Keep It Simple: Depth-based Dynamic Adjustment of Rendering for Head-mounted Displays Decreases Visual Comfort

JOCHEN JACOBS, XI WANG, and MARC ALEXA, TU Berlin

Head-mounted displays cause discomfort. This is commonly attributed to conflicting depth cues, most prominently between vergence, which is consistent with object depth, and accommodation, which is adjusted to the near eye displays.

It is possible to adjust the camera parameters, specifically interocular distance and vergence angles, for rendering the virtual environment to minimize this conflict. This requires dynamic adjustment of the parameters based on object depth. In an experiment based on a visual search task, we evaluate how dynamic adjustment affects visual comfort compared to fixed camera parameters. We collect objective as well as subjective data. Results show that dynamic adjustment decreases common objective measures of visual comfort such as pupil diameter and blink rate by a statistically significant margin. The subjective evaluation of categories such as fatigue or eye irritation shows a similar trend but was inconclusive. This suggests that rendering with fixed camera parameters is the better choice for head-mounted displays, at least in scenarios similar to the ones used here.

CCS Concepts: • **Computing methodologies** → **Rendering**;

Additional Key Words and Phrases: Head-mounted displays, vergence, fatigue

ACM Reference format:
Jochen Jacobs, Xi Wang, and Marc Alexa. 2019. Keep It Simple: Depth-based Dynamic Adjustment of Rendering for Head-mounted Displays Decreases Visual Comfort. *ACM Trans. Appl. Percept.* 16, 3, Article 16 (September 2019), 16 pages.
https://doi.org/10.1145/3353902

1 INTRODUCTION

Immersive virtual reality (VR) has become commonplace with the advent of affordable head-mounted stereo displays (e.g., HTC Vive, Oculus Rift, PlayStation VR, Google Daydream View, etc.). These devices are equipped with two displays, one for each eye, presenting two different images rendered based on the view of each of the eyes. This provides realistic binocular depth cues; in particular, it requires convergence or divergence of the eyes toward the object of interest. However, accommodation is adjusted to the focal plane of the device, which is usually set to a fixed distance of a few meters from the viewer. The inconsistency between vergence and accommodation, the *vergence-accommodation (VA) conflict*, is assumed to be the major source of discomfort experienced by many individuals under prolonged use of head-mounted stereo displays.

While technical solutions are possible to adjust the accommodation (Johnson et al. 2016; Konrad et al. 2016; Padmanaban et al. 2017), they are technically involved and likely unavailable for the consumer market. A variety

Authors' address: J. Jacobs, X. Wang, and M. Alexa, TU Berlin, Marchstrasse 23, Berlin 10587, Germany; emails: jochen.jacobs@campus.tu-berlin.de, {xi.wang, marc.alexa}@tu-berlin.de.

1544-3558/2019/09-ART16 $15.00
https://doi.org/10.1145/3353902

of software approaches have been shown to be ineffective (Koulieris et al. 2017). With the advent of eye tracking built into the display, solutions based on dynamic adjustment of rendering parameters depending on the current vergence situation become tractable. Specifically, we adopt the idea of modifying the interocular distance and vergence of the virtual cameras so that the vergence induced by a rendered object matches the accommodation induced by the display (see Section 3 for the details of our approach). This consistency of vergence and accommodation comes at the expense of dynamic modification of perceived absolute depth. Importantly, however, relative depth perception remains intact.

In an experiment based on a visual search task, we evaluate the visual comfort relative to a baseline method that uses fixed interocular distance and parallel view directions for rendering. We are specifically interested in how matching of vergence and accommodation affects objective measures of discomfort such as pupil diameter and blink rate and subjective assessment based on self-report. As part of this evaluation we also determine if participants notice the dynamic adjustment at all, and if so, how this behavior is judged.

Results show that dynamic adjustment is commonly not noticed by participants and has no significant impact on task performance. Interestingly, it nonetheless decreases visual comfort based on objective measurements by a significant margin. This result is consistent with subjective assessment. We conclude that dynamic adjustment of camera parameters to reduce the VA conflict, even if not consciously noticed by participants, introduces equivalent amount of visual fatigue to the constant inconsistency between vergence and accommodation.

2 RELATED WORK

2.1 Vergence-accommodation Conflict

Vergence describes the type of eye movements when both eyes move in opposite directions (Holmqvist et al. 2011). For example, when we change gaze from distant to close objects, both eyes rotate inward. At the same time, the ciliary muscle changes the shape of the lens such that sharp images are obtained on the retina (Atchison et al. 2000). During this accommodation the power of the eye lenses are adjusted. Vergence and accommodation are coupled, and we can measure the level of accommodation using the vergence angle; however, the accommodation level is not equivalent to the change in lens power. Accommodation is much slower than the movements of the eyes (Lockhart and Shi 2010a), taking about 200–500ms, whereas saccades are around 30–40ms (Holmqvist et al. 2011).

In conventional stereoscopic displays, two different images are presented to the eyes and the disparity between the two images correspond to the depth of scene elements. In other words, objects at different distances correspond to different vergence angles, facilitating depth perception. However, the accommodative images are presented on screen with a fixed distance to the eyes. The resulting unnatural decoupling of vergence and accommodation leads to conflicting cues. Studies have shown that this conflict contributes substantially to the visual discomfort in stereo displays (Lockhart and Shi 2010b; Schor et al. 1999; Shiwa et al. 1996).

2.2 Available Solutions to the Vergence-Accommodation Conflict

Several methods have been proposed to reduce the conflict between vergence and accommodation in stereo viewing conditions, including both algorithm solutions (e.g., depth-of-field (DoF) rendering (Duchowski et al. 2014)) and hardware designs (e.g., light-field displays (Maimone et al. 2013)).

Many available algorithmic solutions have been proposed to reduce the VA conflict by modifying the stereo images (Peli et al. 2001). The essential idea is to adapt the vergence angle to the accommodation level on the screen. In virtual environments, camera parameters, including interocular distance and vergence angle, are adjusted dynamically based on object depth. These methods are commonly called convergence adjustment (Fisker et al. 2013; Sherstyuk et al. 2012). Other methods aim to minimize the conflict by remapping the disparity function of depth such that most scene content is viewed in a comfort zone (Lang et al. 2010), where image-based saliency maps are used to guide the stereoscopic image warping. By tracking the eye movements, the remapping

function can be dynamically adjusted based on where is currently being looked at, and it has been demonstrated to improve depth perception (Kellnhofer et al. 2016). Camera adjustment was proposed in context of interactive stereoscopic applications (Oskam et al. 2011), where large motion of viewers or objects often results in visual artifacts. Linear interpolation of camera parameters was proposed to control the disparities of visual content. The idea was further applied in Koulieris et al. (2016), which trained a decision forest to predict gaze positions in stereo images of video games. Instead of using image-based saliency maps as in Lang et al. (2010), local disparity adjustment based on predicted gaze positions was then used to improve the perceptual experience. Results demonstrated that gaze-based disparity manipulation generalized well, especially for scenes with a large depth range. We refer our readers to work by Terzić and Hansard (2016) for a comprehensive review on available solutions. In this work, we focus on the experimental examination of a vergence-based camera adjustment algorithm and its effects on the vergence-accommodation conflict. Other object-based local disparity adjustments, which remap the range of depth variations in the scene (Lang et al. 2010) or dynamically change the disparities of certain objects (Kellnhofer et al. 2016), are considered as different approaches, as the adjustment algorithms depend on the scene content.

A recent study (Koulieris et al. 2017) proposed a device design to measure the accommodation in head-mounted displays and evaluated the effectiveness of several algorithmic methods and hardware designs in handling the VA conflict. The results showed that only the focus-adjustable-lens design (Johnson et al. 2016; Konrad et al. 2016; Padmanaban et al. 2017) can significantly improve visual comfort, where accommodation is changed effectively.

