Reducing Visual Discomfort of 3D Stereoscopic Displays with Gaze-Contingent Depth-Of-Field

Andrew T. Duchowski*, Donald H. House, Jordan Gestring, Rui I. Wang
Clemson University, Clemson, SC, USA

Krzysztof Krejtz[1,2], Izabela Krejtz[2]
[1]Information Processing Institute, Warsaw, Poland
[2]University of Social Sciences and Humanities, Warsaw, Poland

Radosław Mantiuk, Bartosz Bazyluk
West Pomeranian University of Technology, Szczecin, Poland

Abstract

We report on an experiment testing gaze-contingent depth-of-field (DOF) for the reduction of visual discomfort when viewing stereoscopic displays. Subjective results are compelling, showing, we believe for the first time, that gaze-contingent depth-of-field significantly reduces visual discomfort. When individual stereoacuity is taken into account, objective measurements of gaze vergence corroborate previous reports, showing significant bias toward the zero depth plane, where error is smallest. As with earlier similar attempts, participants expressed a dislike toward gaze-contingent DOF. Although not statistically significant, this dislike is likely attributed to the eye tracker's spatial inaccuracy as well as the DOF simulation's noticeable temporal lag.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Display Algorithms

Keywords: 3D displays, eye tracking, visual psychophysics

1 Introduction

Stereoscopic displays dissociate the natural coupling between vergence and accommodation by rendering images with non-zero disparity at a fixed display distance [Wann et al. 1995; Wilson 1996; Rushton and Riddell 1999]. This dissociation—referred to as the *accommodation-vergence conflict*—has been considered to be the primary reason for discomfort (asthenopia) felt by viewers of stereoscopic displays, with its source tied to eye strain and fatigue [Howard and Rogers 2002; Iwasaki et al. 2009]. Okada et al. [2006] provide empirical evidence that the vergence-accommodation conflict affects the accommodative response, and Shibata et al. [2011] make a compelling case for identifying the vergence-accommodation conflict as the source of visual discomfort. For further review, see Lambooij et al. [2007; 2009].

Typical stereo displays fail to simulate accommodative blur, thereby fixing accommodative demand in the presence of depth-variable disparity. MacKenzie et al. [2010] suggest that multiplane, depth-filtered images may reduce many of the problems caused by the accommodation-vergence conflict.

We show that, by manipulating depth-of-field in a gaze-contingent manner, visual discomfort is reduced. Analysis of vergence error is also provided for those with high stereoacuity.

2 Previous Work

Perceptual Depth Cues. Binocular vision is characterized by gaze vergence, either *fusional* or *accommodative*. Retinal disparity, the horizontal difference of the retinal projections of a point at distance, drives fusional vergence [Shakhnovich 1977] while retinal blur drives accommodative (focal) vergence [Büttner-Ennever 1988]. Both types of vergence are known to be tightly coupled in the human visual system [Fincham and Walton 1957].

Both blur and disparity are complementary cues of depth perception. Depth cue complementarity could also be involved in programming of motor behavior such as eye movements and reaching. For visual space in front of and behind fixation, depth from blur is more precise than depth from disparity, and the visual system relies on the more informative cue when both are available; away from fixation, vision resorts to other depth cues including linear and aerial perspective and familiar size [Held et al. 2012].

Accommodative blur drives depth perception, and because the presence of correct defocus diminishes visual fatigue when viewing stereoscopic stimuli, it is important in the perception of stereoscopic displays ([Hoffman et al. 2008; Hoffman and Banks 2010]).

Most 3D displays lack depiction of focus depth cues, accommodation, and retinal blur, but, in some cases, accommodation demands can be met via specialized optical configurations, e.g., autostereoscopic displays [Akeley et al. 2004; Schowengerdt and Seibel 2006], telecentric optics [Shibata et al. 2005], dual-lens volumetric displays [Love et al. 2009; Shibata et al. 2011], or multilayer display architectures [Maimone et al. 2013]. These specialized optical configurations function to create appropriate focus cues: if the viewer accommodates far, distant parts of the scene are sharply focused and near parts are blurred. If the viewer accommodates near, distant parts are blurred and near parts are sharp.

What is not clear, however, is whether such depth filtering can effectively reduce visual discomfort in stereoscopic displays lacking multiple physical focal planes, i.e., stereoscopic displays composed of a single physical focal plane such as a computer monitor, where depth-filtering is achieved via dynamic depth-of-field (DOF, e.g., slaved to the eye-tracked gaze depth point).

Gaze Tracking for Stereoscopic Displays. Some optical displays obviate the need for tracking gaze [Love et al. 2009], while others assume the observer's gaze position [Maimone et al. 2013]. We track gaze and compute real-time vergence response to different target disparities at fixed accommodative (screen) depth.

Peli et al. [2001] suggested dynamically matching the convergence demand of the displayed object by bringing the fixated object to zero disparity, removing the conflict locally. They, however, did not track gaze to evaluate this early work.

Brooker and Sharkey [2001] used a desktop stereographic display to project a virtual scene in which vergence-derived distance controlled a synthetically generated depth-of-field. They suggested that perceptual performance gains may be achieved by addition of the synthetic DOF but the virtual environment was limited to a display of components interconnected by pipes on a distractive background. We test gaze-contingent depth-of-field in four different scenes.

*e-mail: duchowski@clemson.edu

Laure et al. [2012b; 2012a] showed that limiting high frequencies (e.g., blurring) in areas of large disparities significantly reduced degradation in the punctum maximum of accommodation, easing accommodation after immersion in their stereoscopic virtual world. However, participants expressed an apparent significant preference for the undegraded environment (i.e., no blurring), deemed more aesthetic. A head-mounted eye tracker was proposed as a means of not blurring (monoscopically) fixated objects, but only preliminary anecdotal observations were provided.

Maiello et al. [2013] examined how the time-course of binocular fusion depends on depth cues from blur and stereoscopic disparity in natural images when produced from light field photographs of natural scenes taken with a Lytro camera. They implemented a gaze-contingent stereoscopic display with a natural distribution of blurs and disparities across the retina. Results suggest that distributions of retinal blur facilitate depth perception in natural images.

Gaze-driven Defocus. Depth-of-field improves perception of depth [Wang et al. 2011], and, with gaze-contingent control, also increases immersion (in a first-person-player game) [Hillaire et al. 2008]. Effects on visual comfort are less clear.

