EyeAR: Refocusable Augmented Reality Content through Eye Measurements

**Damien Constantine Rompapas [1],*, Aitor Rovira [1],†, Alexander Plopski [1], Christian Sandor [1],
Takafumi Taketomi [1], Goshiro Yamamoto [1],‡, Hirokazu Kato [1] and Sei Ikeda [2]**

[1] Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma 630-0192, Japan
    aitor.rovira@psych.ox.ac.uk (A.R.); plopski@is.naist.jp (A.P.); sandor@is.naist.jp (C.S.);
    takafumi-t@is.naist.jp (T.T.); goshiro@kuhp.kyoto-u.ac.jp (G.Y.); kato@is.naist.jp (H.K.)
[2] College of Information Science and Engineering, Ritsumeikan University, Kusatsu 525-8577, Japan;
    ikeda.sei.jp@ieee.org
*   Correspondence: damien.rompapas.dk1@is.naist.jp; Tel.: +81-743-72-5336
†   Current address: Department of Psychiatry, University of Oxford, Oxford OX3 7JX, UK.
‡   Current address: Division of Medical Information Technology and Administration Planning,
    Kyoto University Hospital, Kyoto 606-8507, Japan.

Abstract: Augmented Reality (AR) superimposes computer graphics (CG) onto a user's view of the
real world. A key quality problem in this field is to achieve coherence between reality and CG when
the user's eyes refocus or change pupil size. We designed and evaluated a display that improves
coherence by measuring the user's eye state and continuously adapting CG accordingly. Our tabletop
prototype emulates an Optical See-Through Head-Mounted Display, a common AR display device.
In our evaluation, participants observed three pillars at different depths. We then challenged them
to identify a virtual pillar among the three while freely refocusing their eyes. Results show that
our design significantly improved realism. Compared to Light Field Displays, our design aims
to simplify display-optics while providing similar quality. We could only partially achieve this
goal. We discuss the lessons we learned and how we plan to overcome the remaining challenges.
The experimental protocol from our evaluation is useful for display developers as it can be used to
measure the coherence of a display.

Keywords: augmented reality; refocusable augmented reality; optical see-through head-mounted
displays; retinal blur; depth of field; accommodation; ar turing test; perception; psychophysics;
user study

1. Introduction

In 1965, Ivan Sutherland envisioned the ultimate display [1], a display that provides visual and
haptic feedback that can present objects that are indistinguishable from those in the real world.
This vision is similar to the Star Trek Holodeck [2] and has inspired research in the fields of
photorealistic Augmented Reality (AR). 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 [3].

Commercial interest in AR has increased significantly over the last years, paving the way for
Optical See-Through Head-Mounted Displays (OST-HMDs) such as the Google Glass (Mountain
View, CA, USA) [4] and Microsoft HoloLens (Redmond, WA, USA) [5]. OST-HMDs overlay AR
content directly onto a user's view. 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 virtual camera
used to generate computer graphics (CG). Over the years, researchers have proposed a variety of

HMDs that can be refocused [6]. For example, by rendering CG at multiple focal planes, users will experience DoF effects when they focus on objects at different depths. Some displays reproduce the appearance of CG from multiple viewing angles, thus recreating the entire light field (also referred to as Light-Field Displays (LFDs)). These displays are computationally expensive, have small spatial resolution, and introduce unintended noise. Refocusable displays are also difficult to manufacture, for example due to the required resolution of the display and accuracy of the manufacturing process. Most designs were on a conceptual level with bulky proof-of-concept setups. Only a few of these systems have been built into HMDs, and even fewer into OST-HMDs. These prototypes suffer from increased diffraction, contrast, and color issues, as well as bulkiness due to the increased complexity of the optical system [6]. For these reasons, it is no surprise that no refocusable OST-HMDs are currently available commercially. State-of-the-art OST-HMDs present CG on a single focus plane, which is simpler to manufacture, and thus cannot create DoF effects with the built-in optics.

In recent years, there has also been a growing interest in free-focus displays, for example retinal displays [7]. These displays ensure that the content displayed on the OST-HMD always appears in focus to the user. This is especially beneficial for users with eye deformations, like keratoconus or macular degeneration, or those who have to constantly switch focus between CG and the real world, thus leading to increased eye fatigue. As the CG is always presented in focus, it is also not possible to generate DoF effects optically. Therefore, it is necessary to develop a method that creates realistic DoF effects for displays where these cannot be created optically.

Commonly, to project CG from an OST-HMD to the user, the eye is modeled as a pinhole camera, and the image plane coincides with the imaging plane of the HMD [8]. This assumption does not correctly represent the human visual system. Our visual system relies on several depth cues in order to distinguish which objects are closer to us. Teittinen [9] and Ware [10] discuss all 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. Similarly, the size of the pupil diameter is adjusted to control the amount of light entering the eye. Similar to a camera, a wider pupil results in a shorter focus, while a smaller pupil increases the focal distance. Thus, a more accurate representation of our visual system would be a camera with continuously adjusted focal length and aperture size.

In this paper, we present EyeAR, an approach to generate refocusable AR content through the use of eye measurements (see Figure 1). By measuring the user's focus distance and pupil size we can continuously match CG with the user's visual system, thus presenting CG that matches the DoF of the eye. This idea is similar to the work of Kán and Kaufmann [11] who recover the DoF of a camera to present CG that match the camera's DoF. We describe the theory behind EyeAR and how it could be integrated into off-the-shelf OST-HMDs. In this paper, we investigate how well EyeAR can reproduce the DoF on a display with a single focal plane. Contrary to retinal and varifocal displays, the most common type of OST-HMDs presents content in focus only on a single plane. It is thus necessary to understand how well EyeAR performs on these displays and how the disparity between the focus distance and the virtual image plane will affect the results. We use a variant of the AR Turing Test to check if users can distinguish between real and virtual objects and determine the improvement in quality over CG rendered with a pinhole camera eye model.

The realism of rendered content is affected by a variety of cues, such as reflection, occlusion, color consistency, shadow consistency, and DoF consistency, to name a few. As each of these aspects could be used to distinguish between real objects and CG, most visual Turing Tests use photographs to evaluate the quality of rendered content [12,13]. In OST AR, it is not sufficient to ask users to compare photographs of the CG with real objects, as users can also observe the real world. The graphics thus must be affected by the scene illumination, account for the user's focus, and exhibit correct occlusion with real objects.
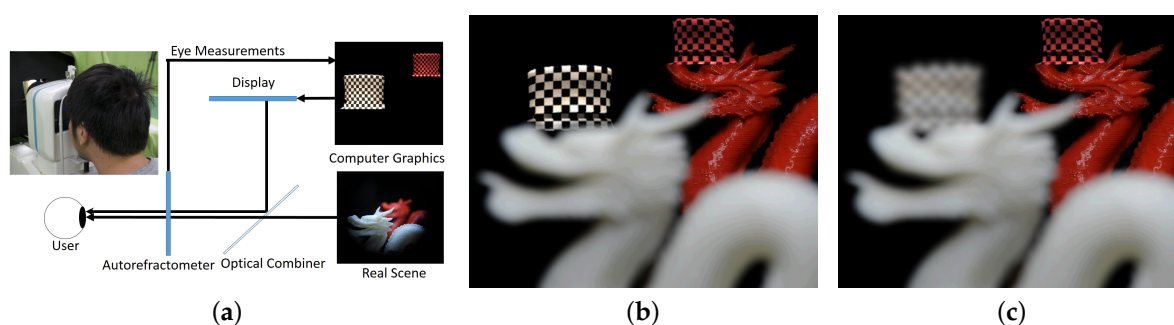
Figure 1. EyeAR aims to embed computer graphics (CG) objects into the real world that are indistinguishable from real objects. (**a**) schema of the EyeAR concept: we measure the user's pupil size and accommodative state and then feed these values into a real-time ray tracer that generates matching CG that is displayed via a half-mirror. In an example application, we augment real dragon statues with virtual hats of the same color; (**b**) conventional optical see-through displays exhibit a Depth of Field (DoF) mismatch between CG and the real scene. Note how the white hat is in focus, while the white dragon is blurred; (**c**) EyeAR accurately matches the DoF of virtual objects, correcting this mismatch.