2.3 Visual Comfort Measurements

Visual discomfort has been a major drawback that poses limitations on the usage of stereoscopic displays. The question of how VA conflict affects visual comfort and fatigue has led to a rich body of literature (Kooi and Toet 2004; Shibata et al. 2011). Visual comfort is mostly assessed by questionnaires of subjective evaluation (Chen et al. 2011; Shibata et al. 2011; Tam et al. 2011). Typically participants are asked to report their fatigue, eye strain, body strain, and headache. Eye tracking data have been used as an indicator of mental fatigue in many studies (Yamada and Kobayashi 2017, 2018; Zhao and Shen 2010). Especially with the improved accessibility of video-based eye tracking techniques, more studies report the characteristics of eye movements as an objective measure of visual comfort (Iatsun et al. 2013; Kim et al. 2011; Morad et al. 2000a). Interestingly, this is not the case in most studies of visual comfort in head-mounted displays.

3 DEPTH-BASED DYNAMIC CAMERA ADJUSTMENT

In this work, we aim to evaluate the effect of depth-based dynamic camera adjustment methods on visual comfort of head-mounted displays (HMDs). The central idea is to avoid the vergence-accommodation conflict by adjusting the camera parameters for rendering so that the vergence matches the accommodation on the physical display. As vergence depends on object depth, the approach requires estimating the depth users are attending to and then slowly adjusting the parameters.

We start with a brief overview of rendering for HMDs. As accurate gaze depth estimation is a prerequisite for any dynamic adjustment, we propose a probability model incorporating uncertainty presented in eye movements data as well as ambiguity with respect to scene geometries. Finally, we introduce two types of camera adjustments and a simple protocol of how to combine them based on the gaze depth.

3.1 Background in Head-Mounted Display Rendering

The major optical components in most HMDs are micro-displays and magnifying lenses, see Figure 1. Through magnification a virtual image is presented at a distance of approximately 2m in front of the eyes. Note that the center point of each eye corresponds to different positions in the visual image. To generate the presentation of the scene, two cameras are placed at eye positions and two images are rendered, one for each eye. Disparity of

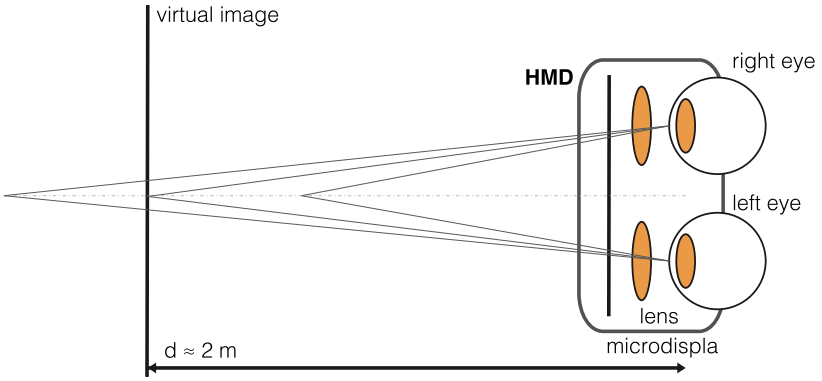


Fig. 1. Illustration of a simplified head-mounted display. Through lens magnification, a virtual image is presented at about 2m in front. When the eyes attend to objects at different depths, the corresponding vergence angle reduces from close to far.

the same object between two images facilitates depth perception. As shown in Figure 1, closer objects correspond to larger vergence angles compared to objects at larger distances.

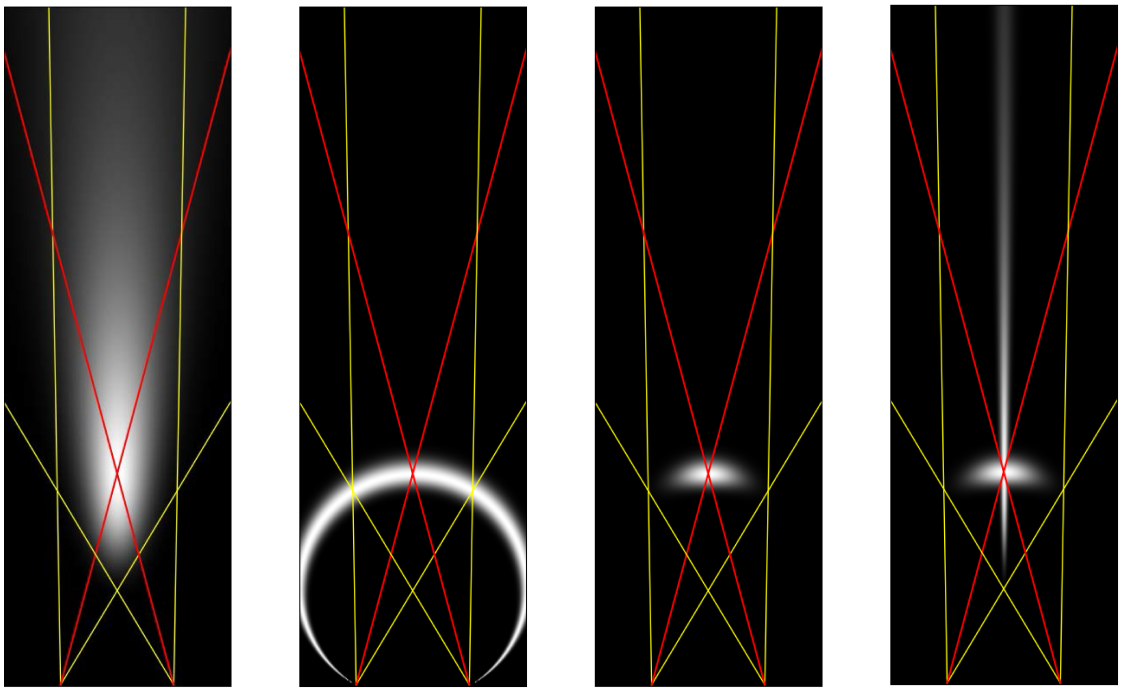
3.2 Accurate Depth Estimation

Our proposed method of real-time dynamic camera adjustment poses a demanding requirement on the accuracy of gaze estimation. The problem of accurately estimating gaze in HMDs is a topic of several current projects (e.g., OpenEDS Challenge (Garbin et al. 2019)). Especially in practice, even a slight shift of the headset would change the underlying gaze mapping functions and result in large errors in the estimated gaze directions.

Vergence based three-dimensional (3D) gaze estimation is an ill-conditioned problem and associates to a large inherent error (Wang et al. 2017). In face, two lines of sight do not even necessarily intersect in space. To better incorporate the uncertainty, we propose a probability model to estimate the gaze positions in 3D. We associate normal distributions to each estimated eye ray direction (i.e., Gaussian distributions with $p(\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\theta-\mu}{\sigma})^2}$) and consider the regions of all possible intersection points, each with its corresponding joint probability. We also explicitly model the dependency between the two eye ray directions. For example, a divergence of two viewing directions is less likely to happen, and its corresponding probability supposes to be low. In such a way, we also assume that errors in the estimated eye ray directions (e.g., caused by camera noise) are dependent, and two eye rays are more likely to have the same directional errors. For each estimated viewing direction represented by θ, we introduce a deviation angle δ to compute all possible intersection points under small perturbation. Given two deviated eye ray directions, we calculate the difference between the two deviation angles $\delta_l - \delta_r$ and estimate how likely this angle difference can be observed according to the distribution p_d. In summary, the probability associated to each intersection point is

$$p(\theta_l, \theta_r) = p_{\theta_l}(\delta_l|\mu, \sigma^2) p_{\theta_r}(\delta_r|\mu, \sigma^2) p_d\left(\delta_l - \delta_r|\mu, \sigma_d^2\right), \tag{1}$$

where θ_l and θ_r are the angles of the left eye ray and the right eye ray. We assume that the deviated eye ray follows a normal distribution around the given ray direction, with the mean $\mu = 0$ and the standard deviation σ. The angles δ_l (or δ_r) represents a deviation angle from the measured left (or right) eye ray. σ_d is the standard deviation for the probability distribution of the difference between the deviation angles. Using a test scene that contains 29 selected target points sampled from the 3D space, we have experimentally determined that $\sigma_d = 0.15\sigma$ produces the best result. The joint probability distribution of the intersection point in 2D is visualized in Figure 2(a) and the probability distribution of the angle difference p_d in Figure 2(b). Multiplying the probability distribution in (a) and (b) results in a joint distribution as shown in Figure 2(c), which largely limits the possible intersection