Masia et al. [2013] categorize computational (including stereoscopic) displays, in terms of the *plenoptic function*, i.e., manipulation of the range of contrast or luminance, color gamut, or angular or spatio-temporal resolution. Although not included in their taxonomy, our gaze-contingent display (GCD) falls in the latter category and is reminiscent of combined optical and temporal superposition schemes. Generally, a GCD maintains the display's high resolution at the point of gaze, while reducing spatial resolution in the periphery [Murphy and Duchowski 2007].

Mantiuk et al. [2011] tested perceived realism of virtual scenes with a gaze-contingent DOF, but did not report significant preference differences between various combination of statically and dynamically presented scenes. We test four different scenes, with and without gaze-contingent DOF. Our gaze-contingent DOF technique differs from that of Mantiuk et al. in one important aspect, namely that their DOF implementation relied on an estimate of the distance between each object in the scene and the lens used in the DOF computation. In our computation, based on a similar reverse-mapped z-buffer technique, we slave the DOF lens directly to gaze depth, i.e., the lens is not dependent on any scene object's depth.

More recently, Bernhard et al. [2014] tested fusion time in gaze-controlled stereoscopic display of random-dot stereograms and found what resembles a bimodal distribution of viewers' response times to the stereoscopically projected stimuli. Their results are generally similar to the results of our stereoacuity test, also conducted with random-dot stereograms: we observed a bimodal distribution of viewers with high and low stereoacuity.

Vinnikov and Allison [2014] tested gaze-contingent DOF over various realistic scenes. They found no impact of different scenes on subjective impressions of DOF and that viewers generally disliked the DOF effect. We show a significant effect of DOF on subjective impressions of visual discomfort, and on an objective measure of vergence error, when taking stereoacuity into account. We also found a similar dislike of DOF, which appears to be an increasingly reported complaint, most likely due to temporal lag of the gaze-contingent effect.

3 Technical Development

Gaze-dependent Depth-Of-Field. Unlike hardware methods addressing variable focus [Liu et al. 2010; MacKenzie et al. 2010], our gaze-contingent approach is software-based. The goal is to maintain whatever is at focal depth in focus, while peripherally blurring what is outside of this depth. Following Mantiuk et al. [2011] we implemented a real-time depth-of-field GPU-based simulation based on the work of Riguer et al. [2004]. Peripheral blur is simulated through estimation of the Circle of Confusion (CoC) radius:

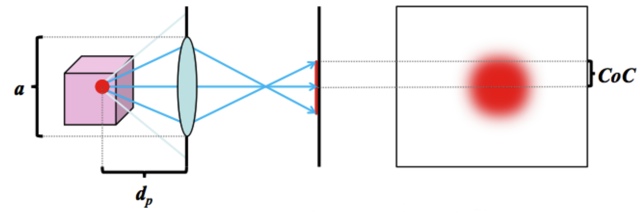


Figure 1: *Depth-of-field rendering.*

$$\text{CoC} = a \cdot \mid \frac{f}{d_0 - f} \mid \cdot \mid 1 - \frac{d_0}{d_p} \mid$$

where $a = 1.0$ is modeled lens aperture diameter, $f = 2.2$ is the lens focal length, d_0 is the distance between the focal plane and the lens (objects in this plane at this distance are in sharp focus), and d_p is the distance from the rendered object to the lens (see Figure 1).

Unlike Mantiuk et al., we do not estimate d_0 as the depth value of the current pixel at the viewer's gaze point, rather, we measure gaze depth z directly, and set the depth-of-field focal plane to this distance. Our approach also differs in that we do not use the Poisson disk to effect the DOF image blur. Instead, we convolve the original image with Gaussian kernels of successively larger sizes to create additional images prior to the final composite.

3D Gaze Depth Estimation. Gaze depth estimation relies on deriving a fixed mapping of 2D coordinates to 3D gaze depth, requiring a 3D calibration procedure. Current binocular eye trackers deliver two eye gaze points, (x_l, y_l) for the left eye and (x_r, y_r) for the right, measured in screen coordinates. The horizontal disparity $\Delta x = x_r - x_l$, between the left and right gaze coordinates, is sufficient to estimate the gaze depth coordinate z. For reviews of 3D gaze estimation see Pfeiffer [2012] and Essig et al. [2012].

The geometry of our approach to gaze depth estimation is derived by Wang et al. [2012]. Assuming constant interocular distance α (in our case set to 6.3 cm, the average for all people regardless of gender), the eyes are assumed to be at the same height $y = 0$, and at equal distances D from the screen. When using spatial triangulation to estimate gaze depth, the two visual axes of the eyes will not necessarily intersect in 3D space. We handle this problem by computing $y_e = (y_l + y_r)/2$, the average of the left and right eye gaze y-coordinates. Given the above geometric assumptions and measurements, and with gaze disparity given by $\Delta x = x_r - x_l$, the disparity induced gaze depth z in our screen centered coordinate frame is given by $z = (\Delta x D) / (\Delta x - \alpha)$.

Gaze depth can be noisy and prone to estimation errors [Duchowski et al. 2011]. Offline filtering followed by a quadratic fit of the recorded data to the depth targets reduces these effects, with a first order online calibration technique sufficient for recovering gaze depth z [Duchowski et al. 2012; Wang et al. 2012]. We follow this approach and also implement the required 3D calibration.

3D calibration is performed by each viewer, following performance of the eye tracker's vendor-supplied 2D calibration. The viewer is then asked to visually pursue a calibration point (a simple sphere) translating through space along a Lissajous-knot path. Collection

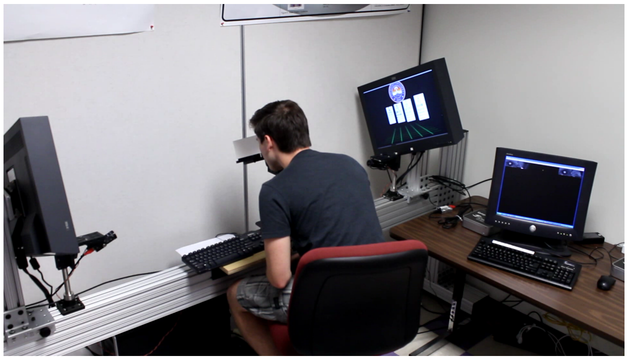


Figure 2: *Wheatstone-style haploscope used in the study.*

of calibration data takes 40 seconds, during which the sphere's position $p(t) = (x(t), y(t), z(t))$ changes with time t in seconds, according to $p(t) = \mathbf{A}\cos(2\pi\,\mathbf{f}\,t + \phi)$, with component amplitude $\mathbf{A} = (9, 5, 20)$ cm, frequencies $\mathbf{f} = (0.101, 0.127, 0.032)$ Hz, and phase angles $\phi = (0°, -90°, 57°)$.