In our setup, we tried to eliminate cues that could help users distinguish between real objects and CG and focus on the impact of the estimated DoF while users freely refocus between objects at different depths. We asked participants to look at three pillars (one of which was virtual) placed at three different depths (see Figure 2). We then challenged users to identify the virtual pillar while allowing their eyes to freely change focal depth. We compare CG rendered with EyeAR and the pinhole eye model. The results of our experiment confirm our expectation that CG rendered based on eye measurements are more realistic than those rendered with the pinhole eye model. We also found that the blue pillar could not be identified as real or CG better than chance if it was placed at the same distance as the display; the other two pillars came close, but were identified slightly above chance. This could be due to the disparity of focal lengths between the real objects and the virtual image plane, the low update rate of the autorefractometer that we used to measure the focus distance, or unintended artifacts, such as minor differences in texture or illumination of the real objects and the generated CG.

This paper describes the core idea and theory that has been presented both as a tabletop prototype [14] and as a partially implemented OST-HMD integration [15]. This paper also describes in detail the AR Turing test from [16].

Contributions:

- We show that, for OST AR, rendering CG based on measurements of the user's eyes improves the perceived realism of the observed graphics. In particular, CG rendered with EyeAR were always perceived to be more realistic than CG rendered with the pinhole eye model.
- EyeAR is the first OST system to create AR CG that are coherent with the real environment while refocusing, albeit only when the position of virtual objects coincided with the focus distance of the display.
- Our experiment is the first instance of an AR Turing Test aimed at testing refocusable content for OST AR setups.

We captured the first person illustrations of our display with a Nikon D700 SLR camera (Shinagawa, Tokyo, Japan) through the display itself and did not apply any post processing. We set the SLR's aperture size and f-stop to typical values obtained from users of user testing.
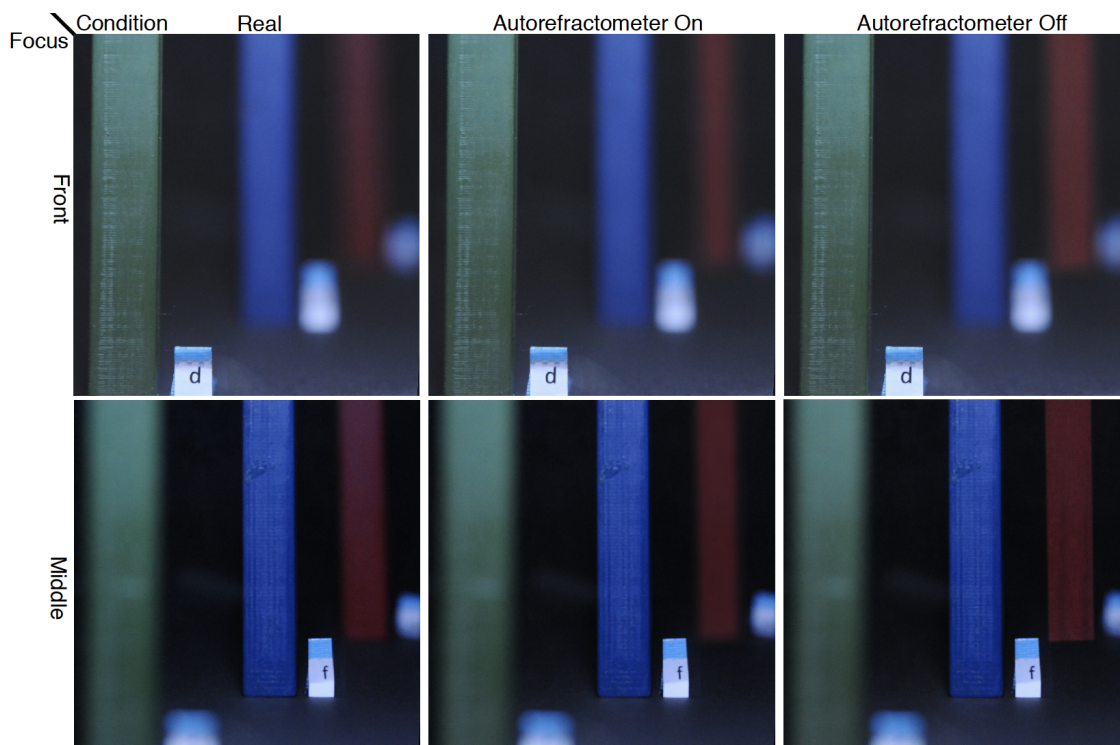
Figure 2. Participants' view during the experiment for two different focus distances (front: green pillar, middle: blue pillar). The green and blue pillars are real; the red pillar is only real in the leftmost column. Note how the red pillars in columns Real and Autorefractometer On closely match. For comparison, column Autorefractometer Off shows views similar to most OST AR displays; note the mismatch compared to the red pillars in the other columns.

2. Related Work

Our work is related to research in OST-HMD design, DoF effects, and Visual Turing Tests that are commonly applied to CG. In this section, we first discuss related work that shares the goal of creating refocusable AR content. Second, we discuss previous studies that evaluate the impact of DoF effects on user perception. Finally, we discuss experimental protocols for visual Turing Tests that validate the realism of CG.

2.1. Displays for Refocusable AR Content

A lot of research has focused on the development of displays capable of reproducing focus cues. Kramida [6] and Hua [17] provide a detailed overview of different approaches and technologies developed to enable refocusing onto virtual content presented at different depths and natural DoF effects.

A common approach to present depth cues is by displaying virtual content on multiple transparent planes [18–20]. Although MacKenzie et al. [21] showed that five focal planes are enough to produce an acceptable range of real-time accommodative cues, multiple focal planes lead to a bulkier HMD design. Our method differs from these in that we target off-the-shelf OST-HMDs that have a single focal plane that cannot present refocusable CG.

Another approach to create refocusable OST-HMDs is to use a refocusable lens, in combination with either a static or a movable display [22,23]. This allows to adjust the distance at which the virtual content appears. However, the lens refocuses the entire display at the same time, thus providing a realistic DoF that requires a very high update rate, which has not been achieved yet, and must be synchronized with the display to prevent unintended effects. In [22], the authors also noted that it is important to evaluate the effectiveness of rendered DoF compared to DoF generated by optical

elements. We applied a similar approach to our method, as we showed in [14]. By matching the position of the display with the distance on which users were focused on, we created realistic DoF effects at different focus distances.

McQuaide et al. [24] used a method that is very similar to ours. They combined a laser projector that shows the user an image that is always in focus with deformable micro-electromechanical system (MEMS) mirrors. They adjusted the focus of the laser beam by manipulating the convergence of the MEMS elements. This required users to refocus to see a sharp image. The idea of our work is very similar in that we assume that the user always observes a sharp image. Our method differs from theirs in that we present the DoF not through optical means, but with CG.

Finally, our method is also very similar to that of Kán and Kaufmann [11]. They obtained the camera's lens parameters and rendered CG using these camera parameters. The resulting CG's DoF matched the DoF properties of the video in each frame. Our work crucially differs from [11]. We created an OST system and also obtain the parameters of the user's eyes instead of a camera.

2.2. Impact of DoF Rendering

The impact of DoF rendering and refocusable displays has been studied in AR, Virtual Reality (VR), and video games. Padmanaban et al. [25] used a display with refocusable lenses to investigate the impact of refocusable graphics on users to go even further and personalize the accommodation to each individual to correct vision impairments.

