Int. J. Human-Computer Studies 91 (2016) 37–51

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Impact of depth of field simulation on visual fatigue: Who are impacted? and how?☆

M. Vinnikov*, R.S. Allison, S. Fernandes

Centre for Vision Research, York University, 4700 Keele Street, Toronto, Canada ON M3J 1P3

ARTICLE INFO

Article history:
Received 11 October 2015
Received in revised form
1 March 2016
Accepted 1 March 2016
Communicated by J. LaViola
Available online 7 April 2016

Keywords:
Depth-of-field
Gaze-contingent displays
User experience
Depth perception
Visual-fatigue

ABSTRACT

While stereoscopic content can be compelling, it is not always comfortable for users to interact with on a regular basis. This is because the stereoscopic content on displays viewed at a short distance has been associated with different symptoms such as eye-strain, visual discomfort, and even nausea. Many of these symptoms have been attributed to cue conflict, for example between vergence and accommodation. To resolve those conflicts, volumetric and other displays have been proposed to improve the user's experience. However, these displays are expensive, unduly restrict viewing position, or provide poor image quality. As a result, commercial solutions are not readily available. We hypothesized that some of the discomfort and fatigue symptoms exhibited from viewing in stereoscopic displays may result from a mismatch between stereopsis and blur, rather than between sensed accommodation and vergence. To find factors that may support or disprove this claim, we built a real-time gaze-contingent system that simulates depth of field (DOF) that is associated with accommodation at the virtual depth of the point of regard (POR). Subsequently, a series of experiments evaluated the impact of DOF on people of different age groups (younger versus older adults). The difference between short duration discomfort and fatigue due to prolonged viewing was also examined. Results indicated that age may be a determining factor for a user's experience of DOF. There was also a major difference in a user's perception of viewing comfort during short-term exposure and prolonged viewing. Primarily, people did not find that the presence of DOF enhanced short-term viewing comfort, while DOF alleviated some symptoms of visual fatigue but not all.

1. Introduction

Stereoscopic displays are no longer the exclusive prerogative of the cinema and research laboratories. People nowadays can afford to buy stereoscopic 3D (s3D) TVs and even stereoscopic monitors, tablets and smart phones. Soon the average user will have an option to choose between 2D and 3D displays for any device they might use. Hence, there is increased interest in developing s3D applications for a variety of displays and devices. It is also important to pay attention to the design of the 3D interface to avoid the issues with quality and perceptual human factors that famously contributed to ending the stereoscopic movie fad of the early- to mid-1950s (Zone, 2007).

Stereoscopic displays spark interest not only among users and developers but also the research community. Over the past few years, extensive research has evaluated different aspects associated with these displays. Yet, many questions remain open, including what types of content are best suited to stereoscopic displays, which tasks benefit the most, how long will the user be able to effectively and comfortably interact with such applications and whether stereoscopic displays can be used effectively by everyone in the general population.

One key problem associated with stereoscopic displays is that stereopsis is only one among many of the cues that help people determine depth. Depth cue omissions can significantly impair depth perception and cause viewers to perceive the observed space flatter than it would appear in real life (Watt et al., 2005; Thompson et al., 2004). So, which cues are missing in typical stereoscopic displays and is it possible to add those cues to improve depth perception?

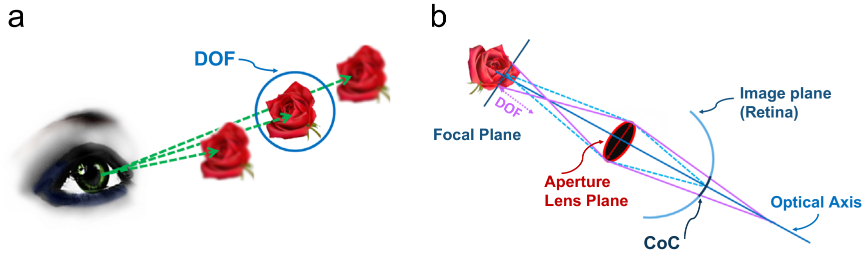
In real life, the clarity of the retinal image of an object depends on its relation to an eye's fixation in the scene (the POR). In other words, the retinal image for a well-focused eye is the sharpest for objects at the focal distance, and is increasingly blurred as the depth of the object from the focal distance increases (Gullstrand, 1910).

