Optimizing Depth Perception in Virtual and Augmented Reality through Gaze-contingent Stereo Rendering

BROOKE KRAJANCICH, Stanford University
PETR KELLNHOFER, Stanford University and Raxium
GORDON WETZSTEIN, Stanford University

Fig. 1. Ocular motion associated with changes of fixation alters the positions of the no-parallax points in both eyes. Rendering models that do not account for ocular motion can create distortions of binocular disparity, as seen in this example. The color-coded error maps illustrate the magnitude of this effect as the difference between angular disparities resulting from classical and our gaze-contingent stereoscopic rendering for two different fixation points. Both shortening (red) and stretching (blue) of disparity gradients can be observed.

Virtual and augmented reality (VR/AR) displays crucially rely on stereoscopic rendering to enable perceptually realistic user experiences. Yet, existing near-eye display systems ignore the gaze-dependent shift of the no-parallax point in the human eye. Here, we introduce a gaze-contingent stereo rendering technique that models this effect and conduct several user studies to validate its effectiveness. Our findings include experimental validation of the location of the no-parallax point, which we then use to demonstrate significant improvements of disparity and shape distortion in a VR setting, and consistent alignment of physical and digitally rendered objects across depths in optical see-through AR. Our work shows that gaze-contingent stereo rendering improves perceptual realism and depth perception of emerging wearable computing systems.

CCS Concepts: • **Hardware → Displays and imagers**; • **Computing methodologies → Computer graphics**; **Mixed / augmented reality**.

Additional Key Words and Phrases: applied perception, rendering, virtual reality, augmented reality

ACM Reference Format:
Brooke Krajancich, Petr Kellnhofer, and Gordon Wetzstein. 2020. Optimizing Depth Perception in Virtual and Augmented Reality through Gaze-contingent Stereo Rendering. *ACM Trans. Graph.* 39, 6, Article 270 (December 2020), 10 pages. https://doi.org/10.1145/3414685.3417820

Authors' addresses: Brooke Krajancich, Stanford University, brookek@stanford.edu; Petr Kellnhofer, Stanford University and Raxium, pkellnhofer@stanford.edu; Gordon Wetzstein, Stanford University, gordon.wetzstein@stanford.edu.

1 INTRODUCTION

Virtual and augmented reality (VR/AR) promise a new medium that offers user experiences with unprecedented levels of immersion and the capability to seamlessly blend digital and physical spaces. Accurate stereoscopic rendering is one of the key requirements for perceptual realism and for being able to place and anchor digital objects in physical environments with optical see-through AR displays. Here, even small amounts of disparity distortion, for example caused by inaccurate stereo rendering, negatively affect the experience and would destroy the seamless blending of virtual and real content.

Current stereo rendering algorithms in VR/AR fall short of accurately modeling the human visual system. An approximation made by almost all existing systems is that the no-parallax point, or center of projection, of the human eye coincides with the center of rotation. Recent work suggests that taking the distance between the centers of rotation and projection of the human eye into account can help improve ordinal depth perception [Konrad et al. 2020]. This is achieved with ocular parallax rendering, i.e. by accounting for the depth-dependent image shifts that occur with ocular motion. In some situations, parallax can convey absolute depth information [Burge et al. 2010], however it is widely known to provide a reliable ordinal (i.e., relative) depth cue, [Kellnhofer et al. 2016b; Yonas et al. 1987]. Kudo et al. [1999] and Konrad et al. [2020] both verified that this is also the case for ocular parallax in monocular viewing conditions but they were not able to show any measurable effects on absolute depth perception in stereoscopic viewing conditions.

Here, we study the effect of ocular motion, and vergence in particular, on disparity distortion in stereoscopic viewing conditions and propose and evaluate a new gaze-contingent stereo rendering approach for VR/AR. The eye model we employ is slightly different

from Konrad's [2020] in that it not only takes the distance between the centers of rotation and projection of the eye into account but also the non-negligible offset between the optical and visual axis. Moreover, we design and conduct a number of user experiments that allow us to make important insights and improvements to existing stereo rendering techniques. First, we experimentally determine the location of the no-parallax point with a small group of subjects and verify that this is well approximated by recently employed model eyes. Second, we experimentally demonstrate that our approach significantly improves disparity distortion in VR settings and perceived alignment of digital and physical objects in AR. With our gaze-contingent stereo rendering technique, we take an important step towards enabling perceptually realistic and seamless experiences with emerging wearable display systems.

Specifically, we make the following contributions:

- We introduce a gaze-contingent stereo rendering algorithm that includes a more accurate model eye than recent work.
- We design and conduct user experiments that demonstrate significant improvements of disparity distortion and enhanced depth perception in VR.
- We design and conduct user experiments that demonstrate significant improvements in perceptual realism and alignment of digital and physical objects with optical see-through AR.

2 RELATED WORK

2.1 Gaze-contingent Rendering and Display

The gaze-contingent rendering and display paradigm has enabled a variety of important computer graphics techniques that adapt to the user's gaze direction [Duchowski et al. 2004; Koulieris et al. 2019]. These include gaze-contingent geometric level-of-detail rendering [Luebke and Hallen 2001; Murphy and Duchowski 2001; Ohshima et al. 1996], foveated rendering [Geisler and Perry 1998; Guenter et al. 2012; Patney et al. 2016; Sun et al. 2017] and display [Kim et al. 2019], tone mapping [Jacobs et al. 2015; Mantiuk and Markowski 2013; Mauderer et al. 2016], sparse-sample image reconstruction [Kaplanyan et al. 2019], depth-of-field rendering [Duchowski et al. 2014; Hillaire et al. 2008; Mantiuk et al. 2011; Mauderer et al. 2014; Vinnikov and Allison 2014] as well as varifocal [Akşit et al. 2017; Dunn et al. 2017; Johnson et al. 2016; Konrad et al. 2015; Liu et al. 2008; Padmanaban et al. 2017] and multifocal [Akeley et al. 2004; Chang et al. 2018; Mercier et al. 2017; Rolland et al. 2000] display techniques.

Seemingly related to our gaze-contingent stereo rendering approach are gaze-contingent disparity manipulation techniques [Hanhart and Ebrahimi 2014; Kellnhofer et al. 2016a; Peli et al. 2001]. These remap the disparities of a stereo image to fit into the zone of comfort of a 3D display [Shibata et al. 2011] to mitigate the vergence–accommodation conflict. This amounts to shifting the 3D scene forward or back such that the fixated object appears well within the zone of comfort, close to the physical screen. Our approach does not manipulate the disparities of another stereo rendering algorithm, but aims at rendering more perceptually accurate disparities in the first place. Disparity remapping could optionally be applied as a post-processing method, although we do not explore this.

The work closest to ours is the recent proposal on ocular parallax rendering by Konrad et al. [2020]. Their algorithms also take the non-negligible distance between centers of rotation and projection of the human eye into account. With this, they were able to demonstrate improvements in ordinal depth perception in monocular viewing conditions, but their experiments showed no effect on binocular vision, which was only explored with a crude blind reaching task. Unlike their model eye, we take the offset between the visual and optical axes into account and focus exclusively on disparity distortion and stereo rendering in binocular viewing conditions. With a better model eye and refined user experiments, we are the first to show significant effects on absolute depth perception and also digital–physical object alignment in VR/AR applications.

2.2 Eye-aware HMD Calibration and Pupil Steering

It is well known that precise calibration of the user's eyes with respect to a head-mounted display (HMD) is important. For example, optical see-through AR systems require knowledge of the eye's location for displaying digital imagery that is seamlessly aligned with the physical content. Otherwise disparity distortion leads to perceptual conflicts with other depth cues as well as with disparity observed in the physical world. To this end, both one-time interactive calibration procedures [Janin et al. 1993; Tuceryan and Navab 2000] and online, computer vision–based techniques [Itoh et al. 2016; Itoh and Klinker 2014; Plopski et al. 2015] have been proposed. Unlike our approach, these calibration techniques assume that the centers of rotation and projection of the eye are the same. Therefore, they do not account for the gaze-contingent disparity changes we study in this work. Kudo and Ohnishi [2000] discuss gaze-contingent optical distortions in HMDs and attribute them in part to ocular parallax. However, they did not propose an improved rendering approach or study its perceptual implications for HMDs.