(a) $p_{\theta_l} p_{\theta_r}$ (b) p_d (c) $p_{\theta_l} p_{\theta_r} p_d$ (d) $p_{\theta_l} p_{\theta_r} p'_d$

Fig. 2. Probability distributions of the estimated position of a fixation target in 2D. Red lines correspond to the given eye ray directions and yellow lines mark the area within two standard deviations of 2σ. (a) The joint distribution of $p_{\theta_l} p_{\theta_r}$, with each modeled as a normal distribution and (b) the distribution of p_d, following a normal distribution with standard deviation of 0.15σ. Product of the combined distribution of $p_{\theta_l} p_{\theta_r} p_d$ is shown in (c); however, if we only consider the absolute differences between angles (without directional information), then error occurs in the estimated probability distribution as shown in (d), where p'_d corresponds to the distribution of differences between unsigned angles.

regions given two eye rays. Apart from the ratio between σ and σ_d, the exact values do not change the position of the maximum response as shown in Figure 2(c). As a common practice in eye tracking, directional angles are considered. Ignoring the sign of rotation angles would lead to artifacts in the joint distribution as visualized in Figure 2(d). As two lines do not necessarily intersect in space, this probability model provides a reasonable distribution, especially when estimating gaze positions in 3D under uncertainty. The last factor we consider is the scene geometry: We assume all intersection points lie on the geometry of the scene (i.e., no gaze points are located in the air); only points visible to both eyes are considered as potential fixation targets, as intended targets are visible to both eyes in our designed scenario and the estimation of gaze depth relies on the vergence angle formed by two eye rays. Standard ray-scene intersection test was used to compute the visibility.

3.3 Interocular Distance and Vergence Angle

We consider two different transformations of the cameras, namely translation and rotation, which correspond to the modifications in interocular distances and vergence angles, formed by the two cameras and the position of the fixation target. Note that each of the parameters can be used to increase or decrease the effective vergence angle for the user of the HMDs. As gaze estimation follows the changes of viewing directions, it is important that scene objects are visible in both of the two images rendered for the two eyes.

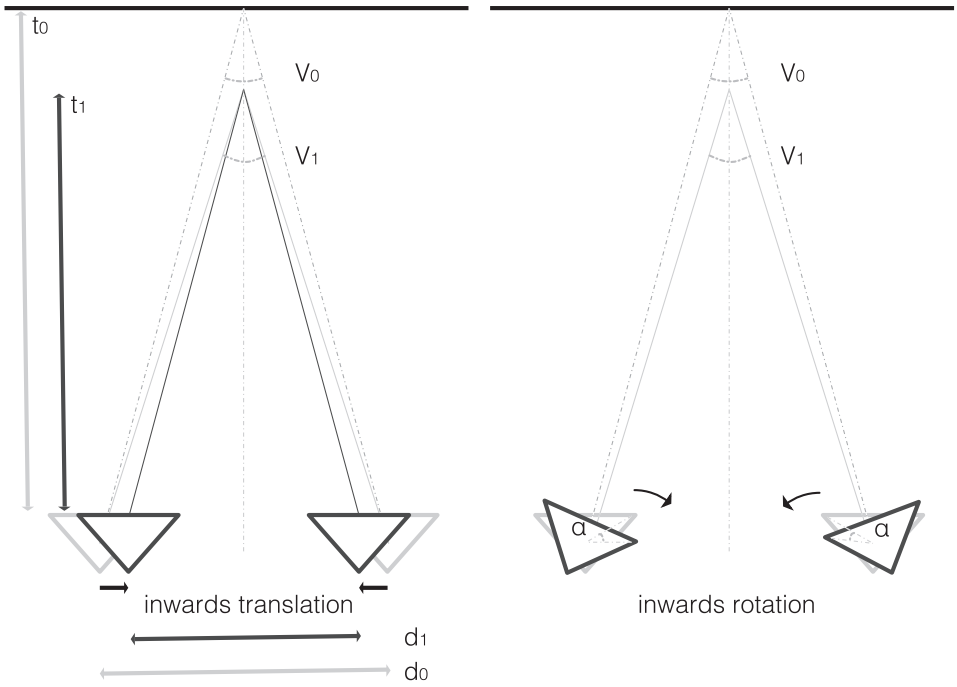


Fig. 3. Illustration of camera adjustments when the fixation target is located in front of the screen. In such a case, we can either decrease the camera distance as shown on the left or rotate both cameras inward as shown on the right. Camera distance is denoted by d and target distance by t. Vergence angle is represented by V. Reducing camera distance from d_0 to d_1 reduces the vergence angle for gaze point in front of the screen at t_1 (see Equation (2)). By rotating both cameras inward by α degree, we effectively reduce the vergence angle from V_1 to V_o when viewing the rendered images parallel (see Equation (3)).

The idea of adjusting camera parameters to improve visual comfort has been proposed in Peli et al. (2001), but only parallel camera setups were considered. In practice, this can be implemented as horizontal shifts of the two images, which results in vergence eye movements. Such adjustment of two rendered images often generates artifacts that lead to reduced visual comfort. As shown in previous studies, inward camera rotation (also called camera toeing-in) introduces vertical disparities (Stelmach et al. 2003; Woods et al. 1993), especially for objects appear close to the corners of the image. For a given camera toeing-in configuration, closer objects result in larger distortion than objects at larger distances. Our idea is to parameterize the solution space in terms of camera parameters by considering changes of both interocular distance and vergence angle. This gives us the freedom to choose between camera translation and rotation, and we propose a simple protocol to combine both transformations such that distortions in the rendered images are minimized (more details in Section 3.4).

Distance between Two Cameras. We only consider the one-dimensional translation along the axis between two cameras. Increased camera distance results in larger disparity in the rendered images, consequently increasing the corresponding effective vergence angle. As an example shown on the left in Figure 3, when fixation targets come closer to the eyes, we move the virtual cameras toward each other. The distance between two cameras is proportional to the depth of the fixation target:

$$\frac{d_1}{t_1} = \frac{d_0}{t_0}. \tag{2}$$

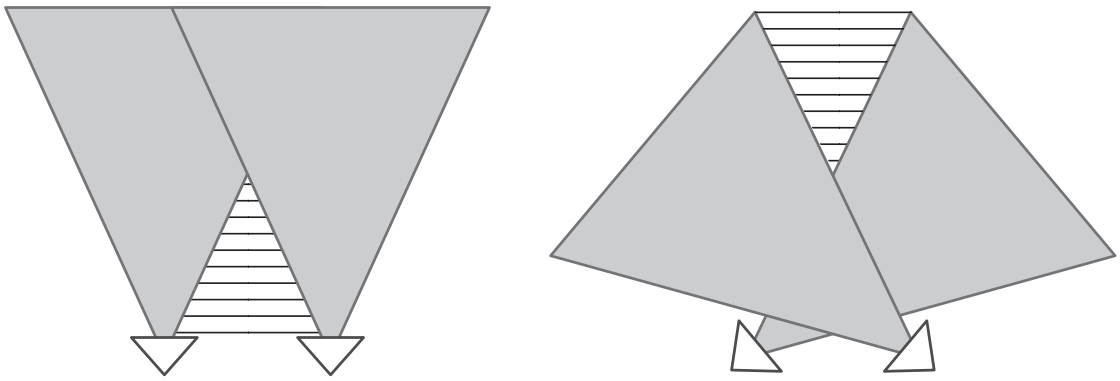


Fig. 4. Extreme parameter settings for the virtual cameras. Gray triangles indicate the view frustums for each camera (small white triangle). Only the intersection area of the gray triangles is useful in practice, covering only a small area away from the cameras for large interocular distance (left), and close to the cameras for extreme inward rotation (right). The areas marked with horizontal lines are completely invisible. Note that these illustrations are exaggerations—real settings are kept within a physiologically plausible range.