Eye-gaze tracking (EGT) has been used in combination with a variety of VR systems to estimate the focus position and generate gaze-contingent DoF effects. Hillaire et al. [26] found that users prefer graphics rendered gaze-contingent DoF effects over the CG in focus. On the contrary, just applying DoF effects [27] did not improve user performance. In particular, some users expressed fatigue and discomfort caused by the DoF effects. Mauderer et al. [28] found that gaze-contingent DoF effects increase realism and help estimate the depth of virtual objects. Vinnikov and Allison [29] found that although DoF effects improve depth perception in monocular systems, users felt discomfort when these were applied to stereoscopic systems.

Hua and Liu [30] found that DoF helps estimate the depth of virtual content in AR, which is similar to a VR scenario. In a similar study, Xueting and Ogawa [31] found that users prefer graphics with too much blur applied to them over the correctly applied amount of blur. Our study differs from these works in that we do not investigate the user's depth perception, but whether they can distinguish between the CG and real objects through an AR Turing test. Furthermore, in [31], the evaluation was performed on images rendered on a display, and not on an OST-HMD. Therefore, users did not experience inconsistencies between the focus distance and the presented CG.

2.3. Visual Turing Tests

In 1950, Alan Turing [32] introduced the Turing Test. In this test, he proposed an "Imitation Game" that can be used to test the sophistication of artificial intelligence (AI) software. In this test, human participants interact with a conversation partner and they have to guess whether the partner is human or an AI simulation. The test is passed if the chance of choosing the AI correctly became equivalent to a random guess.

Similar tests have been proposed in a variety of fields. For example, for visual computing [12], where participants were tasked to distinguish photographs from renderings of 3D reconstructions of buildings. In Computer Graphics, McGuigan [33] performed a Visual Turing Test with the following hypothesis: "The subject views and interacts with a real or computer-generated scene. The test is passed if the subject cannot determine reality from the simulated reality better than a random guess".

Several previous studies proposed restricted versions of this test; Meyer et al. [34] conducted a study which required participants to view a physical setup as well as a setup displayed on a color television. Although this study showed that subjects were unable to distinguish between the physical and displayed setup, this was largely because the physical setup was viewed through a camera.

Borg et al. [13] created a practical CG Turing Test which made use of a box that contained a controlled light source, a removable end which can be replaced with a screen, viewport consisting of a small pin hole and simple, and a real object (which is duplicated in a virtual scene). The participants were asked to look through a small hole using one eye for 10 s and they had to guess if the scene observed was rendered or real. The results showed that participants were generally unable to accurately distinguish the virtual scene from the real scene better than a random guess. A variant of this test was shown by Nvidia (Santa Clara, CA , USA) [35]. They presented two box enclosures with controlled lighting; one scene was a physical scene featuring an electronic drill and the other one featured the same scene as CG. Participants did not view the boxes directly, but were only looking at prerecorded photos.

In spirit, our AR Turing Test follows the protocol of McGuigan's Visual Turing Test. An important difference is that we perform our test in an OST AR setup. To our knowledge, this is the first time that an AR Turing test was performed in such situation. We also reported the design and results of this user study in [16].

3. System Design

We show the workflow of EyeAR in Figure 3. EyeAR is a closed-loop system with the following basic steps:

1. Measure the user's eye
2. Render CG with correct DoF
3. Correct screen-object disparity
4. User observes graphics
5. Goto 1

In the following subsections, we explain the design of each non-trivial step.
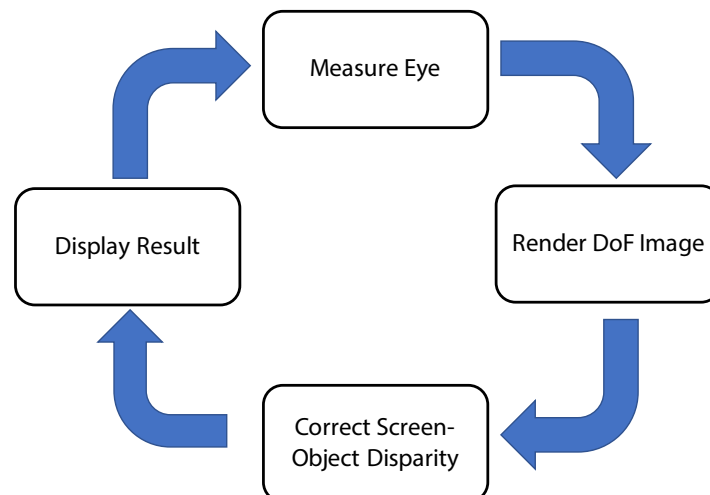
Figure 3. EyeAR is a closed-loop system: Given eye measurements, we generate a realistic DoF image. After correcting the offset between the screen and the virtual object, we present the results on the screen, which triggers a response by the user's eye.

3.1. Measuring the Eye

Accurately measuring a user's focal distance is a very challenging task. Most available EGT devices, e.g., Tobii Glasses (Danderyd, Stockholm, Sweden), estimate the pupil size accurately, but provide only inaccurate estimations of the focus depth from intersection of the estimated gaze directions. Some methods use the variance in the reflection of infrared LEDs on the lens of the eye to estimate the focus depth [36]. Others distinguish between different focus planes [37,38] or estimate

the focus through intersection of the gaze with scene geometry [39]. However, they do not provide continuous measurements that are necessary for correct DoF representation. The most accurate solution to measure the focus distance is an autorefractometer. Although there are different designs, the majority measure the appearance of a ring emitted with near-infrared light on the retina. The autorefractometer adjusts the position of various focusing lenses until the image appears in focus to determine the focus distance of the eye. An autorefractometer can estimate the focus distance, with an average error of only 0.25 D. It is often used in medical examinations and was also used to verify the focus distance in refocusable HMDs [21,23,24]. Due to the required optical elements, even the smallest commercially available autorefractometers are too bulky to fit into an OST-HMD and have a low refresh rate of only 5 Hz.

EyeAR could be applied with any of the methods mentioned above, as long as the system is capable of acquiring accurate measurements of the user's focus distance, e.g., through [36,37]. Although it is possible to build a portable prototype of the system using eye tracking solutions, we opted to create a tabletop prototype that uses the autorefractometer, as it offers the highest accuracy and reliability among the available options.

3.2. Rendering

The aim of our rendering component is to create CG with a DoF that matches the user's view of the real world. Technically speaking, we need to match the camera parameters used to create CG to the parameters of the user's eye.

The human visual system relies on several depth cues in order to distinguish which objects are closer to us than others. Teittinen [9] and Ware [10] discuss all 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. The human visual system always focuses at a distinct depth. Therefore, objects that lie at different depths appear blurred (see Figure 4). They exhibit a circular defocus, commonly called the Circle of Confusion (CoC).
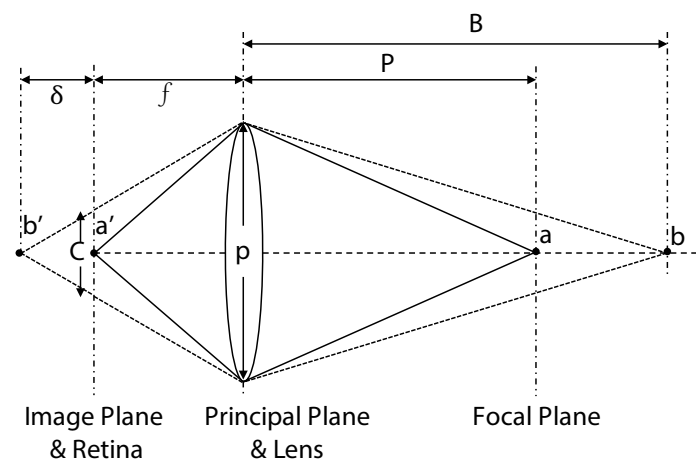
Figure 4. Camera model used in our renderer (adapted from [40,41]). A point *a* lying on the focal plane, located *P* away from the eye, appears in focus as *a′* on the retina. Point *b*, located *B* away from the lens is focused onto *b′* and is blurred by the CoC, *C*.

