EyeAR: Empiric Evaluation of a Refocusable Augmented Reality System

[1]Aitor Rovira [1]Rompapas Damien Constantine [1]Christian Sandor*
[1]Takafumi Taketomi [1]Hirokazu Kato [2]Sei Ikeda
[1]Interactive Media Design Laboratory
Nara Institute of Science and Technology
[2]Mobile Computing Laboratory
Ritsumeikan University

Abstract

We present the evaluation of EyeAR, a display with refocusable content based on user's eyes measurements. We carried out a user study to validate the prototype to verify that participants cannot distinguish between real and virtual objects. Participants looked at three pillars (one of which was virtual) placed at different distances from the user. They had to guess which pillar was the virtual one while freely refocusing. The results partially confirmed that our prototype creates virtual objects that are indistinguishable from real objects.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; I.3.3 [Computing Methodologies]: Picture/Image Generation—Display Methods; H.1.2 [Models and Principles]: User/Machine Systems—Human factors

1 Introduction

In 1965, Ivan Sutherland envisioned the ultimate display [6], a multi-modal display that provides visual and haptic feedback that is indistinguishable from real objects. Current technology is far from being able to create such an ultimate display, but recent research has made progress towards providing indistinguishable visual AR, causing some researchers to call for an AR Turing Test to quantify this progress [5].

Commercial interest in AR has increased significantly over the last few years, paving the way for Optical See-Through Head-Mounted Displays (OST-HMDs). One of the largest obstacles to photorealistic AR on OST-HMDs is the mismatch between the Depth of Field (DoF) properties of the user's eyes and the virtual camera used to generate CG. The human visual system relies on several depth cues in order to distinguish which objects are closer than others. Teittinen [7] and Ware [8] discussed the depth cues in detail; one of them, accommodation, is closely related to DoF and refers to the eye changing its shape to change its focal length, thus bringing objects at different distances into focus. Our goal is to generate CG objects that behave like real objects during accommodative actions of the user's eyes. Up until now, this has only been possible through Light Field Displays (LFDs).

Our contribution is the design of a display with refocusable content and the validation of the prototype with a user study that evaluated whether users could distinguish real and virtual objects. The prototype was based on the measurement of the user's focus distance in real time. The results confirmed our hypotheses partially, as one of the three virtual pillars could not be identified better than chance. The other two pillars came close, but were identified above chance, due to the disparity of the distance between the real objects and the display where virtual objects were rendered.

*e-mail:sandor@is.naist.jp

2 System Design

We used an autorefractometer, a device that computes the focal length by measuring the dioptres of one eye in real time. Alternative approaches approximate this value by either measuring stereo vergence [4], or combining gaze tracking with a depth map [3]. Further details of our prototype can be found in [2].

We designed a box enclosure to prevent variations of illumination of the scene caused by external light sources. Inside the enclosure, there was a lamp to provide a controllable light source and three pillars painted uniformly. One was green, the second was blue, and the last one was red, and were placed at depths 0.25m, 0.375m, and 0.5m respectively on a tilted platform to provide a perspective depth cue. On each trial, we replaced one of the three real pillars for a virtual one, which was rendered either in perfect focus or defocusing it based on the autorefractometer readings, exhibiting DoF.

3 Experiment

The main goal of the user study was to assess whether correctly rendered DoF computer-generated objects using were more difficult to distinguish from the real objects than without EyeAR. We conducted a variant of the Graphics Turing Test [1] where participants looked at three pillars (one of which was virtual) placed at three different depths from the user and they had to identify the virtual pillar while freely refocusing (Fig. 1). Our hypotheses were that with the autorefractometer on, participants will not know which one the virtual pillar is and will try to guess it by chance (**H1**); and with the autorefractometer off, participants will be able to guess the virtual pillar correctly more often than having the autorefractometer on (**H2**).

Twelve participants (6 Female) were recruited with ages between 19 and 45 years, with mean±std. deviation 29.7 ± 8.9. All participants claimed to have normal or corrected-to-normal vision, verified with visual acuity tests. They signed a consent form in order to be part of the experiment and were monetarily compensated for their time.