Some types of near-eye displays, most notably light-field [Hua and Javidi 2014; Huang et al. 2015; Lanman and Luebke 2013] and holographic [Maimone et al. 2017; Padmanaban et al. 2019; Peng et al. 2020; Shi et al. 2017] HMDs, can in theory provide natural parallax within their respective eye box volume. However, this effect has not been studied in the literature and the eye boxes of current holographic displays are too small to support a significant rotational range of the eye. This not only prevents realistic parallax, but can also easily destroy the visual percept itself. One approach to exit pupil expansion is pupil steering [Jang et al. 2018, 2017; Park and Kim 2018] where a user's eyes are tracked and the small exit pupils of the displays are optically steered towards the user's pupils. Although these approaches require eye tracking, we are not aware of prior work in this area that has adjusted the stereo rendering accordingly or analyzed the resulting disparity distortions.

3 EXPERIMENTAL LOCALIZATION OF THE NO-PARALLAX POINT

The no-parallax point of an optical system, such as the human eye, represents the location around which the entire system can be rotated without observing parallax. While exact eye anatomy varies from person to person, several cardinal points and axes are used to describe optical properties common across the population. For the

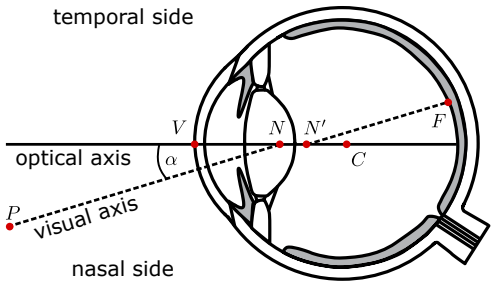

Fig. 2. Illustration of the optical and visual axes and relevant points in the right eye (top view). The optical axis connects the anterior vertex of the cornea (V) and the center of rotation (C). The visual axis connects the point of fixation (P) with the front nodal point (N), which extends through the rear nodal point (N'), to intersect with the fovea (F). The angle α offsets the visual axis on average by 5° in the nasal and 3° in the inferior directions.

purpose of describing our gaze-contingent rendering approach, we refer to six of these cardinal points and two axes, however additional definitions are included in the Supplement. The first axis is referred to as the optical axis, and is anatomically defined by passing through the anterior vertex of the cornea (V) and the center of rotation (C) whereas the visual axis registers the fixated object (P) with the fovea (F) while passing through front (N) and rear (N') nodal points [Atchison and Smith 2000] (Fig. 2). From *ex vivo* examinations, the front nodal point has been estimated to lie 7–8 mm in front of the center of rotation [Atchison 2017]. Note that the location of this point changes with the accommodation state but such variance is relatively minor (less than 0.5 mm [Atchison 2017]), which is particularly true for current fixed-focus VR/AR systems. While it has been postulated that the front nodal point is in fact the no-parallax point of the eye, to our knowledge the only study trying to verify this could not give a confirmation as the larger measured distance hinted at the no-parallax point being located even further forward [Bingham 1993].

In this work, we conduct our own psychophysical study to experimentally determine the position of the no-parallax point for several users. For this purpose, we adapt the general setup proposed by Bingham [1993]. As shown in Figure 3 (a), two surfaces are separated in depth and aligned such that when the users fixate at the rear one, they cannot see its red half. However, with a fixed head and instructed to fixate at the gaze target with an angular displacement θ, the rotation of the eye shifts the no-parallax point of the eye towards the left. This reveals an extent E of the rear surface and the red half becomes visible in the periphery. Note that this is only possible if the center of perspective of the eye, and equivalently the no-parallax point, is located in front of the center of rotation.

The distance between no-parallax point and center of rotation, NC, can then be calculated using the target distances L_1 and L_2:

$$NC = \frac{EL_1}{(L_2 - L_1)\sin\theta + E\cos\theta} \tag{1}$$

A more detailed derivation of this equation is included in the Supplement. We construct this setup with $L_1 = 0.5\,\text{m}$, $L_2 = 1\,\text{m}$ and $\theta = 30°$ in a controlled experiment to determine the largest extent, E, a user can detect. This is equivalent to determining the number of pixels the red region can be shifted towards the right before the user can no longer detect it. The configuration also means that scattering within the ocular media should not increase a user's ability to detect the red region due to the light not being able to enter the eye at or beyond this threshold. After converting pixels to meters, Equation 1 can be used to calculate NC for the user.

Stimuli. We use a 6" Topfoison liquid crystal display (LCD) with a resolution of 1920 × 1080 and an edge-lit light-emitting diode backlight as the far target, as shown in Figure 3 (b). This enables us to easily control and change the displayed extent using an attached laptop, to a precision of 0.069 mm, the pixel pitch of the display. Without eye rotation, the front half surface would completely occlude the red stimulus. Each trial contained both an extent stimulus with red and white regions and a control stimulus, where the full LCD displayed only white. The brightness was defined in the RGB space of the display. The white pixels were rendered at 80% brightness across red, green and blue channels to reduce the apparent brightness to approximately that of the red extent with 100% red alone.

Conditions. All stimuli were presented monocularly to the right eye while the user wears an eyepatch on the left eye. With the user accommodated at the L_2 distance, the eye's limited depth of field can cause the edge of the half surface to appear blurred. As Bingham et al. [1993] described this as a confounding factor, we try to mitigate it using a lamp to illuminate the gaze target to stop down the pupil aperture and maximize the depth of field.

Subjects. Eight adults participated (age range 21–29, 3 female). Due to the demanding nature of our psychophysical experiment, only a few subjects were recruited, which is common for low-level psychophysics (see e.g. [Patney et al. 2016]). All subjects in this and all following experiments had normal or corrected to normal vision, no history of visual deficiency, and no color blindness. All subjects gave informed consent. The research protocol was approved by the Institutional Review Board at the host institution.

Procedure. To start the session, each subject was instructed to use the left and right arrow keys to shift the red portion on the screen such that they could just see it when looking down the center of the targets. After subtracting one pixel, this was used as $E = 0$ for all subsequent trials.

Each trial constituted a two-alternative forced choice (2AFC) test, and subjects were asked to use the keyboard to choose which of the two displays contained the red extent. The keyboard could also be used to toggle between the two displays as the users desired. Most users did so less than 10 times. However once a selection was made, this concluded the trial. No feedback was provided. Subjects were instructed to fixate only as far as the cross target, but were free to look back to the far display if desired.

Subjects completed 60 trials, consisting of 12 displayed extent, E, configurations, each tested 5 times. The experiment took about 20 minutes per subject to complete, including instruction and centerline calibration.

