stereo 3D: Model and Applications. ACM Trans. Graph. (SIGGRAPH) 35, 6 (2016), 176:1–176:12.
8. J. Thatte, J. Boin, H. Lakshman and B. Girod, "Depth augmented stereo panorama for cinematic virtual reality with head-motion parallax," 2016 IEEE International Conference on Multimedia and Expo (ICME), Seattle, WA, 2016, pp. 1-6. doi: 10.1109/ICME.2016.7552858.
9. J. Thatte, B. J. E. I. Girod, Towards Perceptual Evaluation of Six Degrees of Freedom Virtual Reality Rendering from Stacked OmniStereo Representation. **2018**, 352-351-352-356 (2018).
10. S. Eftekharifar, N. F. Troje, in *2018 International Conference on 3D Immersion (IC3D)*. (IEEE, 2018), pp. 1-6.
11. J. E. Cutting, P. M. Vishton, in *Perception of space and motion*. (Elsevier, 1995), pp. 69-117.
12. S. K. Rushton, P. M. J. A. E. Riddell, Developing visual systems and exposure to virtual reality and stereo displays: some concerns and speculations about the demands on accommodation and vergence. **30**, 69-78 (1999).
13. N. Padmanaban, R. Konrad, T. Stramer, E. A. Cooper, G. J. P. o. t. N. A. o. S. Wetzstein, Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. **114**, 2183-2188 (2017).
14. D. J. P. o. t. R. S. o. E. Brewster, 1. On the Law of Visible Position in Single and Binocular Vision, and on the representation of Solid Figures by the Union of dissimilar Plane Pictures on the Retina. **1**, 405-406 (1845).
15. G. P. J. V. R. Bingham, Optical flow from eye movement with head immobilized:“Ocular occlusion” beyond the nose. **33**, 777-789 (1993).
16. I. Hadani, G. Ishai, M. Gur, Visual stability and space perception in monocular vision: mathematical model. *J. Opt. Soc. Am.* **70**, 60-65 (1980).
17. A. P. Mapp, H. Ono, The rhino-optical phenomenon: Ocular parallax and the visible field beyond the nose. *Vision Research* **26**, 1163-1165 (1986).
18. D. A. Leske, E. E. Birch, J. M. J. A. j. o. o. Holmes, Real depth vs randot stereotests. **142**, 699-701 (2006).
19. P. E. Napieralski, B. M. Altenhoff, J. W. Bertrand, L. O. Long, S. V. Babu, C. C. Pagano, J. Kern, T. A. J. A. T. o. A. P. Davis, Near-field distance perception in real and virtual environments using both verbal and action responses. **8**, 18 (2011).
20. A. M. Wing, A. Turton, C. J. J. o. m. b. Fraser, Grasp size and accuracy of approach in reaching. **18**, 245-260 (1986).
21. V. Menon, K. Lim, J. Anderson, J. Johnson, A. J. B. R. M. Pfefferbaum, Instruments,, Computers, Design and efficacy of a head-coil bite bar for reducing movement-related artifacts during functional MRI scanning. **29**, 589-594 (1997).
22. H. H. Schütt, S. Harmeling, J. H. Macke, F. A. J. V. R. Wichmann, Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. **122**, 105-123 (2016).

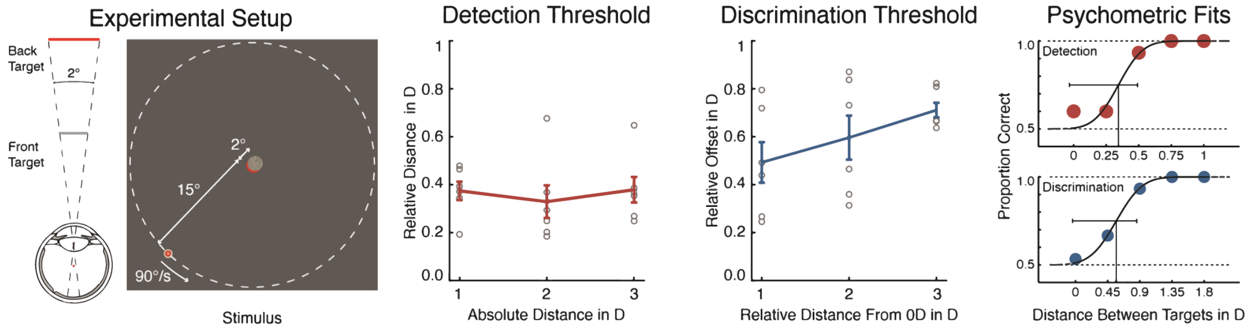

Fig. S1:Detection and discrimination thresholds for ocular parallax in VR. Two user studies were conducted to estimate perceptual thresholds using an HTC Vive Pro head mounted display presenting stimuli of the form shown on the left. A red target subtending 2° of visual angle is completely occluded by a noisy gray target in front of it (left). The relative distances of these targets are varied with conditions described in the text. Detection thresholds (center left) and discrimination thresholds (center right) were estimated from the psychometric functions fitted to the recorded user data (examples shown right).