☆This paper has been recommended for acceptance by J. LaViola.

* Corresponding author.

E-mail addresses: vision-research@margarita.vinnikov.com (M. Vinnikov), allison@cse.yorku.ca (R.S. Allison).

http://dx.doi.org/10.1016/j.ijhcs.2016.03.001



Depth of Field as perceived by an observer.

Depth of Field Geometry: the formation of the image on the retina.

Fig. 1. Depth field: Field. (a) Depth of Field as perceived by an observer. (b) Depth of Field Geometry: the formation of the image on the retina.

The cornea provides most of the optical power contributing to image formation in the eye but the cornea has a fixed focal length. To focus on the POR, the eye must adjust the shape of its intraocular lens to bring objects nearer than infinity into sharp focus; this process is known as accommodation. The ciliary muscles are responsible for adjusting the shape of the lens to accommodate the eye on an object of interest. Conversely, accommodation provides a physiological cue to the distance of an object: by monitoring the focal state associated with the fixated object the observer could obtain an estimate of the distance of the object. Accommodation is a distance cue that usually is not accounted for in stereoscopic displays. While it is possible to provide this cue to an observer in virtual reality, this requires special volumetric displays. Such displays present targets at different optical distances, for instance by displaying graphics on multiple planes at a variety of optical distances (Sucharov, 1998; Suyama et al., 2000; Akeley et al., 2004).

Although, accommodation provides information concerning the distance of the fixated object, it does not provide static information about depth between objects in the display (although change in accommodation across fixations could be informative). However, when the eye accommodates at a given distance, objects that are nearer or further will be subject to defocus blur and image blur is an informative depth cue (Nguyen et al., 2005; Held et al., 2012). In contrast to the real world, most 2D and 3D graphical applications do not render selective image blurring. Movies, games, and still photography do, but frequently for artistic reasons rather than realistic simulation and, in these cases, the blur is determined by the camera's focus not the eye's focus.

The range of distances where objects are perceived to be in focus for an imaging system such as the human eye is typically referred to as the *depth of field* (DOF) (Fig. 1). In 2D photography, the extent of the DOF is determined by the circle of confusion (CoC), which is a blur circle in the image plane (retina for the human eye). The size of CoC depends on size of the aperture (pupil) and the depth relative to the focal plane. As depth increases, the blur due to the CoC eventually becomes detectable (according to some criterion). Hence, the depth of field depends on both the CoC and the resolution of the sensor (visual acuity of the eye). The border of the CoC is not distinct in a real eye but the CoC is a useful model of DOF. The diameter of the CoC (b) can be approximated by different models. For example, Pentland (1987) used a thin-lens model to describe b as follows:

$$b = A\frac{s_0}{d_0}\left|1 - \frac{d_0}{d_1}\right|$$

where A is the pupil diameter, s_0 is the distance from the lens to the retina, d_0 is the distance from the lens to the focal plane and d_1 is the distance from the lens to another object, whose image forms behind an image plane. Consequently, image blur due to lens defocus can be simulated by calculating blur circles for different objects. By adding DOF to 2D displays, one can contribute to depth perception and improve depth qualitatively (Mather, 1997; Vinnikov and Allison, 2014). In addition, Mauderer et al. (2014) found that gaze-contingent DOF increased perceived realism in 3D images.