Here d denotes the distance between the two virtual cameras and t describes the distance of the fixated object to the eyes. We limit the distance of the virtual cameras to a plausible physiological range, as unrealistic adjustments lead to visual artifacts: small camera distance leads to a loss of stereoscopic depth cues with two nearly identical images; large camera distance leads to double vision, and, at extreme distances, close objects become invisible as the required vergence angle becomes too large (see Figure 4, left). Note that only symmetric viewing frustums are considered here. Parallel projection with asymmetric frustum may effectively reduce the invisible areas in front as shown on the left in Figure 4, and it has been used in previous work to correct the keystone distortions (Zelle and Figura 2004).

Vergence Angle between Two Cameras. It is common to set the optical axes orthogonal to the plane formed by the axis between two cameras and the up vector. This means the optical axes are parallel. In our scenario, we want to rotate them. The change in angle of the optical axes of the virtual cameras changes the effective vergence angle of the eyes of the user (in the opposite direction).

Let α be the rotation angle of each camera when the current vergence angle V_1 is modified to V_o, corresponding to the vergence angle when viewing objects at screen distance:

$$\alpha = \frac{|V_o - V_1|}{2}. \tag{3}$$

Similarly to the adjustment of camera distances, large inward rotation leads to double vision and large outward rotation results in discontinuity in the perceived scene. The right illustration in Figure 3 shows an example. Note that when viewing the scene, we still assume the two cameras are parallel to each other but the images are rendered from adjusted view points. If the cameras converge (i.e., rotate inward), then objects that are farther away than the convergence point would require divergent eye movements, which is essentially invisible as shown on the right in Figure 4. Therefore, the cameras can only converge as close as the farthest object. Large camera divergence also leads to large convergence of the eyes, but the largest camera divergence angle required when looking at infinity is only 2°, which is the vergence angle corresponding to screen distance.

3.4 From Vergence Angle to Camera Adjustment

Using the method described in Section 3.2, we estimate the location of the fixation target as well as the corresponding vergence angle, given two viewing directions reported by the eye tracker. Recall that the idea is to

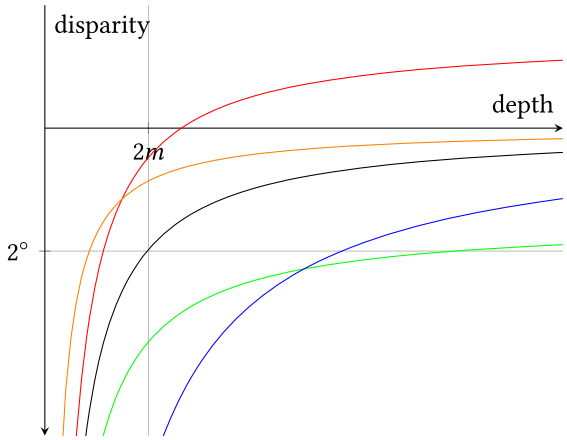


Fig. 5. The disparity function of depth. Black line shows the default disparity map. Orange line draws the mapping when distances between two cameras are small while blue line depicts the mapping when camera distances are large. Green line corresponds to diverging cameras while red line corresponds to converging cameras.

dynamically adjust the camera parameters such that the vergence angle is kept as close as possible to V_o, which is the vergence angle when viewing objects at screen distance. Specifically, we consider the camera parameters of interocular distance and vergence angle.

If fixated objects are in front of the screen, then we can either reduce the interocular distance by moving the cameras closer to each other or rotate them inward to change the vergence angle. Similarly, when attention changes to points that are far away with a large depth value, we can either increase the interocular distance or rotate the cameras outward. Based on the assumption that unexpected changes in the scene may distract users, we aim to minimize such changes while dynamically adjusting the cameras. We experimented with different parameter settings as well as the speed of adjustments.

In practice, each parameter can be adjusted only within a limited range. In principle, adjustment of interocular distance corresponds to the scaling of the total depth range, as shown by the orange and blue curves in Figure 5; adjustment of vergence angle corresponds to shifts in the disparity depth map as depicted by the green and red curves in Figure 5. As discussed in previous section, large camera distance leads to large disparity for closer objects (the blue curve) while converging cameras causes problems to objects that are far away (the red curve). As inward camera translation (i.e., close camera distance) and outward camera rotation (i.e., divergent cameras) up to 2° do not lead to visual artifacts, we optioned for a simple protocol: Only interocular distance is changed when viewing objects in front of the screen, and camera rotation is applied when viewing objects behind the screen. Therefore, the camera distance was adjusted not larger than the average interocular distance; the vergence angle was kept not larger than the angle when focusing on the screen, which corresponds to 2° of visual angle. In such a way, we avoid large noticeable artifacts caused by inward camera rotations, i.e., the vertical disparities that appear close to the image corners of toeing-in cameras. Note that even though no object-based disparity adjustment is included and the relative order of objects remains the same, the perceived absolute distance between objects can be affected by the camera adjustments, as illustrated in Figure 5. The adjustment speed was set constant such that the fixated point changed at a rate of $0.2\frac{D}{s}$, $1D$ (dioptre) $= m^{-1}$. Instead of using arcmin/s, which varies depending on the depth, we measure the change rate in $\frac{D}{s}$, which can be easily used for both types of adjustment.

4 EXPERIMENT

To evaluate the effect of proposed dynamic adjustment of camera parameters on visual comfort, we designed an experiment where participants perform a visual search task. Fixed camera parameters of interocular distance

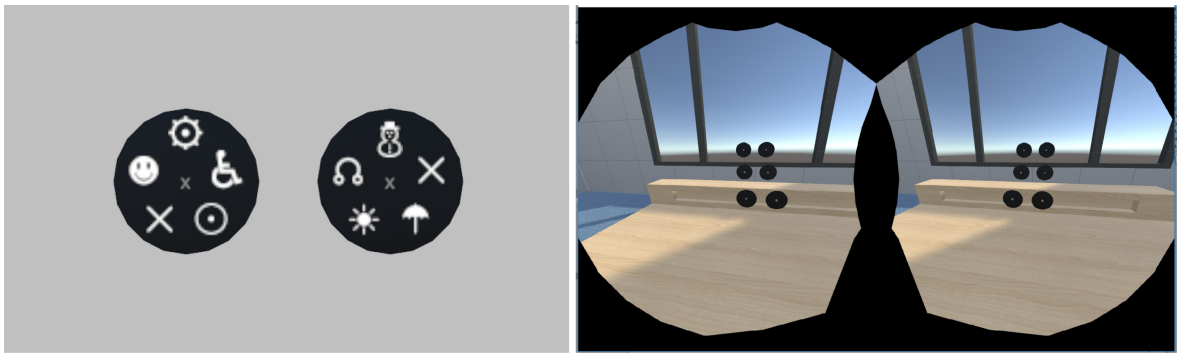


Fig. 6. Experimental scene. The left figure shows one pair of disks, with five symbols on each. Two disks are presented simultaneously to both eyes and participants suppose to find the symbol that appears in both disks (i.e., the cross in this example). The right figure shows the stereo images of tested three depth distances. In the experiment, only one pair at one distance is visible at one time.

equal to 6.9cm and parallel viewing directions serve as a baseline for comparison. We collected both objective eye movement data and subjective evaluations of visual comfort. Additionally, we also collected subjective assessment of the two conditions.

4.1 Participants

Eighteen participants (all students from the university) joined our experiment (5 female, mean age = 24, SD = 3.7). They all had normal or corrected to normal visual acuity. Three of them wore contact lenses. Glasses were not allowed due to concerns about eye tracking accuracy in the HMDs. Fourteen participants reported they had previous experience in VR, and the average rating of their experience between 1 (very bad) and 5 (very good) was 3.8. They were kept naive as to the purpose of the experiment, and their time was compensated at common hourly rates. Written consent was given before the experiment. We also tested the stereoscopic vision of participants using the FLY stereo acuity test (Vision Assessment Corporation). The average stereo acuity is 56s of arc. There was no apparent correlation between the stereo acuity and the objective or subjective measures.

4.2 Apparatus

We used a HTC Vive Pro headset together with its motion tracking system. The touch pads of two HTC Vive Pro controllers were used for the selections in the experiment. Two add-on eye cameras (with frame rate of 120Hz) from Pupil Labs were inserted in the headset to track the eye movements. Unity (Version 2018.3) was used to setup and render the virtual scene, and together with StreamVR they controlled the display and interaction. We used the Unity pupil plugin provided by Pupil Labs to interface with the eye tracker.