Figure 4 shows the camera model used in our renderer. The eye is represented as a camera with a single principal plane with focal length f and a circular aperture p. Point lights on the focal plane such as a are projected onto a point a' on the user's retina. Point b refers to the virtual image of a distant point light that would be focused δ behind the retina, resulting in a CoC with diameter C. C is proportional to both the pupil size p, as well as the ratio of distances of focal plane and point light.

We can also express the second proportional term as the fraction of δ and $\delta + f$, with f being the focal length of the eye, resulting in:

$$C/p = \delta/(f + \delta). \tag{1}$$

Isolating C and substituting $\delta/(f + \delta)$ with ΔF, we obtain:

$$C = \Delta F p. \tag{2}$$

Therefore, in order to create CG that matches the state of the user's eye, we must measure its focal length and pupil size to determine the CoC. We then feed these values into a distributed ray tracer [40].

3.3. Correcting Screen-Object Disparity

Most available OST-HMDs display CG on a fixed focal plane. In some models, the distance of this focal plane can be adjusted, for example in the Brother WD-100 (Bridgewater, NJ, USA). However, the principle remains the same. When the user is not focusing on the same depth as the virtual display, the content on the HMD-screen becomes blurred. As our rendering algorithm assumes that the display position coincides with focus distance, we attempt to correct this disparity.

In our demo at ISMAR2015 [14], we physically adjusted the position of the display to match the focus distance at any given moment. The moving rail in our prototype could adjust the screen position at a speed of 4 cm/s, which we found to be too slow. This coincides with the observations made by Shiomi et al. [42], who found that users could accommodate on a target moving from a position of 50 cm to 1 m at a speed of 10 cm/s.

As an alternative, we tried SharpView [39]. SharpView estimates the amount of blur caused by the difference between a user's focus distance and the screen location. It applies a sharpening effect to the presented image, so that, when the image is blurred by the estimated amount, it coincides with the intended view. In our trials, we found that although SharpView improves the visibility of the texture of the objects, it tends to create unrealistic sharpening artifacts that clearly distinguish CG from real objects. This is likely because the estimated amount of blur did not perfectly match the amount of blur perceived by participants. This coincides with the observation made in [39]. In our trials, participants could always find the virtual objects.

For the reasons explained above, we decided not to correct the disparity problem in this experiment. Instead, we carefully minimized depth variance in our experimental scene, in order to minimize the impact of the screen-object depth disparity (see Figure 2).

4. Experimental Platform

Our experimental platform consisted of the following elements:

Measuring the Eye. We use the Grand Seiko WAM-5500 autorefractometer (Fukuyama City, Hiroshima, Japan). It measures pupil size p and focal length in diopters D at 5 Hz with an accuracy of ± 0.25 D. It also allows us to read these measurements over a serial connection. We then calculate the focus distance as $P = 1/D$ and transmit p and f to the rendering component.

Rendering System. We implemented the rendering pipeline on a desktop computer with an Intel Core i7 3790K processor (Santa Clara, CA, USA), 8 GB RAM and an NVIDIA GTX 980 graphics card.

Our implementation of distributed ray tracing [40] is based on NVIDIA's Optix [43] framework. For each point on the retinal image, we combined ray tracing results from 32 samples uniformly distributed over the pupil. We additionally use bi-directional reflectance functions [44] for materials. To account for the slow measurement speed of the autorefractometer, we linearly extrapolate the values received over the past second to predict the next observation. We then interpolate from the current parameters towards the predicted values. We rendered all images at 30 fps and transferred them to the display in our experimental platform.

Physical Setup. Our experimental setup is shown in Figure 5a. The setup was surrounded by an enclosure to prevent variations of illumination of the scene caused by external light sources. Inside the box enclosure, an LED lamp was placed to provide a controllable area light source. We display the graphics on a Dell Venue 8 (Round Rock, TX, USA) organic light emitting diodes (OLED) tablet with a resolution of 2560 × 1600 and 359 dpi and a 50 R/50 T optical mirror to create an optical see-through display with a fixed display depth. The display was positioned 0.375 m away from the participant. Inside the box enclosure, we placed three 3D printed pillars painted in uniform colors. One was green, the second was blue, and the last one was red, and they were placed at depths of 0.25 m, 0.375 m, and 0.5 m, respectively, on a tilted platform to provide a perspective depth cue (see Figure 5b). In each trial, one of the three real pillars was replaced with a computer-generated counterpart, which was rendered either in perfect focus or using the measurements taken from the autorefractometer, exhibiting DoF. Given the standard error of the autorefractometer of 0.25 D, the expected focus measurements for each pillar were:

- 23.5–26.6 cm for the green pillar,
- 34.2–41.3 cm for the blue pillar, and
- 44.4–57 cm for the red pillar.
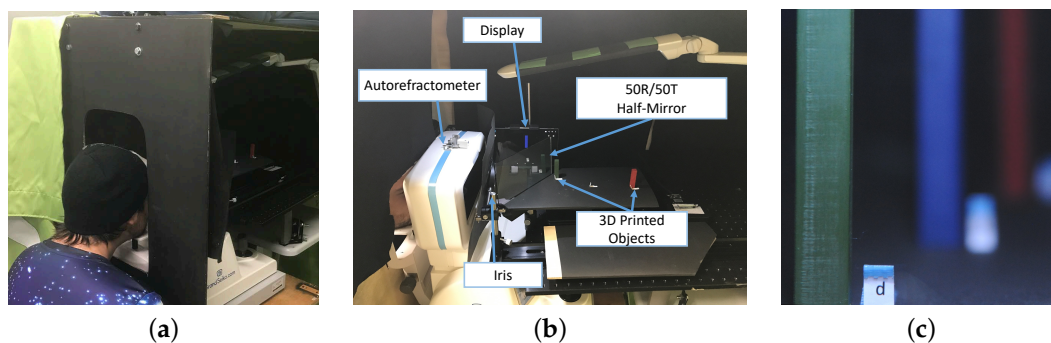
Figure 5. Our experiment setup: (**a**) a user sitting in front of the autorefractometer and looking inside the box enclosure (the side panel was opened for illustrative purposes); (**b**) internal view of the box enclosure; (**c**) user's view.

Liu et al. [23] showed that measurements provided by the autorefractometer fluctuate near the ground-truth focus distance. We thus average the measurements over the past 2 s to account for this fluctuation.

Humans use multiple depth cues such as accommodation, vergence, overlap, and retinal image size to recover the depth of objects [9]. Our setup was designed to eliminate shadows, aerial perspective, overlapping, and linear perspective cues. Additionally, to prevent movement and binocular parallax cues, participants had to wear an eye-patch over their non-dominant eyes and place their heads onto a chin rest. An iris diaphragm behind the autorefractometer ensured that the participant's field-of-view was limited to the experiment setup. To prevent texture depth cues, the pillars had a uniform color. The pillars were aligned manually before each trial and this could cause their position to change slightly. This unnoticeable displacement helped prevent the participants from guessing the virtual pillars due to the aberrations of their eyes. To assist participants in refocusing at different depths, the letters placed next to the pillars provided a texture cue, and the relative size of the pillars provided retinal image size cues. These cues allow participants to adjust the focal length of the eye to refocus at different distances. We could thus measure the participant's accommodation to determine the focus depth.

5. Experiment

The purpose of this experiment was to assess whether correctly rendered DoF computer-generated objects using EyeAR were more difficult to distinguish from the real objects than those generated

without using EyeAR. For this, we conducted a variant of the graphics Turing Test [13]. Our hypotheses stated that:

H1 With the autorefractometer off, participants will guess the virtual pillar correctly more often than when the autorefractometer is on.
H2 With the autorefractometer on, participants will correctly guess the virtual pillar no better than random chance.

5.1. Participants

We recruited twelve participants (6 female, 6 male) between 19 and 45 years, mean 30.2, and standard deviation of 9.2, from both the students at the university and the general public. All participants claimed to have normal or corrected-to-normal vision with the use of contact lenses. We verified this with visual acuity tests under three conditions (see Section 5.2). The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Nara Institute of Science and Technology (Ikoma, Japan). Participants signed a consent form, and were monetarily compensated for their time.