4 Experimental Methodology

Stereoscopic disparity presented without the natural coupling to accommodation disparity can result in visual discomfort ranging from an odd feeling to the irresistible urge to close the eyes [Kooi and Toet 2004]. Discomfort is exacerbated when images are displayed outside the corresponding range of depth of focus (see Yano et al. [2002; 2004]). Tam et al. [2011] note that there are no standard methodologies for the measurement of (subjective) visual comfort for stereoscopic images. Indeed, subjective questionnaires favoring the 5-point Likert scale appear to be the norm [Du et al. 2013].

Our initial goal was to replicate a study by Hoffman et al. [2008], who demonstrated that mismatches in the stimuli to vergence and accommodation cause visual discomfort. With the objective of gauging viewing comfort following similar procedures and subjective impression questionnaires, we ran a pilot study before implementing a number of modifications prior to the main study.

Both pilot and main studies followed a 2×4 within-subjects experimental design with gaze-contingent DOF acting as one fixed factor (at two levels: present or absent) and scene type acting as the second fixed factor (at four levels with four different scenes).

Apparatus. Both pilot and main studies used a Wheatstone-style stereoscope (see Figure 2), consisting of two high-resolution *IBM T221* "Big Bertha" LCD monitors arranged in a haploscope, and driven by two *NVidia Quadro FX 5800* graphics cards installed in an *Intel Xeon E5502 Dualcore* PC running *CentOS*. Both displays are set on a track, sitting 172 cm apart on opposite sides of two small mirrors angled at 45° from the medial axis. The screens are

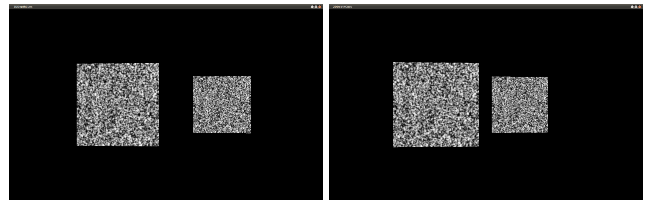


Figure 4: *Random dot stimulus used for pilot pre-screening.*

48 cm wide and 30 cm high (16:10), with a screen resolution of 3840×2400, or 9.2 million pixels, at a viewing distance of 86 cm.

Eye tracking cameras, mounted beneath each monitor, are part of *LC Technologies*' *Eyegaze System* that is used to image the viewer's eyes as seen by the cameras in the mirrors (see Figure 2).

Stimulus and Instructions. In both pilot and main studies, we used four scenes presented stereoscopically (see Figure 3):

- an eye chart scene with four Snellen charts, patterned after the stimulus used by Love et al. [2009], but arranged so that no chart was occluded by another;
- a dragon with five soccer balls, each at different depths, following Duchowski et al.'s [2013] EuroGraphics tutorial;
- a 12×11 grid of spheres, similar to what Duchowski et al. [2012] used, but with an extra row in depth, extending from within the screen at $z = -25$ to $z = 30$ cm in front (the bottom row at $z = 30$ is not visible, resulting in a visible 11×11 grid extending from $z = -25$ to $z = 25$ with the middle row set at $z = 0$ screen depth); and
- a table scene with items (e.g., painting, candlestick, cup, etc.) each at varying depth, similar to scenes used by Mantiuk et al. [2011] and Peli et al. [2001].

Participants (in both pilot and main studies) were instructed to successively gaze at verbally or visually cued scene targets. In the eye chart scene they were asked to read the fourth row of letters on each of the Snellen charts; in the dragon scene they were asked to fixate each of the soccer balls; in the table scene they were asked to fixate the plate, painting, checkerboard, etc. Only the sphere scene contained a visually-cued target: the next sphere to be fixated turned a light pink color. The order of stimulus presentation was counterbalanced following a within-subjects Latin square.

Prior to viewing of the stimulus, a pair of random dot stereograms was shown to participants to ascertain whether they could see which of the two stereograms appeared closer to the viewer (see Figure 4). Results of this stereoacuity pre-screening are discussed below.

Participants. Fifteen participants (11 male and 4 female; aged $\text{M} = 21.53, \text{SD} = 3.89$) took part in the pilot study. Twenty one

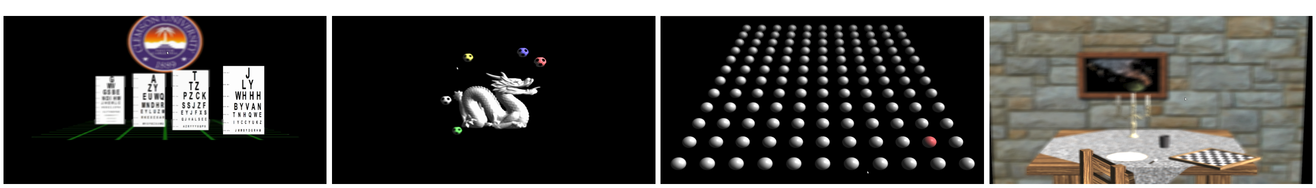


Figure 3: *Graphical stimuli presented stereoscopically with and without gaze-contingent depth-of-field. These images show the effect of DOF with gaze at near distances, e.g., from left to right, looking at the closest eye chart, closest red ball above the dragon's head, pink sphere in the second row, and chair in front of table in the scene with a stone wall at back.*

students took part in the main study, but two were excluded from the analyses due to procedural problems. The final sample included 19 participants (10 female and 9 male) with mean age 24 ($\mathrm{SD} = 5.75$).

For both pilot and main studies, each participant went through two sessions on consecutive days: one made use of gaze-contingent DOF, the other did not. The order of sessions was split evenly among the participants, i.e., half the participants were randomly assigned to start with scene trials with the DOF effect present, and half started with scene trials with the DOF effect absent.

5 Results

We were primarily interested in subjective responses of discomfort when viewing stereoscopically presented scenes. Main study results include analysis of vergence error in two scenes where target depth is guided (spheres) or easily determined (eye charts, where chart depth is determined by ray casting the left and right gaze direction bisector). The main study questionnaire analysis takes into account individual stereoacuity measurements.