4.3 Task

We considered a visual search task to engage participants in the experiments and used the task completion time as an indicator of fatigue, assuming less fatigue corresponds to shorter time. Additionally, task completion time is linked to visual performance. For instance, it indicates whether depth perception is influenced by the dynamic adjustments of the cameras. In the experiments, participants played a Dobble game (also called Spot it!), which is a simple pattern recognition game. As shown on the left in Figure 6, the task was to find a pattern shown on two disks.

In each trial, two disks were presented in front with each showing 5 Unicode characters from the Miscellaneous Symbols block. All 10 symbols were presented simultaneously in front, and the left image in Figure 6 shows an

example. Participants were asked to find the symbol that appeared in both disks, and there was exactly one symbol in common in each trial. Two touch pads on the controllers were used as interface, and participants needed to move the cursors to the selected symbols on both discs simultaneously. It continued to the next trial when the selection was completed. To minimize the bias caused by various gaze positions in the previous trial, we replaced the cursors at the center at the beginning of each trial, and participants were asked to look at them. The discs were colored red for a short time if wrong symbols were selected. Participants were asked to find the matching symbols as quickly as possible. We computed a performance score based on trial correctness and completion time for each participant, and this score was presented to participants as a motivation.

4.4 Protocol

We wanted to compare the effects of dynamic adjustment and fixed parameter rendering on visual comfort and considered these two methods as two different conditions in the experiment. In condition one, camera parameters were dynamically adjusted based on the depth of fixation target; in condition two, fixed camera parameters were used for all participants, and images were rendered by two parallel front-facing cameras with the interocular distance set to 6.9cm. We suspected that the inter-subject variance may be high and decided to collect data of both conditions from each participant. Half of the participants started with the session with camera adjustment and half of them had the session with fixed parameters first.

To test the effects when viewing objects at different depths, we considered three distances for presentation, as shown on the right in Figure 6. The backgrounds of the scene were kept the same for all participants in both conditions. The accommodation on the screen corresponds to a viewing distance of 2m, and we considered a closer distance at 0.46m and a further distance at 50m. To ensure a comparability for the search task, we varied the disk size (as well as the symbol size) such that they spanned the same visual angle of 8.3°.

Three variations in depth lead to six directional jumps in total. We counterbalanced the sequences of jumps over all participants and considered one round of going through all six jumps as one block. Five trials were presented at each distance level to have stable results. Each distance appeared more than once in one block, and in total each block consisted of 35 trials. To motivate participants, we showed them their performance score of each block, as well as the rank of their score in the collected dataset. By doing so, we could also align the time so that each block started at the same time for each participant and the total time wearing the headset was the same for all participants.

One session of six blocks took about 20 minutes in total. Each block started with a calibration of the eye tracker. The calibration of the next block is used as validation of the previous block. We planed to re-calibrate the eye tracker when the validation accuracy was above 2°, but this was never the case. Participants had a trial session at the beginning to familiarize themselves with the procedure. At the end of each session, they were asked to fill out a questionnaire to evaluate their fatigue (detailed questions can be found in the supplementary material). Their eye movements data during the sessions were collected, including eye ray direction, pupil diameter, and occurrence of blink.

4.5 Measures of Visual Comfort

Objective Measures. Task performance was evaluated by the accuracy and duration of task completion, namely whether the matching pair of symbols was correctly selected and how long it took to find the matching pair. To measure fatigue, we considered the common eye movement statistics (Cardona et al. 2011; Kim et al. 2011; Luedtke et al. 1998; Morad et al. 2000b), including blink rate, pupil diameter as well as pupil diameter variation. With increased fatigue, blink rate as well as pupil diameter variability are expected to increase while average pupil diameter is expected to decrease.

Subjective Measures. At the end of each session, participants answered a questionnaire to evaluate their experience including questions about their eyes, vision, focus, headache, and general feeling. After the completion of

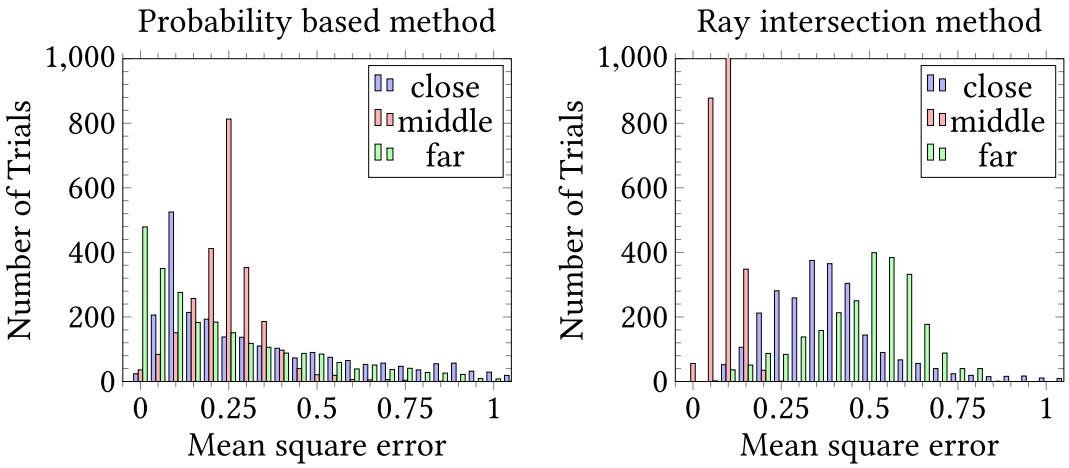


Fig. 7. Histograms of error distribution of estimate fixation targets using the proposed probability-based method (on the left) and the ray intersection method (on the right). We distinguish among three different depths and the proposed method improved the estimation accuracy in general.

both sessions, participants also reported their preference of the two sessions. We followed the standard visual comfort evaluation protocol (Hoffman et al. 2008; Shibata et al. 2011), and participants rated their experience on a 5-point Likert scale (see the supplementary material for details) where 1 indicates positive experience and 5 corresponds to negative experience.

5 RESULTS

We assess the effects of the described depth-based dynamic camera adjustment method on visual comfort compared to fixed camera parameters, following the visual search experiment. Here we report the results of both subjective and objective evaluations.

5.1 Depth Estimation Accuracy

First, we evaluated our proposed probability-based method for depth estimation and compared it to the common strategy of computing the intersection point of two eye rays by finding the point that has the smallest distances to both eye rays. The estimation error was computed by the reciprocal length between the estimated depth and the intended depth, assuming participants were looking at the discs while performing the task. As shown in the equation

$$e = \left(\frac{1}{d_f} - \frac{1}{d_o}\right)^2, \tag{4}$$

d_f is the depth of the estimated fixation target and d_o is the disk depth. To limit the influence of outliers, we have excluded trials where the mean error is larger than $5D$ (10 trials in the whole dataset). Figure 7 shows the results. Comparing to the ray intersection method, our proposed algorithm achieved better results for the close and far targets (error differences between the two methods are $0.12D$ and $0.34D$, $t(2152) = -11.8; p \ll 0.01$ and $t(2183) = -40.5; p \ll 0.01$ respectively, t-test) and worse results for the middle target (error difference of $0.15D$, $t(2195) = 50.1; p \ll 0.01$, t-test). Comparing to the ray intersection method, the proposed probability model seems to benefit from incorporating the noise and ambiguity into the computation and provides an advantage for gaze estimation in 3D.