5.2. Preliminary Tests

Our system requires reliable measurements of the participants' dominant eye with the use of the autorefractometer. We verified that we could read data over a range of diopters. We also confirmed that they were able to focus on the objects within the range used in the experiment. All volunteers had to pass four preliminary tests before taking part in the experiment. The first test allowed us to verify the participant's dominant eye. The second and third verified the participants visual acuity for far and near sight, respectively. The final test verified that the autorefractometer was capable of reading diopters over the experimental depth range (0.25–0.5 m). The total time needed for these tests was 10 min per participant. The tests are described as follows:

Eye-dominance test. Each participant stood 3 m away from a marked object facing towards it. They were then asked to hold their hands 50 cm away from their eyes with the thumb and index finger forming a connected arch, and looked at the marked object through the arch. This caused participants to hold their hands biased towards their dominant eye. We consider the eye used to look at the object their dominant eye. The dominant eye was then used on the later preliminary tests, with the non-dominant eye covered by an eye patch.

Acuity test. Once the dominant eye was determined, each participant carried out two standard tests used in optometry to measure visual acuity. In the first, they stood 2.8 m away from a Snellen chart held at eye level. If the participant was not able to read a letter, they guessed it. The test finished when he or she failed to read more than half of a line or was able to read the entire chart. We only accepted participants with a visual acuity of at least 20/30.

The second acuity test measured participant's short-distance visual acuity. Each participant sat down on a chair 40 cm away from a Rosenbaum chart placed at eye level. They proceeded to read the numbers on each line. If they could not read the line, they had to guess it. The test had the same ending conditions as the previous one, with participants excluded if their determined visual acuity was not better than 20/30.

Operability test. The fourth and last preliminary test verified that the autorefractometer could accurately read the participant's dominant eye. The measurements can sometimes be affected by participants with corrected-to-normal vision with the use of either contact lenses or spectacles, or when the person is not able to focus on the objects placed within the range limit accepted by the autorefractometer, between 0.25 m and 0.5 m. Each participant was asked to look at a chart hanged from a horizontal beam. The dominant eye was measured while the card moved steadily along the beam, starting from 0.5 m away and moving it towards the eye (up to 0.25 m), and then back to the original position over an interval of 3 s.

5.3. Task and Procedure

Participants sat down in front of the autorefractometer and looked at the scene inside the box enclosure. The scene composed of three pillars with letters besides them. Participants were instructed to focus on the pillars and letters beside them for a total of 20 s per trial. At the end of each trial, a researcher occluded the scene and in a separate room the participant wrote on a sheet which pillar they considered to be virtual. While they answered, the researcher changed the pillars according to a sequence randomly generated in advance. The procedure was then repeated twelve times for each participant—from which each permutation of the experimental variables was repeated twice per participant. Each session took 40 min per participant in total.

5.4. Variables

Our experiment was designed as a *within-subjects* experiment. We looked at whether participants correctly guessed the pillar that was computer generated as a binary outcome. We also collected general feedback from each participant with the goal of gaining insight on the results from the evaluation. The independent variables of our experiment were as follows:

VirtualPillar $\in$ { red, green, blue}
　　This refers to the pillar that was virtual.
Autorefractometer $\in$ { On, Off}
　　We evaluated two situations: one adjusting the blurriness of the virtual pillar according
　　to the autorefractometer readings and the other without adjusting it, simulating a pinhole
　　camera rendering.
TrialSequence $\in$ { 1 ... 12}
　　We included the sequence number in which the trials were carried out to observe whether there
　　was a learning effect.

5.5. Results

Table 1 shows the results of the analysis of the recorded data. We use $p < 0.05$ as a criteria to determine statistically significant results.

Table 1. Coefficients and *p*-values of the experimental variables of a linear model fitting using generalized estimating equations with correlation structure = exchangeable. The *p*-values show that Autorefractometer is the main contributor to the model but not the only one, also gender and TrialSequence are significant.

	Estimate	Std. Error	Wald	Pr (>|W|)	95% Conf. Interval
Intercept	−0.561	0.565	0.99	0.3202	−1.67 to 0.55
TrialSequence	0.152	0.058	6.96	0.0083	0.04 to 0.27
Autorefractometer = on	−1.841	0.273	45.65	$< 10^{-10}$	−2.38 to −1.31
VirtualPillar	0.236	0.158	2.23	0.1357	−0.07 to 0.55
Age	0.010	0.013	0.58	0.4475	−0.02 to 0.04
Gender	0.872	0.304	8.23	0.0041	0.28 to 1.47

The results of the regression *support* **H1**. The results show that the number of correct guesses was significantly higher when the autorefractometer was off for all three pillars. We also found that there was a learning effect, making it easier for participants to guess the virtual pillar in later trials. The analysis also revealed that gender was also significant, which was not initially expected.

H2 stated that, with the use of the data collected from the autorefractometer, participants would correctly guess about 33% of the time, the same results as if they tried to guess by chance. This hypothesis 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* **H2**. In all cases, CG rendered

in the "Autorefractometer = on" condition were recognized less often than CG rendered in the "Autorefractometer = off" condition.

Figure 6 shows the percentages of success when guessing which pillar was the virtual one for each experimental condition. At first sight, the plot shows that participants found it easier to guess the virtual pillar when the refractometer was off (81.9% of the times) compared to using the data from the refractometer (44.4%) to adjust the blurriness of the pillars in real time. Concerning the pillars, participants guessed it right more often when the green one was virtual compared to the other two. Looking at **H2**, the total number of correct guesses for "Autorefractometer = on" was 32, $n = 72$. We computed the confidence intervals using different methods to test whether the probability to guess it correctly could be 0.33. The Wilson score interval method returned a sample mean of 0.44 with a 95% confidence interval = 0.335 to 0.559. This result rejects that 0.33 is the expected mean value of the population by a very narrow margin, therefore rejecting **H2**. The Clopper–Pearson binomial interval returns a very similar value (95% conf. int. = 0.327 to 0.566), but it does not reject the hypothesis, although it is still very unlikely that the true mean is 0.33.
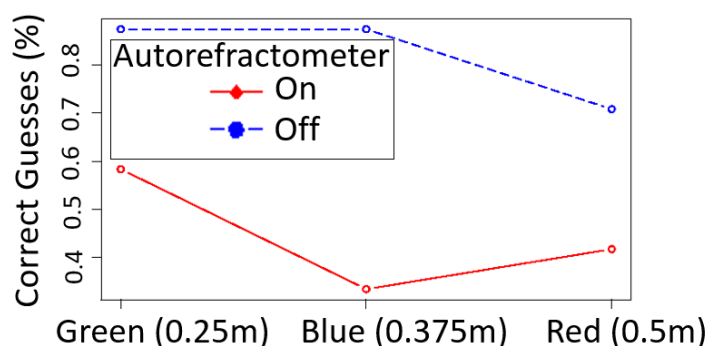
Figure 6. Overall percentage of correct guesses for each pillar when the autorefractometer was on (red line) and off (blue line).

We now take a look at the pillars separately. With "Autorefractometer = on", the green pillar was guessed correctly 14 times ($n = 24$, sample mean = 0.583, 95% conf. int. = 0.366 to 0.779), the blue pillar 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 = 0.417, conf. int. = 0.221 to 0.634). These results reject that green and red pillars were guessed only by chance, but provide evidence that the virtual blue pillar could not be distinguished from its real counterpart.