For the first block of 30 trials, E for a given trial was randomly chosen from 6 evenly spaced values between 20 and 80 pixels (1.38

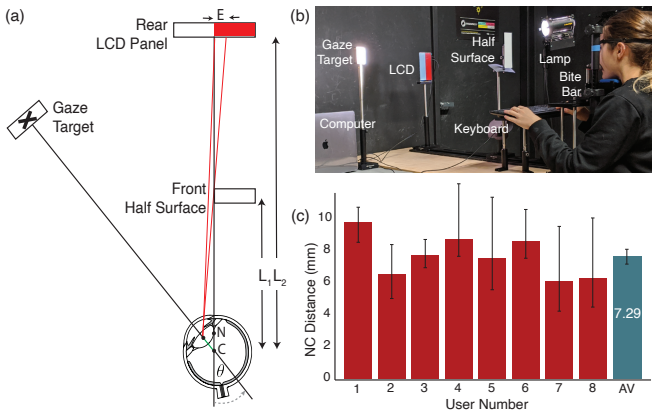

Fig. 3. Psychophysical experiment to measure the position of the no-parallax point. (a) A diagram of the experimental setup. A half surface and an LCD panel are set up at distances L_1 and L_2, respectively. As the eye rotates counterclockwise about its center by an angle, θ, the no-parallax point is translated to the left revealing an extent, E, along the rear surface. The distance NC, shown in green, corresponds to the largest distance the red extent can be shifted towards the right before the user can no longer identify it. (b) Photograph of a user conducting the experiment. The head is kept stationary with a headrest and bite bar. The computer is used to change the stimulus on the LCD panel and record the user's response given by the keyboard. The lamp is used to illuminate the gaze target to reduce depth of field blurring of the edge of the half surface. (c) Results of psychophysical user experiment. The NC distances measured for each of the 8 study participants are shown with 95% confidence interval represented as an error bar. The final column (AV) represents the mean of all participants, 7.29 mm, with error bar showing the standard deviation of 1.25 mm.

and 5.52 mm), covering the range of expected values. For the second block of 30, E for a given trial was randomly chosen from 6 evenly spaced values between the values of E from the previous trial block where the user was getting less than 90% and more than 60% correct. This paradigm was chosen to maximize sampling around the threshold value, without causing observable fatigue.

Analysis. For each E displayed, we compute the proportion of correct responses. Using the *psignifit* Python package [Schütt et al. 2016], we fit a psychometric function to each subject's data using Bayesian inference. Each psychometric function gives us a detection threshold, measured as pixel shifts from the initial set position. The thresholds represent where the psychometric function exceeded a 75% chance for a correct response. Individual psychometric fits are included in the Supplement. This is converted to meters using the pixel pitch of the display (0.069 mm) and NC is then calculated using Equation 1.

Results. The results of this experiment are shown in Figure 3 (c), giving a mean of 7.29 mm for NC, which is within the originally expected range of values (7–8 mm). Surprisingly, we observe a variation of about 3.54 mm among our subjects, indicating that there may be value in measuring and accounting for individual variation. However, the difficulty of measuring a person's NC distance makes such an approach impractical at the moment, so we continue to model an "average observer" as having $NC = 7.29$ mm for the remainder of this paper.

4 STEREOSCOPIC RENDERING WITH THE NO-PARALLAX POINT

In this section, we study the effects of the no-parallax point on binocular vision and the horopter, leading to the prediction of a surprisingly high degree of disparity distortions with conventional stereo displays. We outline a gaze-contingent (GC) stereo rendering pipeline that takes the no-parallax point into account for precise disparity rendering.

4.1 Binocular Vision and the Horopter

The binocular horopter refers to the set of points in space that give rise to the same disparity on the retina [Vojniković and Tamajo 2013]. Thus, the horopter provides a useful tool for analyzing and comparing different models for binocular vision. It is geometrically modeled as an arc on a Vieth-Müller circle formed by the two no-parallax points and the fixation point [Vojniković and Tamajo 2013]. The choice of no-parallax point determines the specific shape of the horopter [Turski 2016]. In Figure 4 we show geometrical horopters for the no-parallax point in the center of rotation, as commonly used in computer graphics (black), and in the nodal point, as used here. The shape is further differentiated by the choice of gaze vector where the angular offset of the visual axis used in our model (green) yields a different horopter than the approximation using the optical axis as used by Konrad et al. [2020] (blue). In the following, we outline an adequate rendering pipeline for our model and then analyze the expected disparity distortions when using other models.

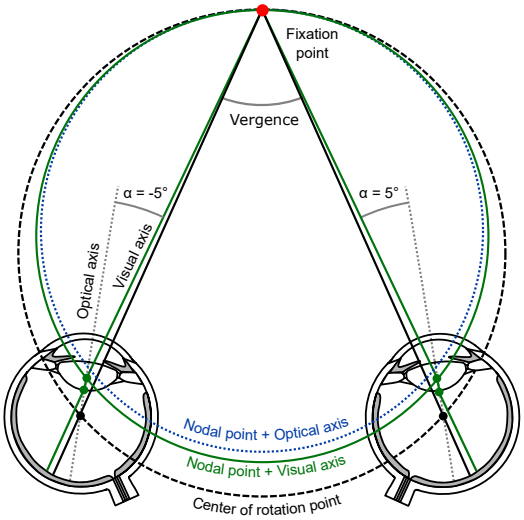

Fig. 4. The horopters predicted for various eye models: the center of rotation model (black dashed), the front nodal point with the gaze vector being the optical axis (blue dotted) and the visual axis (green solid). Note that fixation leads to a different eye rotation for each axis (only visual axis-aligned eyes are shown here). The angle α is exaggerated for clarity.

4.2 Stereoscopic rendering

Traditional stereoscopic rendering models the projection of 3D points into eye coordinates using a matrix–vector multiplication of the matrix $\mathbf{P}_{L/R} \cdot \mathbf{E}_{L/R} \cdot \mathbf{V} \cdot \mathbf{M}$ with a vertex specified in object coordinates. Here, $\mathbf{M}$ is the model matrix, $\mathbf{V}$ is the view matrix, $\mathbf{E}_{L/R}$ is the eye matrix and $\mathbf{P}_{L/R}$ the projection matrix for left and right eye, respectively. Accounting for the no-parallax point requires changes to the eye and projection matrices, which we describe in the following.

Eye matrix. Assume that the centers of rotation of the eyes are $\mathbf{C}_{L/R} = (\mp\frac{\text{IPD}}{2}, 0, 0)$, where IPD is the inter-pupillary distance (IPD). Conventionally, the eye matrices are defined as translations into the eye centers, i.e. $\mathbf{E}_{L/R} = \mathbf{T}(-\mathbf{C}_{L/R})$, using the translation matrix $\mathbf{T}$. To account for the distance between centers of projection and rotation, NC, we calculate the gaze-dependent location of the no-parallax point with respect to $\mathbf{C}_{L/R}$

$$\text{N}_{L/R} = \mathbf{R}\left(-\theta^{(v)}_{L/R}, -\theta^{(h)}_{L/R}, 0\right) \cdot \mathbf{R}\left(-\boldsymbol{\alpha}_{L/R}\right) \cdot \begin{pmatrix} 0 \\ 0 \\ -\text{NC} \end{pmatrix} \tag{2}$$

Here, $\boldsymbol{\alpha}_{L/R} = (-3°, \mp 5°, 0)$ is the offset, in eccentricity angle, between optical and visual axis for the two eyes [Atchison 2017], $\theta^{(h,v)}$ represent the horizontal and vertical gaze angle for each eye, $\mathbf{R}$ is a 3×3 rotation matrix using Euler angles, and we use NC = 7.29 mm from our earlier experiment. The eye matrices then become $\mathbf{E}_{L/R} = \mathbf{T}(-\text{N}_{L/R}) \cdot \mathbf{T}(-\mathbf{C}_{L/R})$. Note that this notation uses a right-handed coordinate system, such as that used by OpenGL.

Projection matrix. We use standard asymmetric off-axis perspective projection matrix $\text{P}_{L/R}$ defined for a magnified virtual image of the microdisplay at distance d [Konrad et al. 2020]. Note that the projection matrix depends on the gaze-dependent position of the no-parallax point (Eq. 2). As illustrated in Figure 5, for 3D points located at distance d no ocular parallax is observed, i.e., this is the zero-parallax plane. Setting the parameter d to match the virtual image distance of an HMD is critical for correct reproduction of disparity with gaze-contingent stereo rendering.

4.3 Disparity distortion

We model the image formed on the retina using the considered binocular projection model to quantify the magnitude of the expected disparity distortion. In Figure 6, we present the differences of the vergence angles predicted for the same fixation points. We further use the disparity perception model by Didyk et al. [2011] to compute just-noticeable differences (JND). Values above 1 JND predict visibility of the predicted distortions. Note, that verging at each fixation point requires eye rotation to achieve required gaze eccentricity. We mark the normal range of horizontal eye rotation (±45° [Shin et al. 2016]) by the red lines. Full binocular field of view (FOV) of 60° is then accessible with additional retinal eccentricities within the Panum fusion area [Atchison and Smith 2000].