Participants sat down and stared at the scene for 20 seconds per trial through the autorefractometer. They had to focus on the pillars and letters beside them. At the end of each trial, the viewpoint from the autorefractometer was blocked and they had to answer which pillar they thought it was the virtual one. This procedure was carried out twelve times per participant, repeating each permutation of the experimental variables twice.

The dependent variable was the binary outcome of whether participants guessed correctly which pillar was computer generated. The independent variables were VirtualPillar and Autorefractometer. The first one refers to which pillar was virtual, either the red, the green, or the blue one. The second variable indicates whether the data from the autorefractometer were used to adjust the blurriness of the virtual pillar.

4 Results

H1 stated that with the use of the data collected from the autorefractometer, participants would guess correctly about 33% of the times, the same results as if they tried to guess by chance. This hypothesis

978-1-5090-3740-7/16 $31.00 © 2016 IEEE
DOI 10.1109/ISMAR-Adjunct.2016.20

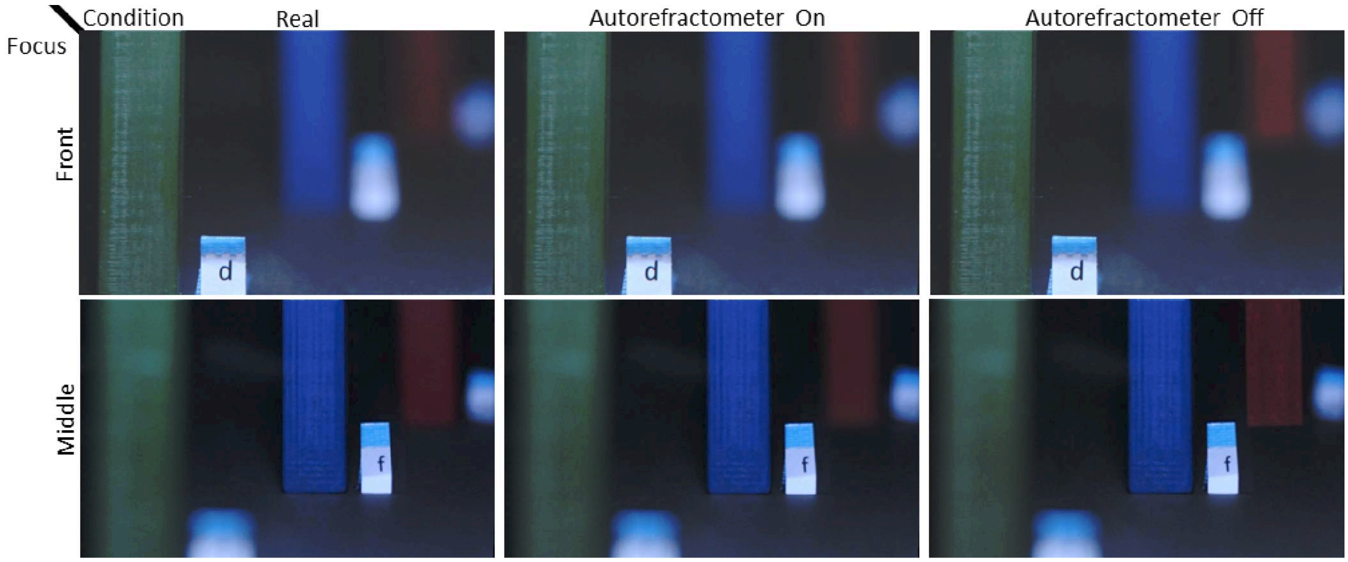

Figure 1: User's view of the pillars. The red pillar looks very similar in reality and when the autorefractometer was on, but very different when the refractometer was off. The red pillar is real only in the first column. The first row shows the user's view when focusing on the front pillar (green), while the second column he focuses on the middle one (blue).