This was a within-subjects experimental design with a binary dependent variable. Then, we decided to estimate a model using generalized estimating equations (GEE) [45,46]. The GEE analysis provides statistical estimation similar to repeated-measures ANOVA, but can achieve higher power with a lower number of repeated measurements [47]. Besides the two main independent variables, we also included the trial sequence (*TrialSequence*), gender, and age to test whether they could also be explanatory variables. A standard criteria to test the model fitting is the Akaike information criteria, but it has been argued that it is not applicable when using GEE, as this method does not provide models based on likelihood. Pan [48] proposed that the quasi-likelihood information criterion (QIC) suits better for this case. Other criteria have been proposed such as the correlation information criterion (CIC) [49], although it is better take into account more than one criterion [50]. Table 1 shows that the main contributor of the linear model with no interaction terms was the experimental variable *Autorefractometer*, although it is not the only one. The trial sequence and gender are also contributors, though they have a lesser effect. On the other hand, VirtualPillar and age do not contribute. This model returned criteria values QIC = 164.91 and a CIC = 4.01. Removing the non-contributors and adding interaction terms for the predictors of the previous model returns that none of them are significant (Autorefractometer•Gender $p = 0.439$; Autorefractometer•TrialSequence $p = 0.804$;

TrialSequence•Gender p = 0.902). This second model returns higher scores on both QIC (169.14) and CIC (6.34), which means that the second model fitting is not as good as the previous one. One last model was tried with only Autorefractometer, TrialSequence, and gender, returns the lowest criterion scores, with QIC = 164.54 and CIC = 3.34, having again Autorefractometer ($p < 10^{-10}$) the strongest predictor, but both TrialSequence (p = 0.022) and Gender (p = 0.013) also need to be taken into account.

5.6. Discussion

The results of our experiment support **H1**. We found that participants were less likely to detect the CG content when rendered based on readings from the autorefractometer. We were surprised by the bad performance of the "Autorefractometer = off" condition for the central pillar. One would expect that if the pillar is rendered in a photorealistic manner, detection results would be identical to placing a picture at the position of the display. Upon further investigation, we concluded that the pillar rendered under the pinhole camera assumption did not appear realistic enough as it looked too sharp. This also indicates that a pinhole eye model is not suitable for rendering realistic objects for AR. In the future, we want to investigate the improvement of realism taking only pupil size into account.

Our results only partially supported hypothesis **H2**. This can be attributed to various factors, such as the disparity between the user's focus and display depth, appearance of the pillars, low update rate of the autorefractometer, and update delay of the resulting CG displayed. These issues could also explain the observed learning effect. Participants were more likely to detect the difference between the real objects and their CG counterparts as they became used to experiments. The results of our experiment suggest that future OST-HMDs, especially HMDs that have a single focal plane, should incorporate the concept of EyeAR to increase the realism of the rendered CG. This may range from using a non-pinhole eye model to generating CG based on measurements of the eye focal length and pupil radius, as reported in this paper.

The most likely reason that **H2** was not supported for the front and back pillars is the depth disparity between the position of the virtual pillars and the display. As discussed in Section 3.3, we did not correct the different depths of the participant's focus and display image, which may have caused unnecessary blur when looking at the front or back pillars (see Figure 7). Although the same condition applied to the central pillar, these effects are more prominent for the front and back pillars, which led to the higher detection rate. In the "Autorefractometer = on" condition, participants could not detect the blue pillar that coincided with the focal plane of the display and were less likely to detect the red pillar in the back than the green pillar in the front. We speculate that EyeAR can create realistic effects within a certain region around the actual focal depth of the display. In the future, it will be necessary to investigate how large this region is, if it is ratio or diopter-based, and at what point the benefits of rendering content with EyeAR are no longer detectable.

| (**a**) Focus front | (**b**) Focus middle | (**c**) Focus back |
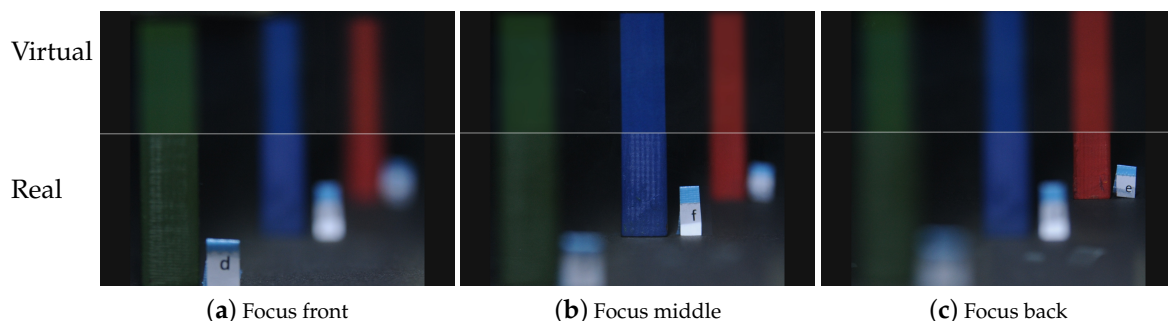
Figure 7. Limitation of our prototype: as the distance to the virtual image plane coincides with the blue pillar, it is the only pillar that can be rendered with only minor artifacts. Note how the red and green pillars exhibit noticeably greater artifacts. The top row shows all virtual pillars, while the bottom row only shows real pillars.

The higher detection rate could also be explained by the slight difference in the appearance of virtual pillars—the texture, environment illumination, refractive optics of the eye, and the sensitivity to different colors, among others. All affected the appearance of the virtual pillars. Although the design of our setup aimed to control the lighting conditions and we optimized the texture and appearance of the pillars to match that of their real counterparts, participants might have noticed slight differences in the color, hue, or texture of the pillars. This is also supported by the learning effects revealed by our analysis. Over the course of the experiment, participants might have become better at detecting the slight differences and cues that distinguished the pillars. We considered switching the pillars to counteract the learning effect and different sensitivity. However, this could have introduced other unintended effects. For example, artifacts introduced when optimizing the appearance of the various pillars could have influenced the results of the evaluation. Overall, our analysis did not reveal that the color of the pillar had a significant effect on the overall results ($p = 0.1357$). While the color might have contributed to easier detection of the front pillar, our results indicate that rendering the content based on the readings from the autorefractometer, or with the pinhole camera model of the eye had the strongest effect ($p < 10^{-10}$). In future experiments, we plan to investigate what effects the pillar color had when content was rendered using the readings from the autorefractometer to adjust its appearance and if rendering content with a non-pinhole eye model can achieve similar results without adjusting the blur effects of CG content.

The slow update rate of the autorefractometer is another factor that could have impacted the results of our study. For one, when participants refocused between the pillars, it could take up to 0.2 s for the system to register this change. Combined with the interpolation that was added to reduce the detection of changes in the rendered content, it created sufficient delay for participants to become aware of it over the course of the experiment. This could also explain the observed learning effects. At the same time, **H2** was supported for the central pillar. This suggests that the delay may not have been noticeable to the participants all the time, but potentially only when participants were observing large changes, e.g., refocusing from the front to back pillar.

Nonetheless, our results show that accounting for focus distance measurements improves the realism of the CG. We expect that faster autorefractometers and improvements to focus depth estimation from eye-gaze tracking cameras will enable focus updates at a speed that is not noticeable by participants. Our results highlight the need for such technology. At the same time, the current measurement speed is enough when participants change focus between objects that are far away and very close by, e.g., at hand's reach. In that case, the maximum refocusing speed of the eye is less than 1.878 ± 0.625 diopter/s [51], which is slower than the update rate of our system. Our analysis also revealed that the gender had a significant impact on the detection rate. Our goal was to evaluate how likely users were to detect content rendered with EyeAR compared to content rendered with a pinhole camera eye model. Therefore, we did not counterbalance the participants in terms of age and gender. This might be a reason for the significance we detected in gender and we will investigate how age and gender affects the results in future work.

6. Conclusions

There is a need for refocusable CG for OST AR in order to create realistic scenes where objects are placed at different depths. In this paper, we have presented the design of a display concept based on the user's eye measurements as an alternative to light field displays. We implemented a tabletop prototype that emulates an OST-HMD setup that is capable of accurately matching the DoF of virtual objects to real objects. We then evaluated our prototype with a user study to verify our claims. Our results strongly support **H1**, which stated that our closed-loop system creates significantly more realistic renderings than a system that does not take into account the user's eyes measurements. Our second hypothesis **H2** was rejected for the pillars in the background and foreground. This is likely due to the reconstruction error in CG caused by the screen-object disparity. Other aspects that

could have contributed to this are the slow update rate of the autorefractometer, color of the pillars, or unintended artifacts.