The left panel compares our full model and the model based on the center of rotation. We observe that even for fixations as far as 2.5 m the difference of vergence angles yields visible disparity differences. This difference further grows with decreasing distance.

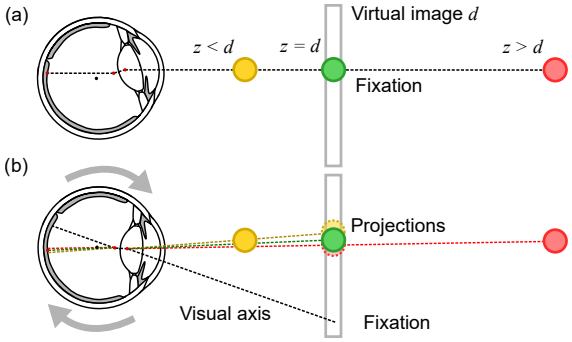

Fig. 5. Ocular parallax for a virtual image at distance d and objects at distances $z < d$ (yellow), $z = d$ (green) and $z > d$ (red). (a) All points project to the same screen coordinates and a single spot on the fovea when the green point is fixated. (b) After a saccade the display projection of the near point moves up, the projection of the far point moves down and the projection of the middle point remains the same. The retinal image changes accordingly.

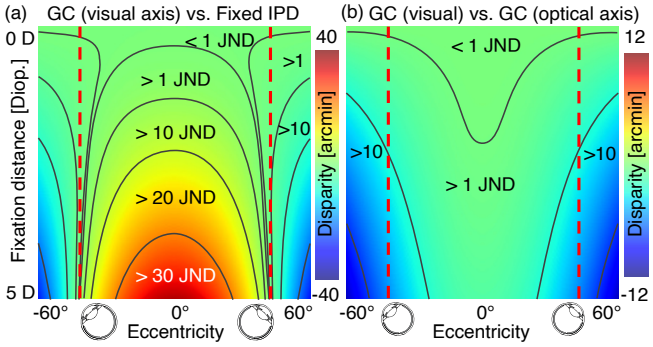

Fig. 6. The difference of eye vergence angles predicted for different fixation points by the gaze-contingent (GC) model with visual axis and either the standard model using center of rotation (a) or the GC model with optical axis (b). The fixations are expressed relative to the midpoint between both no-parallax points in relaxed state. The initial IPD = 64 mm was defined for the eyes looking straight ahead. The isolines mark levels of stereoacuity JNDs from Didyk et al.'s [2011] model for the optimal spatial frequency of 0.4 cpd. Values of JND of 1 and larger predict visibility of a difference in a direct comparison for an average observer. The range of eccentricities covers the full binocular FOV achieved through combination of gaze and retinal eccentricities [Atchison and Smith 2000]. Note, that normal range of eye rotation needed for gaze fixation is delimited by the red bars [Shin et al. 2016].

Additionally, the horizontal eccentricity affects the shape of the distortion field. While the model predicts larger perceived distances in our model for the central visual field, this trend is reversed for eccentricities above $\approx 40°$.

This effect can be practically demonstrated in augmented reality where real objects are superimposed with virtual objects. In Section 6, we show that traditional stereo rendering causes visible misalignment of physical and digitally rendered objects and we

demonstrate how gaze-contingent stereo rendering reduces this issue significantly.

The right panel of Figure 6 further compares our model with a variant that assumes the optical axis, instead of the visual axis, to be a good approximation of the gaze direction (Fig. 4). While the differences for the central visual field are relatively small, the role of the axis is notable for larger viewing angles.

Finally, we also analyze the effect in terms of no-parallax point separation and its change with respect to the initial IPD. For an average IPD of 64 mm and a fixation distance of 30 cm we observe an effective decrease of viewpoint separation to as low as 62.5 mm. That corresponds to a shift from 68 to 52-percentile of the IPD distribution in the female population [Dodgson 2004] which is a deviation that typically requires user adjustment in common VR systems. To illustrate this phenomenon, we explore the produced shape distortions in a VR scenario in Section 5.

4.4 Verifying the Model

Figure 6 predicts the visibility of the difference between both models for an average observer and a theoretical display with $d \rightarrow \infty$. To validate our model we test this hypothesis with a user study using a VR platform. Using the same equipment as in Section 5, we conduct a detection study using a random dot stereogram (RDS) stimulus at varying depths. The users are randomly presented with two RDS stimuli, rendered at the same depth, but placed on top of each other (i.e., vertically) in the center of the visual field (Fig. 7 (a)). Each stimulus is a square 10° in diameter. In random order, one stimulus is rendered with and one without gaze-contingent stereo rendering. As a result, one of the stimuli has a different disparity than the background, which is itself an RDS stimuli rendered at the same depth with our gaze-contingent mode (Fig. 7 (b)). Users are tasked with a 2AFC and use a keyboard to report whether the upper or lower segment contains the patch that protrudes from the background. All stimuli were rendered at a distance of 1, 1.33, 1.5, 1.75, 2, 2.5 or 3 D (inverse meters), with 6 trials at each distance in a randomly shuffled order. A black screen was shown for 3 seconds between trials to assist in eye adjustment. For each of the 7 distance configurations, we computed the proportion of correct responses. Using Bayesian inference methods [Schütt et al. 2016; Wichmann and Hill 2001], we fit a psychometric function to each subject's responses, finding the fixation distance with 75% detection threshold (Fig. 7 (e)). 11 subjects (6 male, 5 female, aged 18–54) took part in the study.

Using a digital single-lens reflex (DSLR) camera, we measured the display distance d of our HTC Vive Pro to be $\approx$ 70 cm. This changes the distribution of depth distortion in Figure 6 such that the 1 JND occurs at a distance of 66 cm (see Fig. 7 (c)). We found that the depth distortion was detectable, on average, at a distance of 62.8 $\pm$ 1.3 cm or 1.59 $\pm$ 0.033 D (Standard Error, SE). This confirms the importance of taking the no-parallax point into account for accurate stereoscopic rendering. While we chose not to additionally burden users with measuring their stereoacuity, the similarity of the mean measured detection distance to the model-predicted expected distance of 1.52 D confirms our model's ability to predict observable disparity distortions of different rendering models (Fig. 7 (d)).

5 DEPTH DISTORTION IN VR

The analysis and experiments in Section 4.3 predict visibility of disparity distortion for rendering that ignores the gaze-contingent shift of the no-parallax point. Here, we explore this issue further and experimentally test a hypothesis that a shape of a 3D object rendered using the traditional stereoscopic rendering will appear distorted as a function of fixation distance. Further, we validate that our gaze-contingent rendering reduces this distortion significantly.

Hardware and Software. We used an HTC Vive Pro VR system, which has a diagonal field-of-view of 145°, a refresh rate of 90 Hz and a 1440×1600 pixel organic light-emitting diode display per eye, resulting in a theoretical central resolution of 4.58 arcmin/pixel. The HTC Vive Pro supports built-in IPD adjustment. Unity was used as the rendering engine for all rendering modes and user experiments.

Stimuli and Conditions. For this experiment, we require a stimulus whose apparent shape does not rely on metric structure, but only on ratios of its dimensions. For this purpose, we emulated the triangle wave experiment performed by Glennester et al. [1996] for measuring stereoscopic depth constancy. As illustrated by the schematic in Figure 8 (a), this stimulus is a triangle wave formed by an RDS pattern. It is rendered such that the amplitude is half the period of the peaks. Thus, if the depth rendering is physiologically correct, the dihedral angle of the peaks and troughs should be at 90° (shown in dark green). However, if the depth space is perceived as stretched, as we predict is the case without gaze-contingent rendering (see Sec. 4), then the angles should appear smaller than 90°. Similarly, a perceived compression of the depth space would increase the angles. A scaled crop of an illustrative anaglyph reproduction can be seen in Figure 8 (b).