5.1 Pilot Study

Replicating Hoffman et al.'s [2008] assessment of subjective visual discomfort, we started with the same two 'symptom' and 'display' questionnaires. The symptom questionnaire contained five questions: (1) How tired are your eyes?; (2) How clear is your vision?; (3) How tired and sore are your neck and back?; (4) How do your eyes feel?; (5) How does your head feel? Participants completed the symptom questionnaire at the end of each session. For each question, participants indicated the severity of their symptoms at that moment on a 5-point Likert scale, e.g., for the fourth question circling 1 if their eyes felt "Very fresh", 2 if "OK", 3 if "Mild ache", 4 if "Moderate ache", and 5 if "Severe strain". Participants completed the display questionnaire after the second session. This questionnaire contained four questions: (1) Which session was most fatiguing?; (2) Which session irritated your eyes the most?; (3) If you felt headache, which session was worse?; (4) Which session did you prefer? A similar 5-point Likert scale was used to collect responses.

Replicating Hoffman et al.'s [2008] analysis, we performed a Wilcoxon signed-rank test on participants' display preferences. Contrary to Hoffman et al., no significant difference between DOF sessions (present or absent) was found for any of the measures, i.e., fatigue ($\mathrm{V} = 32, p = .6$), irritation ($\mathrm{V} = 21.5, p = .32$), headache ($\mathrm{V} = 1.5, p = .13$), or preference ($\mathrm{V} = 28.5, p = .71$).

The symptom questionnaire was analyzed in the same way with the Wilcoxon signed-rank test, revealing a significant difference in vision clarity between sessions (DOF present or absent), $\mathrm{V} = 127.5, p < .01$, suggesting that participants reported greater blurred vision with DOF present ($\mathrm{M} = 2.02, \mathrm{SD} = .93$) than with DOF absent ($\mathrm{M} = 1.67, \mathrm{SD} = .95$). A significant difference between sessions was also found on response to "How tired and sore are your neck and back?", $\mathrm{V} = 210.5, p < .05$. With no DOF, participants reported stronger tiredness in the neck ($\mathrm{M} = 1.92, \mathrm{SD} = .93$) than with DOF ($\mathrm{M} = 1.65, \mathrm{SD} = .82$). Analyses for the remaining three questions revealed no significant differences between sessions.

Discussion. Results of the display questionnaire showed no significant effects of DOF on perceived visual discomfort. Based on participants' informal debriefing comments, we were dismayed to hear that not everyone was able to fully see the depth of the stereoscopically shown objects. We surmised that our stereoacuity prescreening stimulus, in which all participants correctly identified the closer of the two random dot stereograms, was too easily discerned due to the large distance (in depth) between the pair.

We concluded that stereoacuity may be an important factor which we had not accounted for. Indeed, the quality of a person's binocular vision affects the binocular viewing comfort to a limited extent [Kooi and Toet 2004]. Binocular misalignment and excessive stereoscopic disparity are less troublesome for people with reduced (binocular) vision. In other words, those with better stereoacuity are likely to find the vergence-accommodation conflict more troubling. For this reason, the primary alteration we made to the experimental procedure before the main study was to revise the stereoacuity test to ascertain whether our gaze-contingent depth-of-field display could reduce discomfort for those with better stereoacuity.

Furthermore, the only meaningful significant difference between the pilot study sessions found was regarding the experience of blurred vision due to gaze-contingent depth-of-field. Contrary to Hoffman et al., we found no display preference differences.

One of the reasons for the lack of significant differences is the limited scale of the questionnaire responses. A 5-point Likert scale was used with left- and right-hand side responses indicating preference toward absence or presence of gaze-contingent DOF, respectively. However, such limited scales are rarely used for distinguishing differences between two phenomena. An 11-point Thurstone [1928] scale is more sensitive for properly distinguishing human attitudes.

Besides the limited scale, another, potentially greater, problem with Hoffman et al.'s response instrumentation is the assumption that people reporting following two experimental sessions one day apart are able to precisely distinguish between them, especially along several detailed dimensions. This assumes that the experience of vision blur or eye irritation during the first session is very strong or meets some other criteria that allows its consolidation into long-term memory (see Atkinson and Shiffrin [1968]).

For these reasons, we decided on a secondary alteration of the experimental procedure prior to the main study: to redesign the subjective symptom questionnaire from a direct comparison of both sessions to a subjective evaluation of each session immediately following its completion, and to expand the subjective self-reporting scale from a 5-point discrete scale to a 100-point continuous scale. The goal was to use a more sensitive scale for capturing the difference in subjective impressions of gaze-contingent depth-of-field.

5.2 Main Study

In the main study each participant still went through two sessions on consecutive days, just as in the pilot study, but our revised questionnaire posed the following 5 questions after each session: (1) My eyes are irritated.; (2) I have a headache.; (3) I feel fatigued.; (4) Overall, how did you like the session?; (5) Would you like to see more pictures like in this session? Participants used an online continuous slider to indicate their response on a 0-100 scale where 0 meant "not at all" and 100 meant "very much".

Following a revised stereoacuity test, the rest of the procedures remained the same as in the pilot study.

Revised Stereoacuity Test. Inspired by Burge et al.'s [2005] two-interval forced-choice (2IFC) for estimating subjective equality, similar to the two-alternative forced choice (2AFC) test used by Didyk et al. [2011] to estimate just noticeable differences of disparity perception, we constructed a three-alternative forced choice (3AFC) test allowing participants to adjust the depth distance between the two random dot stereograms. Both were displayed in front of a third random dot field to make discerning the distance

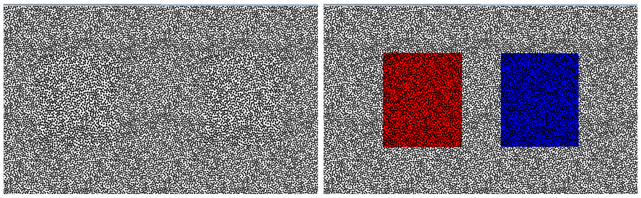


Figure 5: *Random dot stimulus used for stereoacuity test (left), visualization for purposes of illustration showing location of floating stereograms (right).*

between the two "floating" stereograms less obvious (see Figure 5). The stereograms' dots were scaled to eliminate size cues.

The two stereograms started at a depth distance of 30 cm apart, with the closer (left or right) randomized. Participants indicated which was closer by making one of three choices: the left, right, or down arrow key, indicating left, right, or neither (same depth), respectively. A correct response reduced the depth distance between the two stereograms, with the closer of the two once again randomized. Two successive incorrect or same depth responses terminated the test. If the participant could not discern which of the two stereograms was closer at 30 cm during the first session, the second session started by showing the stereograms apart at 15 cm (the reasoning being that perhaps one of the two stereograms was too far out in front of or too far behind the screen plane). The lowest relative depth distance was recorded as the viewer's stereoacuity score.