Referring back to our previously published taxonomy of AR displays [3], EyeAR is an instance of a Personalized AR display, as opposed to the LFD approach, which is an instance of Surround AR. We are convinced that our approach is an interesting alternative to create realistic AR content. The main advantage is that EyeAR does not require complex optical systems and thus addresses many problems in existing refocusable concepts. On the other hand, EyeAR requires very accurate estimation of the user's focus depth and, in the best situation, a display that can adjust its location so that the image plane of the OST-HMD always matches the user's focus distance.

We have identified three main areas for future work: first, and most important, we need to improve the prototype to address the screen-object disparity. Second, we plan to integrate EyeAR into an OST-HMD by either miniaturizing, or replicating, the function of the autorefractometer, for example through use of eye tracking cameras as described in [36,37,39]. Third, we will refine our methodology for conducting AR Turing Tests and carry out several more.

Our rendering algorithm assumes that the position of the display coincides with the position of the virtual object. This effectively restricts its applicable range, as it cannot correct large disparities between the screen and the focus distance. In the future, we want to evaluate the effective range where these disparities become noticeable. Additionally, we aim to investigate how EyeAR could be combined with free-focus OST-HMDs. We imagine that EyeAR could be applied in retinal displays, or with refocusable lenses designed to always present content shown on the OST-HMD in focus. Determining the applicable range of EyeAR could also lead to a combination of EyeAR with multi-focal and varifocal displays in order to reduce the optical complexity of the system.

Alternatively, EyeAR could be combined with SharpView for single focal-plane OST-HMDs. However, we still need to investigate how to build user-specific point-spread functions (PSFs), including 3D PSFs, and model the dynamic accommodation process. In order to achieve the required optical power, we are now considering programmable optical elements. This approach has been successfully demonstrated in the dual problem of increasing the DoF of projectors through coded apertures [52] and fast focal sweep based on a shape-changing lens [53].

In terms of portability, we aim to study how to reduce the size of our system to the point that it can be integrated into an HMD. We envision an EyeAR hardware module that can turn any legacy OST-HMD into a powerful display, perceptually equivalent to a LFD. Additionally, the update rate of our system (5 Hz) can be sufficient to refocus between distances several meters away from the eye, but can lead to noticeable latency when quickly refocusing between objects placed near the eye as in our tabletop setup, where the farthest object was 0.5 m away.

It is also part of our future work to improve the experimental design and standardize a methodology to carry out AR Turing Tests. In our experiment, we used three pillars with plain textures. We aim to study more complex scenes that include objects of different shapes, materials with different surface texture parameters, and models of light scattering. With increasingly complex scenes, experimental measurements could collect ordinal instead of binary answers from participants in order to provide more faceted results. We believe that CG rendered under the pinhole camera model appeared too sharp during our user study and were as a result easily detected. In the future, we plan to compare the realism of CG rendered with a non-pinhole camera model versus EyeAR.

Our evaluation also showed that gender played a significant role in the detection of the virtual objects. It is also part of our future plans to study in more detail how the biological and physiological differences affect the perception of EyeAR.

Author Contributions: D.C.R. and C.S. conceived and designed the prototype described in this paper; D.C.R., A.R. and C.S. designed and performed the experiments; A.R. analyzed the data; D.C.R., A.R., A.P., and C.S. wrote

the paper; C.S., T.T. and G.Y. advised on the technical implementation and experiment design; H.K. assisted with the camera model formulation; S.I. provided the equipment and advised on the implementation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sutherland, I.E. The Ultimate Display. In Proceedings of the IFIP Congress, New York, NY, USA, 24–29 May 1965; pp. 506–508.
2. Wikipedia. Holodeck—Wikipedia, The Free Encyclopedia. Available online: http://en.wikipedia.org/wiki/Holodeck (accessed on 14 May 2017).
3. Sandor, C.; Fuchs, M.; Cassinelli, A.; Li, H.; Newcombe, R.A.; Yamamoto, G.; Feiner, S.K. Breaking the Barriers to True Augmented Reality. *arXiv* **2015**, arXiv:1512.05471.
4. Wikipedia. Google Glass—Wikipedia, The Free Encyclopedia. Available online: https://en.wikipedia.org/wiki/Google_Glass (accessed on 14 May 2017).
5. Wikipedia. Windows Mixed Reality—Wikipedia, The Free Encyclopedia. Available online: https://en.wikipedia.org/wiki/Windows_Mixed_Reality (accessed on 14 May 2017).
6. Kramida, G. Resolving the Vergence-Accommodation Conflict in Head-Mounted Displays. *IEEE Trans. Vis. Comput. Graph.* **2016**, *22*, 1912–1931.
7. Stredney, D.; Weghorst, S.J. *The Virtual Retinal Display: A New Technology for Virtual Reality and Augmented Vision in Medicine*; IOS Press and Ohmsha: Amsterdam, The Netherlands, 1998.
8. Tuceryan, M.; Genc, Y.; Navab, N. Single-Point Active Alignment Method (SPAAM) for Optical See-through HMD Calibration for Augmented Reality. *Presence* **2002**, *11*, 259–276.
9. Teittinen, M. Depth Cues in the Human Visual System. Available online: http://www.hitl.washington.edu/projects/knowledge_base/virtual-worlds/EVE/III.A.1.c.DepthCues.html (accessed on 14 May 2017).
10. Ware, C. *Information Visualization: Perception for Design*, 3rd ed.; Elsevier: Waltham, MA, USA, 2012.
11. Kán, P.; Kaufmann, H. Physically-Based Depth of Field in Augmented Reality. In Proceedings of the Eurographics (Short Papers), Cagliari, Italy, 13–18 May 2012; pp. 89–92.
12. Shan, Q.; Adams, R.; Curless, B.; Furukawa, Y.; Seitz, S.M. The Visual Turing Test for Scene Reconstruction. In Proceedings of the IEEE International Conference on 3D Vision-3DV, Seattle, WA, USA, 29 June–1 July 2013; pp. 25–32.
13. Borg, M.; Johansen, S.S.; Thomsen, D.L.; Kraus, M. Practical Implementation of a Graphics Turing Test. In *International Symposium on Advances in Visual Computing*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 305–313.
14. Rompapas, D.C.; Oshima, K.; Moser, K.R.; Swan, E.J., II; Ikeda, S.; Yamamoto, G.; Taketomi, T.; Sandor, C.; Kato, H. EyeAR: Physically-Based Depth of Field through Eye Measurements. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality Reality, Fukuoka, Japan, 29 September–3 October 2015.
15. Rompapas, D.C.; Rovira, A.; Ikeda, S.; Plopski, A.; Taketomi, T.; Sandor, C.; Kato, H. EyeAR: Refocusable Augmented Reality Content through Eye Measurements. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct) Reality, Merida, Mexico, 19–23 September 2016; pp. 334–335.
16. Rovira, A.; Rompapas, D.C.; Sandor, C.; Taketomi, T.; Kato, H.; Ikeda, S. EyeAR: Empiric Evaluation of a Refocusable Augmented Reality System. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), Merida, Mexico, 19–23 September 2016; pp. 11–12.
17. Hua, H. Enabling Focus Cues in Head-Mounted Displays. *Proc. IEEE* **2017**, *105*, 805–824.
18. Akeley, K.; Watt, S.J.; Girshick, A.R.; Banks, M.S. A Stereo Display Prototype with Multiple Focal Distances. *ACM Trans. Graph.* **2004**, *23*, 804–813.
19. Huang, F.-C.; Luebke, D.; Wetzstein, G. The Light Field Stereoscope: Immersive Computer Graphics via Factored Near-eye Light Field Displays with Focus Cues. *ACM Trans. Graph.* **2015**, *34*, 60:1–60:12.
20. Narain, R.; Albert, R.A.; Bulbul, A.; Ward, G.J.; Banks, M.S.; O'Brien, J.F. Optimal Presentation of Imagery with Focus Cues on Multi-plane Displays. *ACM Trans. Graph.* **2015**, *34*, doi:10.1145/2766909.
21. MacKenzie, K.J.; Hoffman, D.M.; Watt, S.J. Accommodation to Multiple-Focal-Plane Displays: Implications for Improving Stereoscopic Displays and for Accommodation Control. *J. Vis.* **2010**, *10*, 22.