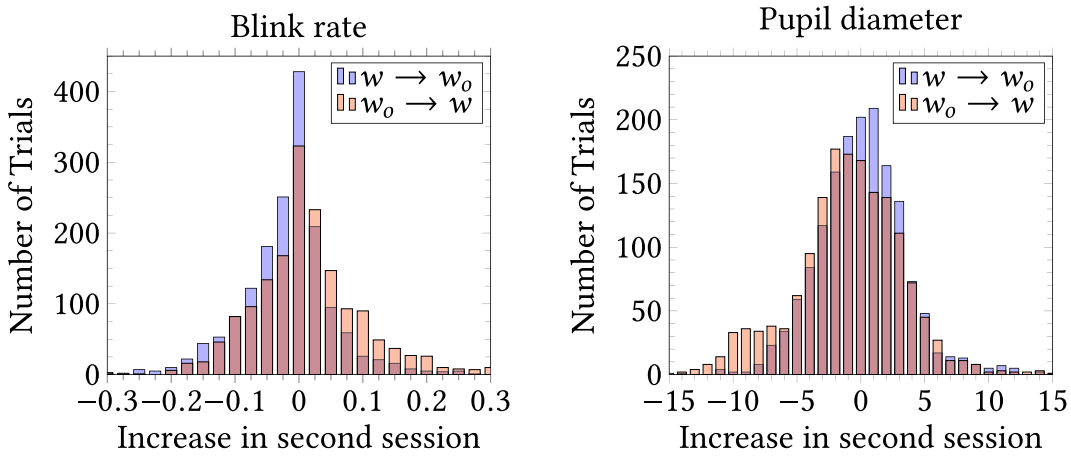


Fig. 8. Histograms of changes in blink rate (left) and pupil diameter (right). We distinguish between two groups depending on the session order (see the legends). w corresponds to the sessions when dynamical adjustment was enabled, and w_o corresponds to the sessions of no adjustment.

5.2 Objective Evaluation of Visual Comfort

Two participants' data were excluded in the following analysis due to large errors in the estimation of fixation targets (based on the average error). Nearly all trials were completed correctly (accuracy = 98.5%, mismatching symbols were selected in only 96 of 6, 681 trials). The average reaction time of one trial (i.e., the trial completion time), when dynamic adjustment was enabled (mean duration = 3.33s, SD = 2.09s), does not differ significantly from trials with fixed camera parameters (mean duration = 3.37s, SD = 2.21s). Similar reaction time in both conditions indicates that the introduced camera adjustment does not significantly influence depth perception, at least not from the perspective of task completion time. Based on participants' self-evaluation, camera adjustment is commonly not noticed, and no noticeable difference between the two sessions was reported.

We observed a large variation among the participants, as indicated by the standard deviations of the reaction time. Therefore, we only performed inter-participant comparison of the eye movements data. We computed the trialwise differences of eye movement statistics for each dataset, and differentiated between the two groups of different session orders (one group started with the adjustment session and the other group started with the session using fixed camera parameters).

For each trial, we computed the average blink rate and pupil diameter, and the differences of two corresponding trials between two sessions. More precisely, the trialwise difference equaled to the increase of the second session from the first one. Positive number corresponded to an increase in the second session, and negative number corresponds to a decrease. Figure 8 shows the histograms of changes in blink rate and pupil diameter. On average, dynamic adjustment led to higher blink rate and smaller pupil diameter. Both are a sign of fatigue. The differences in the resulting distributions are both significant ($t(2907) = -12.21; p \ll 0.01$ and $t(2907) = 4.08; p \ll 0.01$ respectively, t-test). However, we suspect that dynamic changes in the scene may contribute to the increased blink rate; however, this needs to be confirmed in future study. Increase of pupil diameter variance does not differ significantly no matter what the session order was ($t(2907) = -0.83; p = 0.40$, t-test). The literature, where pupil diameter variance was used as a measure of fatigue, were mainly focused on sleep. The level of fatigue after a long time of being awake (Morad et al. 2000a) or shortly before falling asleep (Schumann et al. 2017) leads to a significant different profile of pupil diameter variance. In comparison, pupil diameter variance may not be a valid measure for visual comfort.

Previous studies showed that humans can tolerate certain inconsistency between vergence and accommodation, and there exists a comfort zone for stereoscopic viewing given a fixed accommodation level (Lang et al. 2010; Mendiburu 2012). However, the exact size of the comfortable zone depends on many factors, such as the viewing distance, the illumination, as well as scene content (Shibata et al. 2011). For instance, when viewing objects presented on a screen, the comfort zone in front is considerably larger than the area behind the screen. Depending on the viewing distance t, we consider the region between $33\%t$ in front and $25\%t$ behind as the comfort zone and compute the percentage of time when vergence and accommodation are consistent (i.e., within the comfort zone). On average, vergence and accommodation are consistent for 43% of the time when dynamic adjustments were enabled (67% ± 17% for close targets; 20% ± 12% for middle targets; 42% ± 26% for far targets), whereas only 33% of time when camera parameters were fixed (100% for middle targets and 0% for near and far targets). Note that the amount of time required for the camera adjustment should been considered in such evaluation, as we made a tradeoff between the adjustment speed and noticeable visual changes. In an ideal eye tracking scenario, the in-comfort percentage of time with dynamic adjustment is only 82% (21.7s adjusting time for each block on average; 83% for close targets, 85% for middle targets, and 76% for far targets). While the average consistency is low for the middle area, for some participants it was much better (e.g., the mean in-comfort-percentage of the top four is 38%). Also in these cases there is no correlation between objective measures of comfort and consistency, but an inconsistency was revealed by the objective measures. Compared to the fixed camera parameter configuration, dynamic adjustment leads to significant increases in both blink rate (more fatigue, $t(830) = -10.54; p \ll 0.01$, t-test) and pupil diameter (less fatigue, ($t(830) = -3.16; p = 0.0018$), t-test).

5.3 Subjective Evaluation of Visual Comfort

Similarly, for each participant's subjective evaluation, we reported the difference between the first and second evaluations. No headache was reported regardless of the session order, but the reported eye tiredness indicates a clear preference for the session with fixed camera parameters (see Figure 9(a)). The same trend was observed from the responses to the direct comparisons of two sessions (see Figure 9(b)–(d)). No matter whether it was about fatigue, eye irritation, or depth changes, the majority of participants preferred the session with fixed camera parameters over the session with dynamic adjusted cameras.

5.4 Discussion

Our result goes in line with the previous finding (Koulieris et al. 2017) that improvements in visual comfort achieved by algorithmic solutions are limited and it is difficult to effectively reduce the vergence-accommodation conflict unless physical changes of accommodation is involved.

In the parameter space of two adjustable variables, we only considered a subspace by following a simple protocol. It remains unclear how other combinations of adjustments, for example, by allowing camera translation and rotation at the same time, would affect visual comfort. How to find an effective protocol is also an interesting question.

One of our design goals was to minimize the changes while adjusting the cameras such that observers do not notice any differences or artifacts. It has to be balanced with the adjustment speed. In the experiment, when gaze point changes from the close distance to the middle depth, it took 8.6s until cameras reached the stable configuration; when changing from middle to far-away distance, it took 2.4s. These relative slow adjustments result in no noticeable changes based on participants report; however, it may, on the other hand, induce fatigue. Future studies are required to investigate this further.

Even though our experiment requires eye movement changes among three different depths, visual comfort is mainly evaluated in a static scene once the eyes have been adjusted to the targets. It is possible that the dynamic adjustment method could reduce the vergence-accommodation conflict more effectively for a dynamic scene, where fixation targets continuously move in depth, for instance.