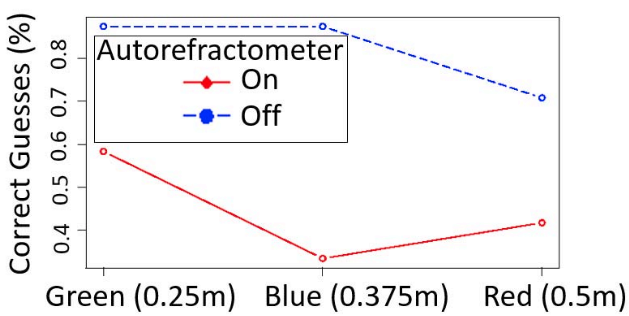

Figure 2: Overall percentage of correct guesses for each pillar.

was *rejected* for the green (58.3%) and red (41.7%) pillars. However, the results for the blue pillar (33.3% of correct guesses) were *compatible with* **H1**.

The results from a regression analysis *support* **H2**, showing that the number of correct guesses was significantly greater when the refractometer was off for all three pillars ($p < 0.001$, conf. interval -2.38 to -1.31). The results also shows a learning effect ($p = 0.008$, conf. interval 0.04 to 0.27), making it easier for participants to guess correctly at the later trials. Additionally, it also revealed that gender is significant ($p = 0.004$, conf. interval 0.28 to 1.47).

Figure 2 shows the percentages of success when guessing which pillar was the virtual one for each experimental condition. When the autorefractometer was off, the overall percentage of correct guess was 81.9%, and when the autorefractometer was on, it was 44.4%. Looking at the pillars separately, participants guessed it correctly more often when the green one was virtual compared to the other two.

Looking at the pillars separately, the green one was guessed correctly 14 times (n=24, sample mean 0.583, 95% conf. int. 0.366 to 0.779), the blue one 8 times (n=24, sample mean 0.333, conf. int. 0.156 to 0.553), and the red one 10 times (n=24, sample mean 0.139, conf. int. 0.069 to 0.241). These results reject that green and red pillar were guessed only by chance, but provide evidence that the virtual blue pillar could not be distinguished from its real counterpart.

5 Discussion

We found support for **H1** when the virtual pillar was the blue one, but not for the other two. We speculate that the distance between each pillar and the semi-transparent display is very important to correct the display-eye focal depth difference. The greater was the error, the easier it was to guess the virtual pillar correctly.

The regression analysis results support **H2**, showing that Autrorefractometer was the strongest predictor in the model. The learning effect and the gender of the participant were also lesser predictors.

In the future, we aim to address the screen-object disparity and reduce the size of the autorefractometer to integrate it in an OST-HMD. We will also carry out more tests with objects of different shapes and textures in more complex scenes.

Acknowledgements

This research was funded by the Japan Society for the Promotion of Science under awards Grant-in-Aid for Scientific Research #15H02737 and Challenging Exploratory Research #15K12084.

References

[1] M. Borg, S. S. Johansen, D. L. Thomsen, and M. Kraus. Practical implementation of a graphics turing test. In *Advances in Visual Computing*, pages 305–313. 2012.

[2] D. Constantine Rompapas, K. Oshima, S. Ikeda, T. Taketomi, G. Yamamoto, C. Sandor, and H. Kato. EyeAR : Physically-Based Depth of Field through Eye Measurements. *Proc. 14th Int. Symp. on Mixed and Augmented Reality*, 2015.

[3] Y. Itoh and G. Klinker. Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization. In *IEEE Symposium on 3D User Interfaces*, pages 75–82, 2014.

[4] A. Maimone, G. Wetzstein, M. Hirsch, D. Lanman, R. Raskar, and H. Fuchs. Focus 3d: Compressive accommodation display. *ACM Transactions on Graphics*, 32(5):153:1–153:13, 2013.

[5] C. Sandor, M. Fuchs, A. Cassinelli, H. Li, R. A. Newcombe, G. Yamamoto, and S. K. Feiner. Breaking the barriers to true augmented reality. *arXiv preprint arXiv:1512.05471*, 2015.

[6] I. E. Sutherland. The ultimate display. In *Proceedings of the IFIP Congress*, pages 506–508, 1965.

[7] M. Teittinen. Depth cues in the human visual system. *The Encyclopedia of Virtual Environments*, 1, 1993.

[8] C. Ware. *Information visualization: perception for design*. Morgan Kaufmann, 2013.