During the user study two identical patterns were shown horizontally side-by-side, one with and one without GC rendering, at a depth of either 0.3, 0.5 or 0.7 m. These depths were chosen such as to increment to the measured display distance of the HTC Vive Pro (0.7 m), where no disparity distortion should be observed. We refer to the rendering without GC as fine-tuned (FT) rendering, since we first set the subject's IPD using the physical knob provided on the device, and then allow the user to further tune the horizontal separation of both virtual images in projection matrices. This was done by rendering a single pattern at a fixed far distance and instructing the users to tweak the separation until the stimulus exhibited 90° angles. The distance of 2 m was chosen as a compromise, where the effect of GC rendering diminishes yet binocular disparity is still a relevant depth cue. The GC mode was identical to the FT mode, but with the modifications described in Section 4.2. Finally, to ensure fair comparison of these rendering modes, we shift the center of projection in the FT mode from the center of rotation to the position of the GC no-parallax point of a user looking towards optical infinity. This ensures that the only difference between the two modes comes from the gaze-contingent movement of the no-parallax point and not from an arbitrary initial position bias.

Procedure. Before starting trials, each of the nine adult subjects (age range 18–54, 4 female) completed the calibration procedure described above to set-up the fine-tuned rendering mode. Each trial then constituted 2AFC, where one of the three tested depths was

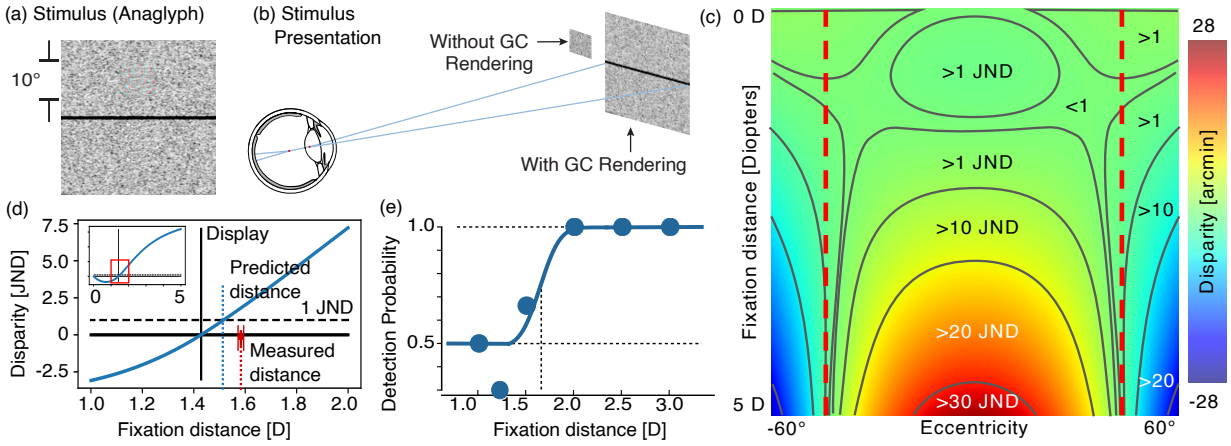

Fig. 7. Verifying the detection threshold for gaze-contingent rendering. (a) The stereoscopic stimulus visualized in anaglyph. (b) Conceptual side view (schematic) of the stimulus presentation. The stimulus rendered without gaze-contingent (GC) rendering appears to pop out from the background, unlike the other stimulus (unseen under the black line), which are both rendered with GC rendering. (c) The predicted disparity differences between models with and without ocular parallax for a VR display with a display distance $d = 0.7$ m. The red bars delimit normal range of horizontal eye rotation which restricts the range of gaze fixation eccentricities [Shin et al. 2016]. (d) The JNDs for different fixation distances of the central vision around the display (the inset shows larger distance range). The red interval marks the detection threshold and SE interval measured in our experiment. (e) An example of psychometric function fit for one user.

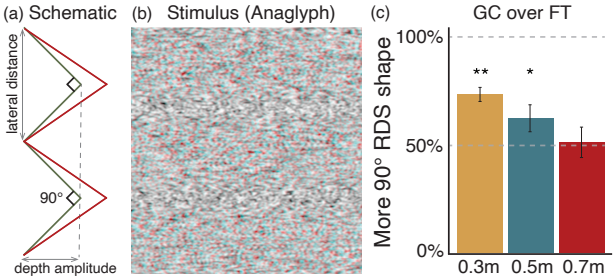

Fig. 8. Evaluating shape distortion of virtual content. Subjects simultaneously viewed two identical triangle wave random dot stereogram (RDS) stimuli, one rendered with fine-tuned (FT) IPD and the other with gaze-contingent (GC) rendering. (a) A schematic of a cross-section of the stimulus. Designed to evaluate shape distortion caused by incorrect depth scaling, the dimensions of the RDS triangles are calculated such that the amplitude of the peaks (in depth) is twice the lateral distance (period of the pattern). If the depth space is correct, the dihedral angle of the peaks should be at 90° (green), but if the depth space is stretched (as it is without gaze-contingent rendering), the angles should appear smaller (red). (b) An illustrative anaglyph rendering of the stimulus (not to scale). Both stimuli were rendered at a target depth of either 0.3, 0.5 or 0.7 m and we asked subjects to indicate which of the two contained angles is closer to 90°. (c) The percentage of times that the gaze-contingent mode was chosen as more accurate per distance. Despite the seemingly small effect size, shape distortion is detectable, in particular for closer distances. Error bars represent Standard Error (SE) and significance is indicated at the $p < 0.05$ and 0.001 levels with * and ** respectively.

randomly chosen for rendering, and subjects were asked to choose which of the two randomly ordered patterns (left or right) exhibited angles closer to 90°. A total of 24 trials were conducted, taking each user approximately 10 minutes to complete the study.

Results. The results of the comparisons averaged across users and trials are plotted in Figure 8 (c). At 0.3 and 0.5 m, the GC rendering was chosen as closer to the target of 90° in 73.6% and 62.5% of trials, respectively. This is significantly more than FT ($p < 0.001$, respective $p < 0.05$, one-tailed binomial test). The visibility of the difference decreases towards the display distance $d = 0.7$ m where GC was only preferred at near chance level of 51.4%.

These results suggest that accounting for the gaze-contingent no-parallax point is important for correct depth scaling needed to properly convey relative distance and shape of objects within a scene, particularly when a user is verging to a close object or familiar shape, such as a cube. Judging the angle at which two planes meet requires higher-level reasoning and combination of both absolute and relative depth cues. We expect that the distortion can be even easier to detect in tasks where the relative displacement of two surfaces alone is a sufficient cue. We explore this hypothesis in the following AR alignment study.

6 ALIGNMENT INACCURACY IN AR

Many applications in AR desire accurate alignment of digital and physical objects. For example, a surgeon aligning medical data to a patient will want to rely on it being displayed in the correct place. As such, accurate depth rendering is critical. Section 4.3 predicts displacements of virtual objects when the position of the no-parallax point is not taken into account. Here, we experimentally verify visibility of this effect in an AR environment. We further test a hypothesis that our gaze-dependent rendering can noticeably improve the accuracy of alignment between the virtual and real objects.

Hardware and Software. We used a Microsoft HoloLens 1 optical see-through AR headset, which has a diagonal field-of-view of 34°, a refresh rate of 60 Hz and a 1280×720 waveguide display per eye, resulting in a theoretical central resolution of 1.39 arcmin/pixel. As with the VR user experiment, we again used Unity to render all modes and control the user experiment.