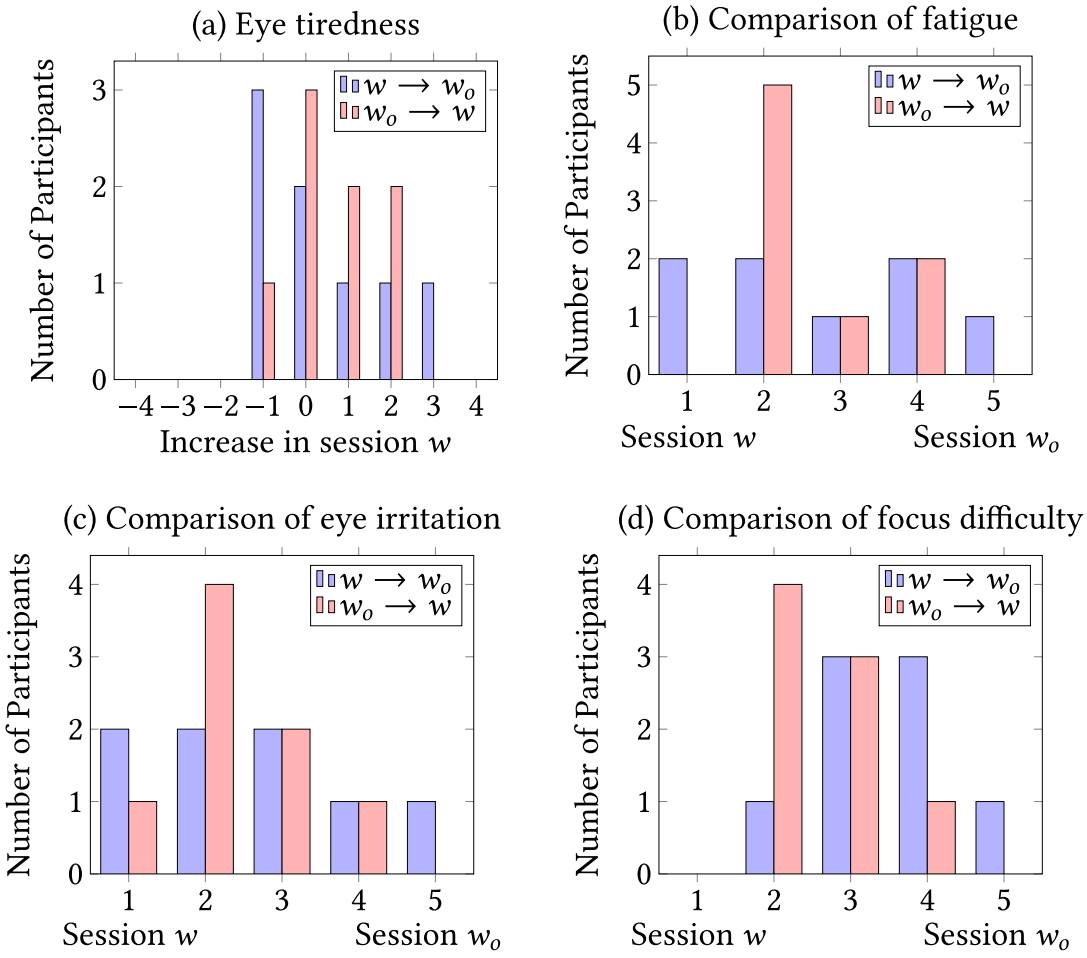


Fig. 9. Subjective responses. (a) Increase in eye tiredness. In the 5-point rating scale, 1 corresponds to fresh and 5 corresponds to irritated. Therefore, positive values correspond to the amount of increased eye tiredness in the second session, and negative rates indicate the amount of decrease in eye tiredness. Plots in (b)–(d) show the responses when participants were asked to evaluate (b) which session was more fatigue, (c) which session was more irritating to the eyes, and (d) which session was easier to change in depths. w corresponds to the sessions when dynamical adjustment is enabled and w_o corresponds to the sessions of no adjustment.

Essentially, we evaluated visual comfort in sessions of 20 minutes. Very likely fatigue gets stronger over time. It is not clear how dynamic adjustments would affect results over longer periods of time. Additionally, we want to point out that the accuracy of eye tracking plays an important role in such experiments. Even though the proposed probability-based gaze estimation method achieves better results for targets that are close or far away, the accuracy for target objects at the screen distance drops down, which leads to a large inconsistency between vergence and accommodation when viewing such objects.

6 CONCLUSION

We have implemented and evaluated a new way to resolve a prolonged vergence-accommodation conflict in HMDs. The approach could be considered as having the potential to reduce the visual discomfort associated with HMDs; however, our experiments indicate that this is not the case. At least in the tested scenario, both

subjective and objective evaluations suggest to keep camera parameters fixed. This is a useful result, because the dynamic adjustment requires additional equipment and processing. However, many factors could influence the results, and further studies are required to understand how well these findings can be generalized.

The camera adjustment naturally provides a two-parameter space for the adjustment. In addition, the speed of adjustment could be varied within the limits of the processing time of eye tracking equipment. We have opted for the adjustment that we believe interferes as little as possible with human perception. It may be fruitful in the future to experiment with other protocols for adjustment.

Another question arising from the experiment is to further look into what exactly causes discomfort. The vergence-accommodation conflict may contribute to discomfort, but perhaps other causes are so far underestimated.

ACKNOWLEDGMENTS

We thank all participants who joined our experiment. We also thank Minjung Kim, Andreas Ley, Ronny Hänsch, and Amelie Froessl, who joined our pilot study and gave us valuable feedback.

REFERENCES

David A. Atchison, George Smith, and George Smith. 2000. Optics of the human eye. Butterworth-Heinemann Oxford.

Genís Cardona, Carles Garcá, Carme Serés, Meritxell Vilaseca, and Joan Gispets. 2011. Blink rate, blink amplitude, and tear film integrity during dynamic visual display terminal tasks. *Curr. Eye Res.* 36, 3 (2011), 190–197. DOI: https://doi.org/10.3109/02713683.2010.544442 PMID: 21275516.

Wei Chen, Jérôme Fournier, Marcus Barkowsky, and Patrick Le Callet. 2011. New stereoscopic video shooting rule based on stereoscopic distortion parameters and comfortable viewing zone. In *IS&T/SPIE Electronic Imaging*, Andrew J. Woods, Nicolas S. Holliman, and Neil A. Dodgson (Eds.), Vol. 7863. 78631O. DOI: https://doi.org/10.1117/12.872332

Andrew T. Duchowski, Donald H. House, Jordan Gestring, Rui I. Wang, Krzysztof Krejtz, Izabela Krejtz, Radosław Mantiuk, and Bartosz Bazyluk. 2014. Reducing visual discomfort of 3D stereoscopic displays with gaze-contingent depth-of-field. In *Proceedings of the ACM Symposium on Applied Perception (SAP'14)*. ACM, New York, NY, 39–46. DOI: https://doi.org/10.1145/2628257.2628259

Martin Fisker, Kristoffer Gram, Kasper Kronborg Thomsen, Dimitra Vasilarou, and Martin Kraus. 2013. Automatic convergence adjustment for stereoscopy using eye tracking. *Eurographics*.

Stephan J. Garbin, Yiru Shen, Immo Schuetz, Robert Cavin, Gregory Hughes, and Sachin S. Talathi. 2019. OpenEDS: Open eye dataset. *CoRR* abs/1905.03702 (2019). arXiv:1905.03702 http://arxiv.org/abs/1905.03702

David M. Hoffman, Ahna R. Girshick, Kurt Akeley, and Martin S. Banks. 2008. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *J. Vis.* 8, 3 (03 2008), 33–33. DOI: https://doi.org/10.1167/8.3.33

Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press, Oxford.

Iana Iatsun, Mohamed-Chaker Larabi, and Christine Fernandez-Maloigne. 2013. Investigation of visual fatigue/discomfort generated by S3D video using eye-tracking data. In *Stereoscopic Displays and Applications XXIV*, Vol. 8648. International Society for Optics and Photonics, 864803.

Paul V. Johnson, Jared A. Q. Parnell, Joohwan Kim, Christopher D. Saunter, Gordon D. Love, and Martin S. Banks. 2016. Dynamic lens and monovision 3D displays to improve viewer comfort. *Opt. Expr.* 24, 11 (May 2016), 11808–11827. DOI: https://doi.org/10.1364/OE.24.011808

Petr Kellnhofer, Piotr Didyk, Karol Myszkowski, Mohamed M. Hefeeda, Hans-Peter Seidel, and Wojciech Matusik. 2016. GazeStereo3D: Seamless disparity manipulations. *ACM Trans. Graph.* 35, 4 (10 2016), 68:1–68:13. DOI: https://doi.org/10.1145/2897824.2925866

Donghyun Kim, Sunghwan Choi, Sangil Park, and Kwanghoon Sohn. 2011. Stereoscopic visual fatigue measurement based on fusional response curve and eye-blinks. In *Proceedings of the 2011 17th International Conference on Digital Signal Processing (DSP'11)*. 1–6. DOI: https://doi.org/10.1109/ICDSP.2011.6004999

Robert Konrad, Emily A. Cooper, and Gordon Wetzstein. 2016. Novel optical configurations for virtual reality: Evaluating user preference and performance with focus-tunable and monovision near-eye displays. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI'16)*. ACM, New York, NY, 1211–1220. DOI: https://doi.org/10.1145/2858036.2858140

Frank L. Kooi and Alexander Toet. 2004. Visual comfort of binocular and 3D displays. *Displays* 25, 2-3 (2004), 99–108.