Stereoacuity Results. Stereoacuity is treated as a between-subject factor at two levels (high vs. low), representing the best result out of two measures taken before each session. Small stereoacuity scores close to 0 indicate that participants could reduce the visual depth distance between the two random dot stereograms to near equivalence. Large scores suggest they could not estimate the relative depth difference between the two stereograms.

Stereoacuity distribution is bimodal with highest score frequencies below 1 and equal to 15 cm (see Figure 6). We decided to treat scores as a two-level factor following a median split. The median score (0.47 cm) reflects the threshold distance between the two stereograms on which we split participant responses (scores ≤ 0.47 indicate high stereoacuity, scores > 0.47 indicate low stereoacuity). Of the 19 participants in the study, 9 scored high and 10 scored low.

Visual Discomfort. After each session participants answered questions about their impression of fatigue, headache, preference, eye irritation, and desire to see more of similar scenes.

Initial examination of responses via boxplots revealed a number of outliers which were thought to bias the differences of means between DOF sessions. We replaced outliers[1] with median values of the given variables—this does not magnify the mean response difference (see Aguinis et al. [2013]). Four cases were so replaced for the fatigue response, four for the headache response, and two for the indicated desire of seeing more similar scenes.

A 2×2 mixed-design analysis of variance (ANOVA) was used to gauge the impact of DOF on fatigue, with experimental session as a within-subjects factor (DOF vs. no DOF) and stereoacuity as a between-subjects factor (low vs. high). The dependent variable was fatigue, self-reported by participants after each session. ANOVA

[1] Outliers are defined above the 3^{rd} or below the 1^{st} quartile $\pm$ 1.5 $\times$ the inner quartile range, respectively, as per R's boxplot function.

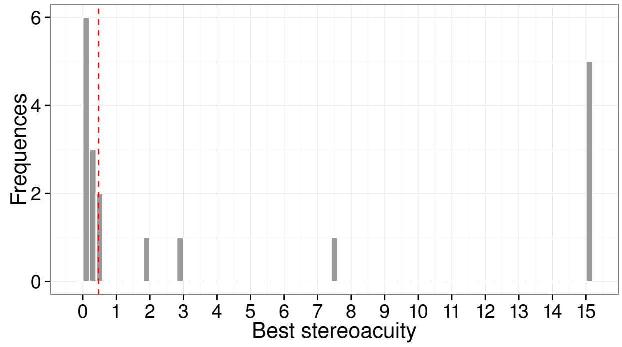


Figure 6: *Frequency distribution of best stereoacuity scores, with dotted line indicating median.*

was calculated with Type-II sum of squares, revealing a statistically significant main effect of experimental session, $F(1,17) = 6.34, p < .05, \eta^2 = .185$ (see Figure 7(a)). Participants after the DOF session reported lower fatigue ($\text{M} = 5.95, \text{SE} = 1.71$) than after the session without DOF ($\text{M} = 18.53, \text{SE} = 4.13$). Analysis also showed that neither main effect of stereoacuity ($F(1,17) < 1$) nor interaction ($F(1,17) < 1$) were statistically significant.

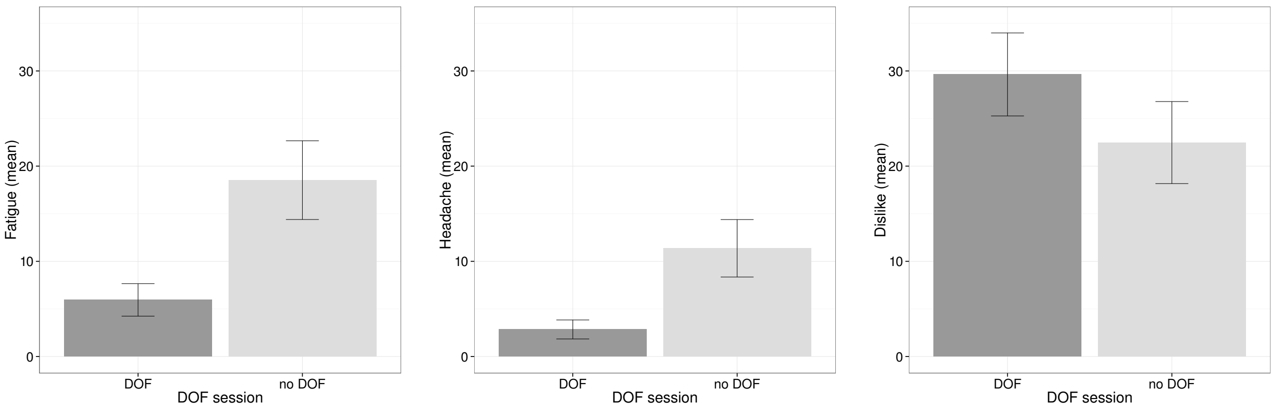
A 2×2 mixed-design ANOVA was performed with impression of headache treated as a dependent variable. As expected, results showed a statistically significant main effect of experimental session, $F(1,17) = 7.61, p < .02, \eta^2 = .168$ (see Figure 7(b)). Participants after the DOF session self-reported lower values for headache ($\text{M} = 2.84, \text{SE} = 1.71$) than after the session with DOF absent ($\text{M} = 18.02, \text{SE} = 4.13$). There was no significant main effect of stereoacuity ($F(1,17) < 1$) nor interaction ($F(1,17) < 1$).

The same type of analysis was repeated on other self-reported symptoms: eye irritation, preference which we report as dislike (subtracting original preference scores from 100), and desire of seeing more similar scenes. Only preference responses showed a weak main effect of experimental session at a statistical tendency level, $F(1,17) = 3.06, p = .098, \eta^2 = .039$. As seen in Figure 7(c), participants tended to dislike the presence of DOF ($\text{M} = 29.63, \text{SE} = 4.36$) rather than its absence ($\text{M} = 22.47, \text{SE} = 4.31$). However, neither main effect of stereoacuity ($F(1,17) = 1.20, p > .1$) nor interaction ($F(1,17) = 1.04, p > .1$) was significant.

Analyses of variance of eye irritation and desire to see similar scenes revealed no significant effects.