Stimuli and Conditions. The stimuli consisted of a single 8 cm tall flat surface, textured with a playing card image (see Fig. 9 (a)), displayed at target fixation distance of either 0.5, 1.0, 1.5 or 2.0 m. Again, these distances were chosen such as to increment to the display distance of the Microsoft HoloLens (2 m), where no disparity distortion should be observed. A physical target was placed at the same distance from the user (as measured from the user's eyes in the physical world) but with a small lateral displacement, such that the virtual and physical objects would appear side by side (see Fig. 9 (b)). In the experiment, subjects viewed the rendered stimulus in three rendering conditions: conventional (HL, i.e., as provided by the Windows Mixed Reality SDK in Unity), fine-tuned (FT), and gaze-contingent (GC). For the HL rendering mode, we implemented the online instructions provided by Microsoft for rendering to the HoloLens with Unity. We let the rendering be set up by the supplied Windows Mixed Reality SDK and only adjusted the IPD setting for each user through their Developer Portal interface. For the FT rendering mode, we followed a similar procedure for adjusting virtual image separation as in Sec. 5. In this case, the manufacturer-provided calibration was fine-tuned for each subject by aligning the card stimulus at a calibration distance of 2 m. Finally, the GC rendering mode was identical to the FT mode, but with the same modifications as in Section 5. This was again motivated by the desire to show that even a more accurate calibration of the IPD is insufficient to remove the misalignment observed at closer distances if the position of the no-parallax point is not taken into account. While wearing the headset, an SR Research head rest was used to keep the subject's head fixed with respect to the physical targets throughout the study.

Procedure. Each set of trials began with the IPD fine-tuning task required to set up the FT rendering for each of the thirteen participants (age range 18–54, 7 female). Each trial constituted a 2AFC, where one of the three target depths was randomly chosen, and subjects were asked to choose which of the two selected modes provided the best alignment in depth with the physical target, which was placed by the researcher before the stimulus was shown. Subjects had the ability to freely switch between the modes using a keyboard key before making a selection, though most users only made a single switch per trial. A total of 12 trials were conducted comparing FT and HL rendering, followed by a short rest break. After which, the calibration was repeated, and another 12 trials were conducted comparing FT to GC rendering.

Results. The results of the comparisons averaged across users and trials are plotted in Figure 9 (c) and (d). At all measured distances the FT rendering achieves significantly better alignment of the rendered and physical stimulus than the HL rendering (100%, 100%, 94.2% and 86.5% of trials, $p < 0.001$, one-tailed binomial test). Some users found it harder to judge the difference between the two modes as the planes moved further away, but overall, it can be seen that fine-tuning user's IPD measurement by calibration almost consistently improved alignment compared to the conventional approach.

Moreover, additional improvement of alignment was observed in the GC rendering mode which achieved a significant preference over the FT for distances of 0.5 m (96.2%, $p < 0.001$) and 1.0 m (71.2%, $p < 0.05$). While it was more difficult to detect differences for larger distances (57.7% at 1.0 m) the results of the experiment confirm our hypothesis that accounting for the gaze-contingent shift of the no-parallax point is crucial for accurate reproduction of stereoscopic disparity. Although fine tuning of the IPD proved helpful, the gaze-contingent rendering was required to ensure good alignment of virtual and physical objects across distances in AR. While the shift may become indistinguishable for far away objects, gaze-contingent stereo rendering could be critical in several near AR tasks, including AR-assisted surgery, maintenance, and training.

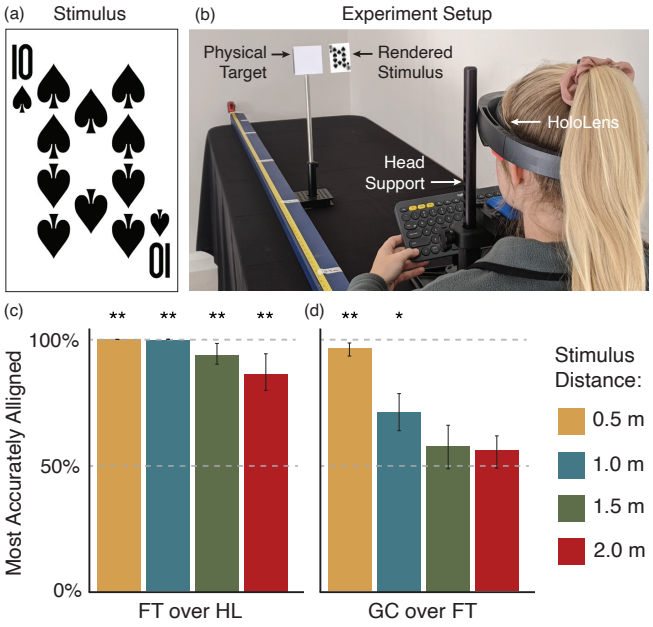

Fig. 9. Evaluating alignment of real and virtual content. Subjects viewed a playing card (a) rendered at a target depth of either 0.5, 1.0, 1.5 or 2.0 m next to a physical reference. In the first set of trials (c), this stimulus was presented with either native HoloLens (HL) or fine-tuned (FT) rendering and we asked subjects to indicate which rendering mode provided the most accurate alignment with the physical target. A photograph of the experiment set-up is shown in (b) (the card is added for illustrative purposes). In the second set of trials (d), subjects were asked to compare fine-tuned and gaze-contingent (GC) rendering. Results of these comparisons show the percentage of times the first member of the pair was chosen over the second. It can be seen that using an initial calibration procedure to accurately measure the subject's IPD significantly improved alignment compared to the standard HoloLens approach for all distances. Furthermore, GC rendering was able to further improve alignment at closer distances indicating that it is most critical for arm's reach viewing. Error bars represent Standard Error (SE) and significance is indicated at the $p < 0.05$ and 0.001 levels with * and ** respectively.

7 DISCUSSION

In summary, we study the disparity distortion induced by ignoring the gaze-contingent location of the no-parallax point in the human eye. Using several user studies, we experimentally validate the location of the no-parallax point and demonstrate that modeling it accurately during stereo rendering significantly reduces disparity and shape distortion in a VR setting and significantly improves consistent alignment of physical and digital objects in an optical see-through AR setting.

The results of our experiments show that disparity distortions are easier to detect in the AR alignment task (Sec. 6) than in the VR shape matching task (Sec. 5). This is expected as the human visual system is sensitive to even small disparity changes between a physical reference and a digitally rendered object [Didyk et al. 2011]. On the other hand, the shape judgment task required subjects to interpret the relative disparity in the context of estimated object distance. Without a real-world reference, the relatively poor absolute depth cue of eye convergence likely increased the difficulty of the task [Richards and Miller 1969].

7.1 Limitations and Future Work

Gaze-contingent stereo rendering relies on robust gaze tracking, which is available in most modern AR systems, but not yet in all VR displays. However, since the magnitude of parallax changes gradually with eye rotation, we do not require extreme accuracy in gaze prediction. For the model situation in Fig. 7 and a central vision fixation distance of 1 m, a 1° differential tracking error between gaze angles of both eyes results in a disparity rendering error of 12 " which is a difference on the limit of human stereoacuity in ideal conditions [Didyk et al. 2011]. The technique is more sensitive to latency as a delayed response could produce visible jumps of disparity. Thus without implementing significant temporal smoothing, HMD eye tracking therefore remains a challenging problem. Furthermore, any stereo rendering approach, including ours, is only as good as the optics and calibration of the headset and the accuracy of user-specific parameters. Variation in lens distortion as the eye rotates off axis and across the lens (commonly referred to as pupil swim) can cause its own disparity distortion. This is not something our approach inherently corrects for, but it could be used in combination with existing pupil swim correction approaches [Ratcliff et al. 2020]. Similarly, in practical use cases there is likely to be per-user variation in the parameters of our eye model or even inaccuracies when measured on a per-user basis. Even so, our model generally pushes the disparity in the empirically correct direction.

While our studies demonstrate statistically significant effects, they all use task-specific stimuli. Studies in more complex environments where the user's cognitive load is higher may provide an interesting setting for future user experiments. Moreover, it would be interesting to explore the adaption of this technique for varifocal and multifocal displays, and the interaction of gaze-contingent stereo rendering with other depth cues.

8 CONCLUSION

Virtual and augmented reality systems strive to create experiences imperceptible from physical ones. The emergence of wearable eye tracking in these systems has seen a number of techniques, such as foveated or gaze-contingent varifocal displays modes, improve user experience and visual comfort. With our work, we demonstrate that eye tracking can also enable us to optimize stereo rendering to account for dynamic gaze effects, and thus improve depth perception and perceptual realism in emerging wearable computing systems.