George-Alex Koulieris, Bee Bui, Martin S. Banks, and George Drettakis. 2017. Accommodation and comfort in head-mounted displays. *ACM Trans. Graph.* 36, 4, Article 87 (Jul. 2017), 11 pages. DOI: https://doi.org/10.1145/3072959.3073622

George Alex Koulieris, George Drettakis, Douglas Cunningham, and Katerina Mania. 2016. Gaze prediction using machine learning for dynamic stereo manipulation in games. In *Proceedings of the 2016 IEEE Conference on Virtual Reality (VR'16)*. 113–120. DOI: https://doi.org/10.1109/VR.2016.7504694

Manuel Lang, Alexander Hornung, Oliver Wang, Steven Poulakos, Aljoscha Smolic, and Markus Gross. 2010. Nonlinear disparity mapping for stereoscopic 3D. *ACM Trans. Graph.* 29, 4, Article 75 (Jul. 2010), 10 pages. DOI: https://doi.org/10.1145/1778765.1778812

Thurmon E. Lockhart and Wen Shi. 2010a. Effects of age on dynamic accommodation. *Ergonomics* 53, 7 (2010), 892–903. DOI: https://doi.org/10.1080/00140139.2010.489968 PMID: 20582770.

Thurmon E. Lockhart and Wen Shi. 2010b. Effects of age on dynamic accommodation. *Ergonomics* 53, 7 (10 2010), 892–903. DOI: https://doi.org/10.1080/00140139.2010.489968

Holger Luedtke, Barbara Wilhelm, Martin Adler, Frank Schaeffel, and Helmut Wilhelm. 1998. Mathematical procedures in data recording and processing of pupillary fatigue waves. *Vis. Res.* 38, 19 (1998), 2889–2896. DOI: https://doi.org/10.1016/S0042-6989(98)00081-9

Andrew Maimone, Gordon Wetzstein, Matthew Hirsch, Douglas Lanman, Ramesh Raskar, and Henry Fuchs. 2013. Focus 3D: Compressive accommodation display. *ACM Trans. Graph.* 32, 5, Article 153 (Oct. 2013), 13 pages. DOI: https://doi.org/10.1145/2503144

Bernard Mendiburu. 2012. *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Routledge.

Yair Morad, Hadas Lemberg, Nehemiah Yofe, and Yaron Dagan. 2000a. Pupillography as an objective indicator of fatigue. *Curr. Eye Res.* 21, 1 (2000), 535–542. DOI: https://doi.org/10.1076/0271-3683(200007)2111-ZFT535 PMID: 11035533.

Yair Morad, Hadas Lemberg, Nehemiah Yofe, and Yaron Dagan. 2000b. Pupillography as an objective indicator of fatigue. *Curr. Eye Res.* 21, 1 (2000), 535–542.

Thomas Oskam, Alexander Hornung, Huw Bowles, Kenny Mitchell, and Markus Gross. 2011. OSCAM—optimized stereoscopic camera control for interactive 3D. *ACM Trans. Graph.* 30, 6, Article 189 (Dec. 2011), 8 pages. DOI: https://doi.org/10.1145/2070781.2024223

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. *Proc. Natl. Acad. Sci. U.S.A.* 114, 9 (2017), 2183–2188. DOI: https://doi.org/10.1073/pnas.1617251114

Eli Peli, Reed Hedges, Jinshan Tang, Dan Landmann, T Reed Hedges, Jinshan Tang, and Dan Landmann. 2001. A binocular stereoscopic display system with coupled convergence and accommodation demands. *SID Symp. Dig. Techn. Pap.* 32, 1 (10 2001), 1296–1299. DOI: https://doi.org/10.1889/1.1831799

Clifton M. Schor, Lori A. Lott, David Pope, and Andrew D Graham. 1999. Saccades reduce latency and increase velocity of ocular accommodation. *Vis. Res.* 39, 22 (10 1999), 3769–3795. DOI: https://doi.org/10.1016/S0042-6989(99)00094-2

Andy Schumann, Juliane Ebel, and Karl-Jürgen Bär. 2017. Forecasting transient sleep episodes by pupil size variability. *Curr. Direct. Biomed. Eng.* 3, 2 (2017), 583–586. DOI: https://doi.org/10.1515/cdbme-2017-0121

Andrei Sherstyuk, Arindam Dey, Christian Sandor, and Andrei State. 2012. Dynamic eye convergence for head-mounted displays improves user performance in virtual environments. In *Proceedings of the Symposium on Interactive 3D Graphics and Games (I3D'12)*. ACM, 23–30. DOI: https://doi.org/10.1145/2159616.2159620

Takashi Shibata, Joohwan Kim, David M. Hoffman, and Martin S. Banks. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *J. Vis.* 11, 8 (07 2011), 11–11. DOI: https://doi.org/10.1167/11.8.11

Shinichi Shiwa, Katsuyuki Omura, and Fumio Kishino. 1996. Proposal for a 3-D display with accommodative compensation: 3DDAC. *J. Soc. Inf. Displ.* 4, 4 (10 1996), 255–261. DOI: https://doi.org/10.1889/1.1987395

Lew B Stelmach, Wa James Tam, Filippo Speranza, Ronald Renaud, and Taali Martin. 2003. Improving the visual comfort of stereoscopic images. In *Stereoscopic Displays and Virtual Reality Systems X*, Vol. 5006. International Society for Optics and Photonics, 269–283. DOI: https://doi.org/10.1117/12.474093

Wa James Tam, Filippo Speranza, Sumio Yano, Koichi Shimono, and Hiroshi Ono. 2011. Stereoscopic 3D-TV: Visual comfort. *IEEE Trans. Broadcast.* 57, 2 (6 2011), 335–346. DOI: https://doi.org/10.1109/TBC.2011.2125070

Kasim Terzić and Miles Hansard. 2016. Methods for reducing visual discomfort in stereoscopic 3D: A review. *Sign. Process.: Image Commun.* 47 (2016), 402–416. DOI: https://doi.org/10.1016/j.image.2016.08.002

Xi Wang, David Lindlbauer, Christian Lessig, and Marc Alexa. 2017. Accuracy of monocular gaze tracking on 3D geometry. In *Eye Tracking and Visualization*, Michael Burch, Lewis Chuang, Brian Fisher, Albrecht Schmidt, and Daniel Weiskopf (Eds.). Springer International Publishing, Cham, 169–184.

Andrew J. Woods, Tom Docherty, and Rolf Koch. 1993. Image distortions in stereoscopic video systems. In *Stereoscopic Displays and Applications IV*, Vol. 1915. International Society for Optics and Photonics, 36–49. DOI: https://doi.org/10.1117/12.157041

Yasunori Yamada and Masatomo Kobayashi. 2017. Fatigue detection model for older adults using eye-tracking data gathered while watching video: Evaluation against diverse fatiguing tasks. In *Proceedings of the 2017 IEEE International Conference on Healthcare Informatics (ICHI'17)*. 275–284. DOI: https://doi.org/10.1109/ICHI.2017.74

Yasunori Yamada and Masatomo Kobayashi. 2018. Detecting mental fatigue from eye-tracking data gathered while watching video: Evaluation in younger and older adults. *Artif. Intell. Med.* 91 (2018), 39–48. DOI: https://doi.org/10.1016/j.artmed.2018.06.005

John M. Zelle and Charles Figura. 2004. Simple, low-cost stereographics: VR for everyone. *SIGCSE Bull.* 36, 1 (Mar. 2004), 348–352. DOI: https://doi.org/10.1145/1028174.971421

Sanyuan Zhao and Tingzhi Shen. 2010. Driver fatigue detection based on eye status. In *Proceedings of the 2010 International Conference on Multimedia Technology*. 1–4. DOI: https://doi.org/10.1109/ICMULT.2010.5630864

Received July 2019; accepted July 2019