Vergence Error. To explore the potential cause for subjective assessment of reduced visual discomfort, we compared mean vergence error when fixating targets at differing depths in the presence or absence of gaze-contingent DOF. Because those with poor stereoacuity may not have resolved stereo disparity, for analysis of vergence error we considered only those with high stereoacuity scores. ANOVA of vergence error (mean signed distance between target and smoothed gaze depth) was performed on data from the sphere grid and eye charts stimuli, since these scenes facilitated estimation of target depth.

Spheres were fixated at five depths: -20, -10, 0, 10, and 20 cm (negative depths behind the screen). A 2×5 (experimental session $\times$ depth) within-subjects ANOVA was performed with vergence error as the dependent variable, revealing a statistically significant main effect of depth, $F(4,32) = 49.13, p < .001, \eta^2 = 0.802$ (see Figure 8). Pairwise comparisons with Bonferroni correction revealed that there were no significant differences between neigh-



(a) Mean responses concerning fatigue. *(b) Mean responses concerning headache.* *(c) Mean responses concerning preference.*

Figure 7: *Mean subjective responses of all viewers, regardless of stimulus or stereoacuity score.*

boring depths (−20 vs. −10, −10 vs. 0, 0 vs. 10, 10 vs. 20) but all other differences are statistically significant ($p < .01$). The analysis also revealed a significant interaction effect between experimental session and depth, $F(4, 32) = 4.11, p < .01, \eta^2 = 0.067$. The main effect of DOF (presence or absence) was not significant.

A similar ANOVA was performed for the eye chart stimulus, where targets were presented at four depths: -15, -5, 5, and 15 cm. Analysis revealed a main effect of depth, $F(3, 20) = 38.43, p < .001, \eta^2 = 0.715$ (see Figure 9). Similar to the previous analysis, pairwise comparisons with Bonferroni correction showed that vergence error did not differ significantly between -15 vs. -5 cm nor between -5 vs. 5 cm. All other differences were significant ($p < .05$). Additionally, the analysis revealed a main effect of experimental session (DOF), $F(1, 20) = 14.16, p < .01, \eta^2 = 0.088$. Participants with DOF present showed lower vergence error ($\mathrm{M} = 2.73, \mathrm{SE} = 1.82$) compared to when DOF was absent ($\mathrm{M} = 4.75, \mathrm{SE} = 1.62$). Interaction between depth and session (DOF) was significant, $F(3, 20) = 3.95, p < .05, \eta^2 = 0.073$.

6 Discussion

Results clearly show that the illusion of depth-of-field reduces visual discomfort (as manifested by subjective impressions of fatigue and headache). As with previous studies [Mantiuk et al. 2011; Leroy et al. 2012b; Vinnikov and Allison 2014], participants expressed a dislike of the defocus blur, albeit at a statistical tendency level. This is likely due to a noticeable lag in the DOF's focus window catching up to the viewer's gaze, an effect of the real-time Butterworth filtering of the gaze depth estimate. We note (anecdotally) that without this temporal smoothing, however, gaze depth is too noisy to provide a pleasing impression of depth defocus. Using Bernhard et al.'s [2014] means of estimation, given our display's refresh rate (30 Hz), the eye tracker's sampling rate (60 Hz), and filter width (3), a lower bound on latency is $33 + 16 + (3 \times 16) = 97$ ms. Display updates as late as 60 ms after eye movement completion do not significantly increase the detectability of image blur and/or motion transients due to the update [Loschky and Wolverton 2007], suggesting that our system is at least 37 ms overbudget.

Does gaze-contingent depth-of-field reduce visual discomfort be-

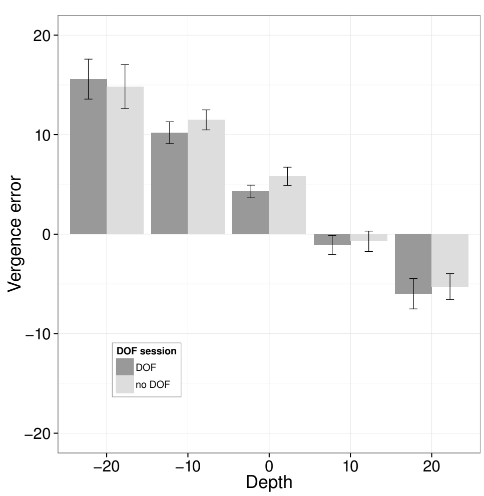


Figure 8: *Mean vergence error of those with high stereoacuity looking at sphere grid (error bars represent standard error values).*

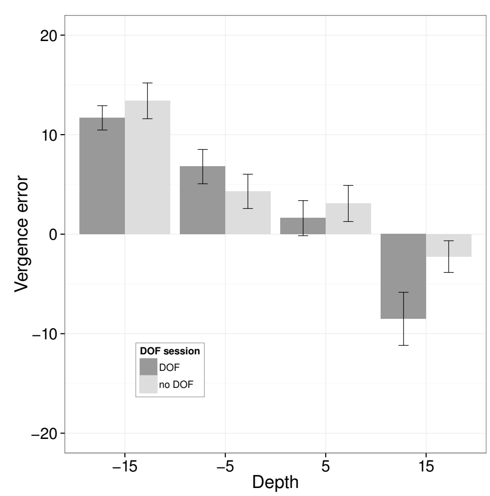


Figure 9: *Mean vergence error of those with high stereoacuity looking at eye charts (error bars represent standard error values).*

cause vergence error to stereoscopically presented targets is reduced? For those with high stereoacuity (able to resolve stereo disparity of two random dot stereograms at about 0.5 cm or less), analysis hints at an effect of depth-of-field on vergence error (on the eye chart scene but not the sphere grid), but this causal effect is not clear. However, Figures 8–9 show a curious trend in vergence error in the presence of gaze-contingent depth-of-field: behind the screen ($z < 0$), vergence error appears reduced, while in front of the screen ($z > 0$) it appears exacerbated. While not statistically significant, these trends motivate further exploration of gaze-contingent depth-of-field.

7 Conclusion & Future Work

Gaze-contingent depth-of-field reduces visual discomfort of stereoscopic scenes for those with high stereoacuity. A statistically weak dislike was expressed for the real-time blurring effect. It is likely that while depth-of-field plays a part in reducing the vergence-accommodation conflict through the illusion of accommodative defocus, to do so more effectively it would need to be better synchronized with binocular gaze, i.e., with improved spatio-temporal fidelity (minimal spatial eye-tracking error and temporal lag).