ACKNOWLEDGMENTS

B.K. was supported by a Stanford Knight-Hennessy Fellowship. G.W. was supported by an Okawa Research Grant and a Sloan Fellowship. Other funding for the project was provided by NSF (award numbers 1553333 and 1839974) and a PECASE by the ARO. The authors would also like to thank Professor Anthony Norcia, for advising on an appropriate experiment to measure depth distortion, and Robert Konrad, for providing additional insights on ocular parallax.

REFERENCES

Kurt Akeley, Simon J. Watt, Ahna R. Girshick, and Martin S. Banks. 2004. A Stereo Display Prototype with Multiple Focal Distances. *ACM Trans. Graph. (SIGGRAPH)* 23, 3 (2004), 804–813.

Kaan Akşit, Ward Lopes, Jonghyun Kim, Peter Shirley, and David Luebke. 2017. Near-eye varifocal augmented reality display using see-through screens. *ACM Transactions on Graphics* 36, 6 (Nov. 2017), 189:1–189:13.

David Atchison and George Smith. 2000. *Optics of the human eye.* Butterworth Heinemann.

David A. Atchison. 2017. Schematic Eyes. In *Handbook of Visual Optics, Volume I - Fundamentals and Eye Optics*, Pablo Artal (Ed.). CRC Press, Chapter 16.

Geoffrey P. Bingham. 1993. Optical flow from eye movement with head immobilized: "Ocular occlusion" beyond the nose. *Vision Research* 33, 5 (March 1993), 777–789.

Johannes Burge, Charless C. Fowlkes, and Martin S. Banks. 2010. Natural-Scene Statistics Predict How the Figure–Ground Cue of Convexity Affects Human Depth Perception. *Journal of Neuroscience* 30, 21 (May 2010), 7269–7280.

Jen-Hao Rick Chang, B. V. K. Vijaya Kumar, and Aswin C. Sankaranarayanan. 2018. Towards multifocal displays with dense focal stacks. *ACM Transactions on Graphics* 37, 6 (Dec. 2018), 198:1–198:13.

Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, and Hans-Peter Seidel. 2011. A Perceptual Model for Disparity. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2011, Vancouver)* 30, 4 (2011).

Neil A. Dodgson. 2004. Variation and extrema of human interpupillary distance. In *Stereoscopic Displays and Virtual Reality Systems XI*, Mark T. Bolas, Andrew J. Woods, John O. Merritt, and Stephen A. Benton (Eds.), Vol. 5291. International Society for Optics and Photonics, SPIE, 36 – 46.

Andrew T. Duchowski, Nathan Cournia, and Hunter A. Murphy. 2004. Gaze-Contingent Displays: A Review. *Cyberpsychology & behavior* 7 (2004), 621–34.

Andrew T. Duchowski, Donald H. House, Jordan Gestring, Rui I. Wang, Krzysztof Krejtz, Izabela Krejtz, Radoslaw Mantiuk, and Bartosz Bazyluk. 2014. Reducing Visual Discomfort of 3D Stereoscopic Displays with Gaze-contingent Depth-of-field. In *Proc. ACM Symp. on Appl. Perc. (SAP)*. 39–46.

David Dunn, Cary Tippets, Kent Torell, Petr Kellnhofer, Kaan Akşit, Piotr Didyk, Karol Myszkowski, David Luebke, and Henry Fuchs. 2017. Wide Field Of View Varifocal Near-Eye Display Using See-Through Deformable Membrane Mirrors. *IEEE TVCG* 23, 4 (2017), 1322–1331.

Wilson S. Geisler and Jeffrey S. Perry. 1998. Real-time foveated multiresolution system for low-bandwidth video communication. In *Human Vision and Electronic Imaging III*, Vol. 3299. International Society for Optics and Photonics, 294–305.

Andrew Glennerster, Brian J. Rogers, and Mark F. Bradshaw. 1996. Stereoscopic depth constancy depends on the subject's task. *Vision Research* 36, 21 (Nov. 1996), 3441–3456.

Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. 2012. Foveated 3D Graphics. *ACM Trans. Graph. (SIGGRAPH Asia)* 31, 6 (2012), 164:1–164:10.

Philippe Hanhart and Touradj Ebrahimi. 2014. Subjective evaluation of two stereoscopic imaging systems exploiting visual attention to improve 3D quality of experience. In *Proc. SPIE vol. 9011*. 0D–1–11.

Sebastien Hillaire, Anatole Lecuyer, Remi Cozot, and Gery Casiez. 2008. Using an Eye-Tracking System to Improve Camera Motions and Depth-of-Field Blur Effects in Virtual Environments. In *IEEE Virtual Reality Conference*. 47–50.

Hong Hua and Bahram Javidi. 2014. A 3D integral imaging optical see-through head-mounted display. *Optics Express* 22, 11 (2014), 13484–13491.

Fu-Chung Huang, Kevin Chen, and Gordon Wetzstein. 2015. The Light Field Stereoscope: Immersive Computer Graphics via Factored Near-Eye Light Field Display with Focus Cues. *ACM Trans. Graph. (SIGGRAPH)* 34, 4 (2015).

Yuta Itoh, Toshiyuki Amano, Daisuke Iwai, and Gudrun Klinker. 2016. Gaussian Light Field: Estimation of Viewpoint-Dependent Blur for Optical See-Through Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 22, 11 (Nov. 2016), 2368–2376.

Yuta Itoh and Gudrun Klinker. 2014. Interaction-free calibration for optical see-through head-mounted displays based on 3D Eye localization. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. 75–82.

David Jacobs, Orazio Gallo, Emily A. Cooper, Kari Pulli, and Marc Levoy. 2015. Simulating the Visual Experience of Very Bright and Very Dark Scenes. *ACM Trans. Graph.* 34, 3, Article 25 (2015), 15 pages.

Changwon Jang, Kiseung Bang, Gang Li, and Byoungho Lee. 2018. Holographic Near-Eye Display with Expanded Eye-Box. *ACM Trans. Graph.* 37, 6, Article 195 (Dec. 2018), 14 pages.

Changwon Jang, Kiseung Bang, Seokil Moon, Jonghyun Kim, Seungjae Lee, and Byoungho Lee. 2017. Retinal 3D: Augmented Reality near-Eye Display via Pupil-Tracked Light Field Projection on Retina. *ACM Trans. Graph. (SIGGRAPH Asia)* 36, 6 (2017).

Adam L Janin, David W Mizell, and Thomas P Caudell. 1993. Calibration of head-mounted displays for augmented reality applications. In *Proc. IEEE Virtual Reality*. 246–255.

Paul V. Johnson, Jared AQ. Parnell, Joohwan Kim, Christopher D. Saunter, Gordon D. Love, and Martin S. Banks. 2016. Dynamic lens and monovision 3D displays to improve viewer comfort. *OSA Opt. Express* 24, 11 (2016), 11808–11827.

Anton S Kaplanyan, Anton Sochenov, Thomas Leimkühler, Mikhail Okunev, Todd Goodall, and Gizem Rufo. 2019. DeepFovea: neural reconstruction for foveated rendering and video compression using learned statistics of natural videos. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–13.

Petr Kellnhofer, Piotr Didyk, Karol Myszkowski, Mohamed M. Hefeeda, Hans-Peter Seidel, and Wojciech Matusik. 2016a. GazeStereo3D: Seamless Disparity Manipulations. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 35, 4 (2016).

Petr Kellnhofer, Piotr Didyk, Tobias Ritschel, Belen Masia, Karol Myszkowski, and Hans-Peter Seidel. 2016b. Motion parallax in stereo 3D: model and applications. *ACM Transactions on Graphics* 35, 6 (Nov. 2016), 176:1–176:12.