22. Dunn, D.; Tippets, C.; Torell, K.; Kellnhofer, P.; Aksit, K.; Didyk, P.; Myszkowski, K.; Luebke, D.; Fuchs, H. Wide Field of View Varifocal Near-Eye Display Using See-Through Deformable Membrane Mirrors. *IEEE Trans. Vis. Comput. Graph.* **2017**, *23*, 1322–1331.

23. Liu, S.; Cheng, D.; Hua, H. An Optical See-Through Head Mounted Display with Addressable Focal Planes. In Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality, Cambridge, UK, 15–18 September 2008; pp. 33–42.

24. McQuaide, S.C.; Seibel, E.J.; Kelly, J.P.; Schowengerdt, B.T.; Furness, T.A. A Retinal Scanning Display System that Produces Multiple Focal Planes with a Deformable Membrane Mirror. *Displays* **2003**, *24*, 65–72.

25. Padmanaban, N.; Konrad, R.; Stramer, T.; Cooper, E.A.; Wetzstein, G. Optimizing Virtual Reality for all Users through Gaze-Contingent and Adaptive Focus Displays. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2183–2188.

26. Hillaire, S.; Lécuyer, A.; Cozot, R.; Casiez, G. Using an Eye-Tracking System to Improve Camera Motions and Depth-of-Field Blur Effects in Virtual Environments. In Proceedings of the IEEE Virtual Reality, Reno, NE, USA, 8–12 March 2008; pp. 47–50.

27. Hillaire, S.; Lécuyer, A.; Cozot, R.; Casiez, G. Depth-of-Field Blur Effects for First-Person Navigation in Virtual Environments. *IEEE Comput. Graph. Appl.* **2008**, *28*, 47–55.

28. Mauderer, M.; Conte, S.; Nacenta, M.A.; Vishwanath, D. Depth Perception with Gaze-contingent Depth of Field. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, ON, Canada, 26 April–1 May 2014; pp. 217–226.

29. Vinnikov, M.; Allison, R.S. Gaze-contingent Depth of Field in Realistic Scenes: The User Experience. In Proceedings of the ACM Symposium on Eye Tracking Research and Applications, Safety Harbor, FL, USA, 26–28 March 2014; pp. 119–126.

30. Hua, H.; Liu, S. Depth-Fused Multi-Focal Plane Displays Enable Accurate Depth Perception. *Proc. SPIE Opt. Des. Test. IV* **2010**, doi:10.1117/12.871245.

31. Lin, X.; Takefumi, O. Blur with Depth: A Depth Cue Method Based on Blur Effect in Augmented Reality. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, Adelaide, Australia, 1–4 October 2013; pp. 1–6.

32. Turing, A.M. Computing Machinery and Intelligence. *Mind* **1950**, *59*, 433–460.

33. McGuigan, M. Graphics Turing Test. *arXiv* **2006**, arXiv:cs/0603132.

34. Meyer, G.W.; Rushmeier, H.E.; Cohen, M.F.; Greenberg, D.P.; Torrance, K.E. An Experimental Evaluation of Computer Graphics Imagery. *ACM Trans. Graph.* **1986**, *5*, 30–50.

35. Nvidia. How We Created Demos that Blur the Line Between Real and Rendered. Available online: http://blogs.nvidia.com/blog/2015/08/11/photorealistic/ (accessed on 14 May 2017).

36. Rosales, P.; Dubbelman, M.; Marcos, S.; van der Heijde, R. Crystalline Lens Radii of Curvature from Purkinje and Scheimpflug Imaging. *J. Vis.* **2006**, *6*, 1057–1067.

37. Lee, J.W.; Cho, C.W.; Shin, K.Y.; Lee, E.C.; Park, K.R. 3D Gaze Tracking Method Using Purkinje Images on Eye Optical Model and Pupil. *Opt. Lasers Eng.* **2012**, *50*, 736–751.

38. Toyama, T.; Orlosky, J.; Sonntag, D.; Kiyokawa, K. A Natural Interface for Multi-focal Plane Head Mounted Displays using 3D Gaze. In Proceedings of the International Working Conference on Advanced Visual Interfaces, Como, Italy, 27–29 May 2014; pp. 25–32.

39. Oshima, K.; Moser, K.R.; Rompapas, D.C.; Swan, E.J., II; Ikeda, S.; Yamamoto, G.; Taketomi, T.; Sandor, C.; Kato, H. Improved Clarity of Defocussed Content on Optical See-Through Head-Mounted Displays. In Proceedings of the IEEE Symposium on 3D User Interfaces, Greenville, SC, USA, 19–23 March 2016.

40. Cook, R.L.; Porter, T.; Carpenter, L. Distributed Ray Tracing. *Proc. ACM Siggraph Comput. Graph.* **1984**, *18*, 137–145.

41. Green, D.G.; Powers, M.K.; Banks, M.S. Depth of Focus, Eye Size and Visual Acuity. *Vis. Res.* **1980**, *20*, 827–835.

42. Shiomi, T.; Uemoto, K.; Kojima, T.; Sano, S.; Ishio, H.; Takada, H.; Omori, M.; Watanabe, T.; Miyao, M. Simultaneous Measurement of Lens Accommodation and Convergence in Natural and Artificial 3D Vision. *J. Soc. Inf. Disp.* **2013**, *21*, 120–128.

43. Parker, S.G.; Bigler, J.; Dietrich, A.; Friedrich, H.; Hoberock, J.; Luebke, D.; McAllister, D.; McGuire, M.; Morley, K.; Robison, A.; et al. Optix: A General Purpose Ray Tracing Engine. *ACM Trans. Graph.* **2010**, *29*, 66.

44. Nicodemus, F.E. Directional Reflectance and Emissivity of an Opaque Surface. *Appl. Opt.* **1965**, *4*, 767–775.

45. Liang, K.-Y.; Zeger, S.L. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika* **1986**, *73*, 13–22.

46. Agresti, A. *Categorical Data Analysis*; Springer: Berlin/Heidelberg, Germany, 2011; p. 710.

47. Ma, Y.; Mazumdar, M.; Memtsoudis, S.G. Beyond Repeated Measures ANOVA: Advanced Statistical Methods for the Analysis of Longitudinal Data in Anesthesia Research. *Reg. Anesth. Pain Med.* **2012**, *37*, 99–105.

48. Pan, W. Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics* **2001**, *57*, 120–125.

49. Hin, L.-Y.; Wang, Y.-G. Working-Correlation-Structure Identification in Generalized Estimating Equations. *Stat. Med.* **2009**, *28*, 642–658.

50. Evans, S.; Li, L. A Comparison of Goodness of Fit Tests for the Logistic GEE Model. *Stat. Med.* **2005**, *24*, 1245–1261.

51. Lockhart, T.E.; Shi, W. Effects of Age on Dynamic Accommodation. *Ergonomics* **2010**, *53*, 892–903.

52. Grosse, M.; Wetzstein, G.; Grundhöfer, A.; Bimber, O. Coded aperture projection. *ACM Trans. Graph.* **2010**, *29*, 1–12.

53. Iwai, D.; Mihara, S.; Sato, K. Extended Depth-of-Field Projector by Fast Focal Sweep Projection. *IEEE Trans. Vis. Comput. Graph.* **2015**, *21*, 462–470.