Although depth-of-field parameters were fixed to test its feasibility, our results suggest inclusion of peripheral defocus blur as an additional parameter in predictive models of disparity perception (e.g., see Didyk et al. [2011; 2012]) or visual comfort [Du et al. 2013]. Future applications benefitting from more comfortable use of stereoscopic displays include games, simulators for surgical training, scientific visualization, and tele-robotic control.

8 Acknowledgments

This work was supported in part by US NSF Award #0722313 and in part by the Polish National Science Centre Decision DEC-2013/09/B/ST6/02270.

References

AGUINIS, H., GOTTFREDSON, R. K., AND JOO, H. 2013. Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. *Organizational Research Methods 16*, 2, 270–301.

AKELEY, K., WATT, S. J., GIRSHICK, A. R., AND BANKS, M. S. 2004. A Stereo Display Prototype with Multiple Focal Distances. *ACM Transactions on Graphics 23*, 3 (August), 804–813.

ATKINSON, R. C., AND SHIFFRIN, R. M. 1968. Human mmeory: A proposed system and its control processes. In *The psychology of learning and motivation: Advances in research theory*. Academic Press, New York, NY, 89–195.

BERNHARD, M., DELL'MOUR, C., HECHER, M., STAVRAKIS, E., AND WIMMER, M. 2014. The Effects of Fast Disparity Adjustment in Gaze-Controlled Stereoscopic Applications. In *ETRA '14: Proceedings of the 2014 Symposium on Eye Tracking Research & Applications*, ACM, New York, NY.

BROOKER, J. P., AND SHARKEY, P. M. 2001. Operator performance evaluation of controlled depth of field in a stereographically displayed virtual environment. In *SPIE 4297: Steroscopic Displays and Virtual Reality Systems VIII*, IS&T/SPIE Electronic Imaging, 408.

BURGE, J., PETERSON, M. A., AND PALMER, S. E. 2005. Ordinal configural cues combine with metric disparity in depth perception. *Journal of Vision 5*, 534–542.

BÜTTNER-ENNEVER, J. A., Ed. 1988. *Neuroanatomy of the Oculomotor System*, vol. II of *Reviews of Oculomotor Research*. Elsevier Press, Amsterdam, Holland.

DIDYK, P., RITSCHEL, T., EISEMANN, E., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2011. A perceptual model for disparity. *ACM Transactions on Graphics 30*, 4 (July), 96:1–96:10.

DIDYK, P., RITSCHEL, T., EISEMANN, E., MYSZKOWSKI, K., SEIDEL, H.-P., AND MATUSIK, W. 2012. A luminance-contrast-aware disparity model and applications. *ACM Transactions on Graphics 31*, 6 (Nov.), 184:1–184:10.

DU, S.-P., MASIA, B., HU, S.-M., AND GUTIERREZ, D. 2013. A metric of visual comfort for stereoscopic motion. *ACM Transaction on Graphics 32*, 6 (Nov.), 222:1–222:9.

DUCHOWSKI, A. T., PELFREY, B., HOUSE, D. H., AND WANG, R. 2011. Measuring Gaze Depth with an Eye Tracker During Stereoscopic Display. In *Applied Perception in Graphics & Visualization (APGV)*, ACM.

DUCHOWSKI, A. T., PELFREY, B., HOUSE, D. H., AND WANG, R. 2012. Eye-Tracked Vergence Response During Active-Stereo Display. In *The 5th Hamlyn Symposium on Medical Robotics*.

DUCHOWSKI, A. T., KREJTZ, K., KREJTZ, I., MANTIUK, R., AND BAZYLUK, B. 2013. *Tut5: An Eye on Perceptual Graphics: Eye-Tracking Methodology*. EuroGraphics, Girona, Spain. EuroGraphics Half Day Tutorial.

ESSIG, K., PRINZHORN, D., MAYCOCK, J., DORNBUSCH, D., RITTER, H., AND SCHACK, T. 2012. Automatic Analysis of 3D Gaze Coordinates on Scene Objects Using Data From Eye Tracking and Motion-Capture Systems. In *ETRA '12: Proceedings of the 2012 Symposium on Eye Tracking Research & Applications*, ACM, New York, NY.

FINCHAM, E. F., AND WALTON, J. 1957. The reciprocal actions of accommodation and vergence. *J. Physiol. Lond. 137*, 488–508.

HELD, R. T., COOPER, E. A., AND BANKS, M. S. 2012. Blur and Disparity Are Complementary Cues to Depth. *Current Biology 22*, 426–431.

HILLAIRE, S., LÉCUYER, A., COZOT, R., AND CASIEZ, G. 2008. Using an Eye-Tracking System to Improve Camera Motions and Depth-of-Field Blur Effects in Virtual Environments. In *Proceedings of the IEEE Virtual Reality Conference*, IEEE, 47–50.

HOFFMAN, D. M., AND BANKS, M. S. 2010. Focus information is used to interpret binocular images. *Journal of Vision 10*, 5, 1–17.

HOFFMAN, D. M., GIRSHICK, A. R., AKELEY, K., AND BANKS, M. S. 2008. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision 8*, 3 (June), 1–47.

HOWARD, I. P., AND ROGERS, B. J. 2002. *Seeing in Depth*, vol. II: Depth Perception. I Porteous, University of Toronto Press, Thornhill, ON, Canada.

IWASAKI, T., KUBOTA, T., AND TAWARA, A. 2009. The tolerance range of binocular disparity on a 3D display based on the physiological characteristics of ocular accommodation. *Displays 30*, 44–48.

KOOI, F. L., AND TOET, A. 2004. Visual Comfort of Binocular and 3-D Displays. In *Encyclopedia of Optical Engineering*, R. G. Driggers, Ed. Marcel Dekker, Inc., New York, NY.

LAMBOOIJ, M. T. M., IJSSELSTEIJN, W. A., AND HEYNDERICKX, I. 2007. Visual Discomfort in Stereoscopic Displays: A Review. In *Stereoscopic Displays and Virtual Reality Systems XIV*, A. J. Woods, N. A. Dodgson, J. O. Merritt, M. T. Bolas, and I. E. McDowall, Eds., SPIE.

LAMBOOIJ, M. T. M., IJSSELSTEIJN, W. A., FORTUIN, M. F., AND HEYNDERICKX, I. 2009. Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. *Journal of Imaging Technology and Science 53*, 3, 1–14.

LEROY, L., FUCHS, P., AND MOREAU, G. 2012. Real-Time Adaptive Blur for Reducing Eye Straing in Stereoscopic Displays. *Transactions on Applied Perception 9*, 2, 9:1–9:18.