Jonghyun Kim, Youngmo Jeong, Michael Stengel, Kaan Akşit, Rachel Albert, Ben Boudaoud, Trey Greer, Joohwan Kim, Ward Lopes, Zander Majercik, et al. 2019. Foveated AR: dynamically-foveated augmented reality display. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–15.

Robert Konrad, Anastasios Angelopoulos, and Gordon Wetzstein. 2020. Gaze-contingent ocular parallax rendering for virtual reality. *ACM Transactions on Graphics (TOG)* 39, 2 (2020), 1–12.

Robert Konrad, Emily Cooper, and Gordon Wetzstein. 2015. Novel Optical Configurations for Virtual Reality: Evaluating User Preference and Performance with Focus-tunable and Monovision Near-eye Displays. In *Proc. SIGCHI*.

George A. Koulieris, Kaan Akşit, Michael Stengel, Rafał K. Mantiuk, Katerina Mania, and Christian Richardt. 2019. Near-Eye Display and Tracking Technologies for Virtual and Augmented Reality. *Computer Graphics Forum* 38, 2 (2019).

Hiroaki Kudo and Noboru Ohnishi. 2000. Effect of the sight line shift when a head-mounted display is used. In *Proc. EMBS International Conference*, Vol. 1. 548–550.

Hiroaki Kudo, Masaya Saito, Tsuyoshi Yamamura, and Noboru Ohnishi. 1999. Measurement of the ability in monocular depth perception during gazing at near visual target-effect of the ocular parallax cue. In *Proc. IEEE International Conference on Systems, Man, and Cybernetics*, Vol. 2. 34–37.

Douglas Lanman and David Luebke. 2013. Near-eye Light Field Displays. *ACM Trans. Graph. (SIGGRAPH Asia)* 32, 6 (2013), 220:1–220:10.

Sheng Liu, Dewen Cheng, and Hong Hua. 2008. An Optical See-through Head Mounted Display with Addressable Focal Planes. In *Proc. ISMAR*. 33–42.

David Luebke and Benjamin Hallen. 2001. Perceptually driven simplification for interactive rendering. In *Rendering Techniques 2001*. Springer, 223–234.

Andrew Maimone, Andreas Georgiou, and Joel S. Kollin. 2017. Holographic near-eye displays for virtual and augmented reality. *ACM Transactions on Graphics* 36, 4 (July 2017), 85:1–85:16.

Radosław Mantiuk, Bartosz Bazyluk, and Anna Tomaszewska. 2011. Gaze-dependent depth-of-field effect rendering in virtual environments. In *Serious Games Development and Appl.* 1–12.

Radoslaw Mantiuk and Mateusz Markowski. 2013. Gaze-Dependent Tone Mapping. In *ICIAR*.

Michael Mauderer, Simone Conte, Miguel A. Nacenta, and Dhanraj Vishwanath. 2014. Depth Perception with Gaze-Contingent Depth of Field. In *Proc. SIGCHI*. 217–226.

Michael Mauderer, David R. Flatla, and Miguel A. Nacenta. 2016. Gaze-Contingent Manipulation of Color Perception. *Proc. SIGCHI* (2016).

Olivier Mercier, Yusufu Sulai, Kevin Mackenzie, Marina Zannoli, James Hillis, Derek Nowrouzezahrai, and Douglas Lanman. 2017. Fast Gaze-contingent Optimal Decompositions for Multifocal Displays. *ACM Trans. Graph. (SIGGRAPH Asia)* 36, 6 (2017).

Hunter Murphy and Andrew Duchowski. 2001. Gaze-Contingent Level Of Detail Rendering. *EuroGraphics* 2001 (01 2001).

Toshikazu Ohshima, Hiroyuki Yamamoto, and Hideyuki Tamura. 1996. Gaze-directed adaptive rendering for interacting with virtual space. In *Proc. IEEE VR*. IEEE, 103–110.

Nitish Padmanaban, Robert Konrad, Tal Stramer, Emily A. Cooper, and Gordon Wetzstein. 2017. Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *PNAS* 114, 9 (2017), 2183–2188.

Nitish Padmanaban, Yifan Peng, and Gordon Wetzstein. 2019. Holographic Near-Eye Displays Based on Overlap-Add Stereograms. *ACM Trans. Graph. (SIGGRAPH Asia)* 6 (2019). Issue 38.

Jae-Hyeung Park and Seong-Bok Kim. 2018. Optical see-through holographic near-eye-display with eyebox steering and depth of field control. *Opt. Express* 26, 21 (2018), 27076–27088.

Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG)* 35, 6 (Nov. 2016), 179:1–179:12.

Eli Peli, T Reed Hedges, Jinshan Tang, and Dan Landmann. 2001. A Binocular Stereoscopic Display System with Coupled Convergence and Accommodation Demands. In *SID Symposium Digest of Technical Papers*, Vol. 32. 1296–1299.

Yifan Peng, Suyeon Choi, Nitish Padmanaban, Jonghyun Kim, and Gordon Wetzstein. 2020. Neural Holography. In *ACM SIGGRAPH Emerging Technologies*.

Alexander Plopski, Yuta Itoh, Christian Nitschke, Kiyoshi Kiyokawa, Gudrun Klinker, and Haruo Takemura. 2015. Corneal-Imaging Calibration for Optical See-Through Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 21, 4 (2015), 481–490.

Joshua Ratcliff, Alexey Supikov, Santiago Alfaro, and Ronald Azuma. 2020. ThinVR: Heterogeneous microlens arrays for compact, 180 degree FOV VR near-eye displays. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (May 2020), 1981–1990.

Whitman Richards and John F Miller. 1969. Convergence as a cue to depth. *Perception & Psychophysics* 5, 5 (1969), 317–320.

Jannick P. Rolland, Myron W. Krueger, and Alexei Goon. 2000. Multifocal planes head-mounted displays. *OSA Appl. Opt.* 39, 19 (2000), 3209–3215.

Heiko H Schütt, Stefan Harmeling, Jakob H Macke, and Felix A Wichmann. 2016. Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research* 122 (2016), 105–123.

Liang Shi, Fu-Chung Huang, Ward Lopes, Wojciech Matusik, and David Luebke. 2017. Near-eye Light Field Holographic Rendering with Spherical Waves for Wide Field of View Interactive 3D Computer Graphics. *ACM Trans. Graph. (SIGGRAPH Asia)* 36, 6, Article 236 (2017), 236:1–236:17 pages.

Takashi Shibata, Joohwan Kim, David M. Hoffman, and Martin S. Banks. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision* 11, 8 (2011), 11–11.

Y Shin, HW Lim, MH Kang, M Seong, H Cho, and JH Kim. 2016. Normal range of eye movement and its relationship to age. *Acta Ophthalmologica* 94 (2016).

Qi Sun, Fu-Chung Huang, Joohwan Kim, Li-Yi Wei, David Luebke, and Arie Kaufman. 2017. Perceptually-guided foveation for light field displays. *ACM Transactions on Graphics* 36, 6 (Nov. 2017), 192:1–192:13.

Mihran Tuceryan and Nassir Navab. 2000. Single point active alignment method (SPAAM) for optical see-through HMD calibration for AR. In *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*. 149–158.

Jacek Turski. 2016. On binocular vision: The geometric horopter and Cyclopean eye. *Vision research* 119 (2016), 73–81.

Margarita Vinnikov and Robert S. Allison. 2014. Gaze-contingent Depth of Field in Realistic Scenes: The User Experience. In *Proc. Symp. on Eye Tracking Res. and Appl. (ETRA)*. 119–126.

Božo Vojniković and Ettore Tamajo. 2013. Horopters–Definition and Construction. *Collegium antropologicum* 37, 1 (2013), 9–12.

Felix A Wichmann and N Jeremy Hill. 2001. The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & psychophysics* 63, 8 (2001), 1293–1313.

Albert Yonas, Lincoln G. Craton, and William B. Thompson. 1987. Relative motion: Kinetic information for the order of depth at an edge. *Perception & Psychophysics* 41, 1 (01 Jan 1987), 53–59.