LEROY, L., FUCHS, P., AND MOREAU, G. 2012. Visual fatigue reduction for immersive stereoscopic displays by disparity, content and focus-point adapted blur. *IEEE Transactions on Industrial Electronics*, 3998–4004.

LIU, S., HUA, H., AND CHENG, D. 2010. A Novel Prototype for an Optical See-Through Head-Mounted Display with Addressable Focus Cues. *Transactions on Visualization and Computer Graphics 16*, 381–393.

LOSCHKY, L. C., AND WOLVERTON, G. S. 2007. How Late Can You Update Gaze-Contingent Multiresolutional Displays Without Detection? *Transactions on Multimedia Computing, Communications and Applications 3*, 4 (December).

LOVE, G. D., HOFFMAN, D. M., HANDS, P. J. W., GAO, J., KIRBY, A. K., AND BANKS, M. S. 2009. High-speed switchable lens enables the development of a volumetric stereoscopic display. *Optics Express 17*, 18, 15716–15725.

MACKENZIE, K. J., HOFFMAN, D. M., AND WATT, S. J. 2010. Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control. *Journal of Vision 10*, 8, 1–20.

MAIELLO, G., CHESSA, M., SOLARI, F., AND BEX, B. 2013. Stereoscopic fusion with gaze-contingent blur. In *ECVP '13: European Conference on Visual Perception*.

MAIMONE, A., WETZSTEIN, G., HIRSCH, M., LANMAN, D., RASKAR, R., AND FUCHS, H. 2013. Focus 3D: Compressive accommodation display. *ACM Transaction on Graphics 32*, 5 (Oct.), 153:1–153:13.

MANTIUK, R., BAZYLUK, B., AND TOMASZEWSKA, A. 2011. Gaze-Dependent Depth-of-Field Effect Rendering in Virtual Environments. In *SGDA '11: Serious Games Development and Applications*, Springer-Verlag, 1–12.

MASIA, B., WETZSTEIN, G., DIDYK, P., AND GUTIERREZ, D. 2013. A review of computational displays: Pushing the boundaries of optics, computation, and perception. *Computers & Graphics*. In Press.

MURPHY, H., AND DUCHOWSKI, A. T. 2007. Hybrid Image-/Model-Based Gaze-Contingent Rendering. In *Applied Perception in Graphics & Visualization (APGV)*, ACM.

OKADA, Y., UKAI, K., WOLFFSOHN, J. S., GILMARTIN, B., IIJIMA, A., AND BANDO, T. 2006. Target spatial frequency determines the response to conflicting defocus- and convergence-driven accommodative stimuli. *Vision Research 46*, 475–484.

PELI, E., HEDGES, R., TANG, J., AND LANDMANN, D. 2001. A Binocular Stereoscopic Display System with Coupled Convergence and Accommodation Demands. In *SID International Symposium, Digest of Technical Papers*, vol. 53.2, Society for Information Display, 1296–1299.

PFEIFFER, T. 2012. Measuring and visualizing attention in space with 3d attention volumes. In *ETRA '12: Proceedings of the 2012 Symposium on Eye Tracking Research & Applications*, ACM, New York, NY.

RIGUER, G., TATARCHUK, N., AND ISIDORO, J. 2004. Real-Time Depth of Field Simulation. In *ShaderX^2: Shader Programming Tips & Tricks with DirectX 9*, W. F. Engel, Ed. Wordware Publishing, Inc., Plano, TX, 539–556.

RUSHTON, S. K., AND RIDDELL, P. M. 1999. Developing visual systems and exposure to virtual reality and stereo displays: some concerns and speculations about the demands on accomodation and vergence. *Applied Ergonomics 30*, 69–78.

SCHOWENGERDT, B. T., AND SEIBEL, E. J. 2006. True 3D Scanned Voxel Displays Using Single and Multiple Light Sources. *Journal of the Society for Information Display 14*, 2, 135–143.

SHAKHNOVICH, A. R. 1977. *The Brain and Regulation of Eye Movement*. Plenum Press, New York, NY.

SHIBATA, T., KAWAI, T., OHTA, K., MIYAKE, N., YOSHIHARA, Y., AND IWASAKI, T. 2005. Stereoscopic 3-D display with optical correction for the reduction of the discrepancy between accommodation and convergence. *Journal of the Society for Information Display 13*, 8, 665–671.

SHIBATA, T., KIM, J., HOFFMAN, D. M., AND BANKS, M. S. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision 11*, 8, 1–29.

TAM, W. J., SPERANZA, F., YANO, S., SHIMONO, K., AND ONO, H. 2011. Stereoscopic 3D-TV: Visual Comfort. *IEEE Transactions on Broadcasting 57*, 2 (June), 335–346.

THURSTONE, L. L. 1928. Attitudes Can Be Measured. *American Journal of Sociology 33*, 529–554.

VINNIKOV, M., AND ALLISON, R. S. 2014. Gaze-Contingent Depth of Field in Realistic Scenes: The Uesr Experience. In *ETRA '14: Proceedings of the 2014 Symposium on Eye Tracking Research & Applications*, ACM, New York, NY.

WANG, J., BARKOWSKY, M., RICORDEL, V., AND LE CALLET, P. 2011. Quantifying how the combination of blur and disparity affects the perceived depth. In *Electronic Imaging*, SPIE.

WANG, R., PELFREY, B., DUCHOWSKI, A. T., AND HOUSE, D. H. 2012. Online Gaze Disparity via Binocular Eye Tracking on Stereoscopic Displays. In *Second Joint 3DIM/3DPVT Conference: 3D Imaging, Modeling, Processing, Visualization & Transmission (3DimPVT 2012)*, IEEE.

WANN, J. P., RUSHTON, S., AND MON-WILLIAMS. 1995. Natural Problems for Stereoscopic Depth Perception in Virtual Environments. *Vision Research 35*, 19, 2731–2746.

WILSON, J. R. 1996. Effects of Participating in Virtual Environments: A Review of Current Knowledge. *Safety Science 23*, 1, 39–51.

YANO, S., AND EMOTO, M. 2002. Two factors in visual fatigue caused by stereoscopic HDTV images. In *Proceedings of SPIE 4864, Three-Dimensional TV, Video, and Display*, vol. 157.

YANO, S., EMOTO, M., AND MITSUHASHI, T. 2004. Two factors in visual fatigue caused by stereoscopic HDTV images. *Displays 25*.