Accommodation and defocus blur are less important for the cinema as they are effective cues to distance only for relatively near targets, say less than two meters away. Presentation of such images in the cinema is rare since these would correspond to objects presented at extreme depths with respect to the screen. However, smaller stereoscopic displays are typically viewed at closer distances. In a such scenario, people often rely on an additional distance cue, which is *vergence*. Vergence is a physiological distance cue associated with the movements of the eyes. In vergence, the two eyes rotate in opposite directions. A principal function of vergence is to align the high-resolution fovea of both eyes on a target of interest to get a sharp binocular image of the object. As a result, observers need to increasingly cross, or converge, their eyes as the distance to an object of interest decreases. Typically, convergence and accommodation are tightly coupled (Schor, 1979). However, this is not the case for stereoscopic displays. The problem arises from the fact that, when an observer views a stereoscopic 3D display, she needs to converge her eyes to fuse stimuli located off the screen, while accommodating her eyes at the screen distance. This is known as an *accommodation-vergence conflict* (Fig. 2). This conflict can lead to a range of negative side effects, such as discomfort, eye-strain, headache, and visual fatigue (Luebke, 2003; Mon-Williams and Wann, 1998; Wann et al., 1995; Lambooij et al., 2009; Hoffman et al., 2008). One solution is to try to null the conflict by a quickly adjusting binocular disparities to keep objects of interest near the screen plane (Bernhard et al., 2014). However, as with accommodative displays, such a solution is often limited by the number of discrete physical screens (leading to a limited number of real distances). Such displays also cause unnatural shifting of the rendered scene relative to the screen as points of interest change. Finally such accommodation displays have noticeable artefacts associated with fast disparity adjustments. Hence, a possible solution to alleviate the impact of the negative side effects is to provide an artificial simulation of defocus blur (Brooker et al., 2001; Villarruel, 2006). This would not resolve the accommodative-vergence conflict but would provide the natural relationship between retinal image blur and binocular disparity.

Such an approach requires an invisible user interface that responds to the user's action, in this case their gaze movements, in real-time without any explicit user intent. In order to provide the blur cues present in the real world, the interface needs to measure the POR and update the display in a naturalistic fashion. Ideally, such a simulation will provide a more natural and comfortable interface that users could tolerate for reasonably long periods of intensive use. A DOF simulation has to be congruent to the user's

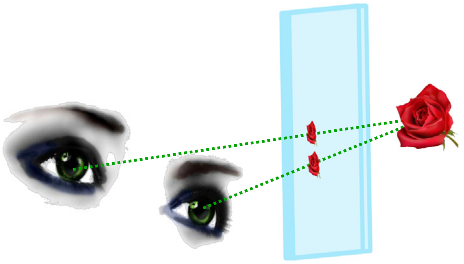


Fig. 2. Accommodation-vergence conflict: the image demonstrates the conflict between binocular vergence and accommodation in a stereoscopic display. The eyes should focus or accommodate at the relatively short distance of the screen to obtain a clear image, however the virtual rose is presented stereoscopically at a further distance and the eyes should converge here when looking at it.

POR at every instant. The best way to achieve this is through eye-tracking applications known as (GCD). Gaze-contingent DOF simulations estimate the POR in 3D space and then alter the visual scene accordingly. Such techniques could be very useful in games and applications that require interacting with 3D content. For example, Hillaire et al. (2008 a,b) used gaze-contingent DOF in the context of navigation and user experience. They observed that users had a positive user experience with DOF added.

This paper is based on our preliminary work (Vinnikov and Allison, 2014), where we evaluated the benefits of DOF as a depth cue and determined the crucial perceptual factors that might impact the strength of DOF as a cue. Specifically, here we investigate whether simulation of DOF could improve a user's depth perception, general comfort, and image quality preferences. Furthermore, we extended this to older adults to see whether there were significant differences between older and younger adults that reflected the marked decline in accommodative ability with age. In this context, the primary goals of this paper were to evaluate the impact of gaze-contingent DOF simulation on:

- *Depth perception*: As we noted above, simulating natural depth of field cues could reduce blur-disparity conflicts and possibly improve depth perception. Previous data on the benefits of DOF simulation are mixed. Our earlier work (Vinnikov and Allison, 2014) indicated that DOF simulation might improve qualitative depth perception. Further evidence for beneficial effects comes from Mauderer et al. (2014), who found that dynamic, as opposed to static, DOF simulation could provide limited information to users about depth order. Leroy et al. (2012), on the other hand, used head-coupled DOF in the context of a pointing task. They examined the impact of blur on both pointing performance and visual fatigue and found no impact on task performance when adding blur. However, their system was head-contingent and not gaze-contingent. The cue conflict between blur and disparity is by definition a binocular phenomenon. Therefore one might expect DOF simulation to be particularly effective for stereoscopic displays. On the other hand even for monocular displays simulated DOF provides an additional depth cue for the user. Previous investigators have not directly assessed the impact of gaze-contingent DOF on depth perception in monocular versus binocular displays. Therefore, in this paper we evaluated the possible perceptual benefits of gaze-contingent DOF in the context of natural gaze interactions with stereoscopic and non-stereoscopic environments.
- *Impact on fatigue/stress in extended viewing sessions* (> 30 min): The reduced blur-disparity conflict provided by simulating natural depth of field cues could improve visual comfort. There are at least two different aspects or stages of visual discomfort. The first stage, often referred to as visual stress, is usually associated with an unpleasant sensation that results almost immediately upon exposure to a stimulus. Consequently, researchers that are interested in how a display or stimulus produces visual stress usually show images to the participants for several seconds (0.5–15 s is typical) and have them judge or compare the images in terms of visual comfort (Kooi and Toet, 2004; Blum et al., 2010; O'Hare and Hibbard, 2013; Vinnikov and Allison, 2014). The second stage is known as visual fatigue and builds up over time, often in response to prolonged visual stress. Visual fatigue often involves visual discomfort, both when the stimulus is present and when it is removed. To study visual fatigue, extended exposures of 30 min to several hours are required (Hoffman et al., 2008; Mon-Williams and Wann, 1998; Ukai and Howarth, 2008). The benefits of DOF simulation on a user's performance during extended sessions (more than 30 min), where the effects of cue conflict are expected to be manifest most significantly, have not been established. Furthermore, the short and long term effects of visual stress on visual discomfort and visual fatigue have not always been distinguished. We believe that it is critical to make such a distinction to fully understand the relationship between stereopsis and DOF on user comfort. To address this issues, we both measure short-term visual comfort and evaluate the effects of stereopsis and dynamic DOF cues on visual fatigue with extended viewing.
- *Subjective ratings of image qualities*: Simulated DOF also is a more realistic depiction of the scene if the simulation is high-fidelity and gaze-contingent. It is, therefore, important to evaluate the impact of simulated DOF in terms of user subjective experience and to see if the perceived realism or aesthetic qualities are affected. For example, Hillaire et al. (2008b) conducted a subjective evaluation, where they asked participants to rate their experience navigating in the virtual environment without DOF, with static DOF fixed to the centre of the screen and finally with gaze-contingent DOF. They found that with the gaze-continent DOF, participants had the best user experience in terms of: rendering realism, fun, depth perception and immersion. Mantiuk et al. (2011) built a display system which incorporated gaze-contingent DOF. They then used their system to test different levels of simulated blur as well as static versus contingent simulations in terms of their impact on subjective ratings of realism. They found that people gave static DOF simulations very low scores, whereas gaze-contingent simulations with medium blur scored the highest. Subjects noticed the DOF manipulation and disliked it. Similarly, Duchowski et al. (2014b) found that while visual fatigue was reduced with DOF people expressed a dislike of DOF on the preference measures. Although Duchowski et al. attributed this to noticeable temporal lag, it is possible that other factors, such as an individual's ability to accommodate and see clearly at certain distances, affected user's ratings. Hence, we evaluate the role of gaze-contingent DOF on subjective experience of the 3D scene. Given that older users have grown accustomed to not being able to readily accommodate everywhere in a scene it is possible the influence of blur on realism in these users differs from young adults. Thus, we looked at these judgements in both younger and older adults.

2. System concept

Fig. 3 shows the system concept for the gaze-contingent DOF system that we developed and used in the present studies. To simulate DOF, we incorporated gaze-contingent image processing and used Unity3D as our 3D interactive graphics platform. As a result, we were able to use different virtual environments with custom shading models and scripts to simulate different static and kinetic scenes. We could also implement different modalities for

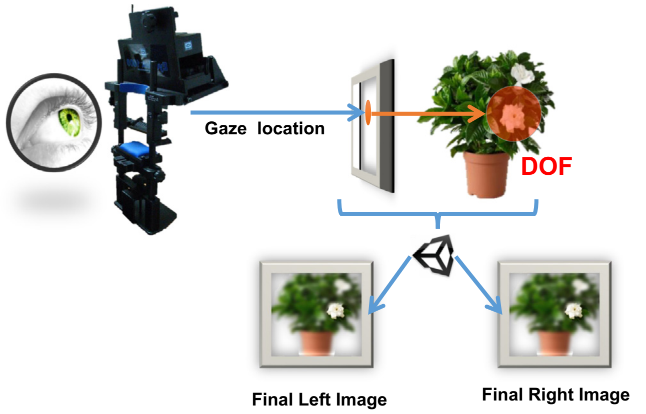


Fig. 3. System architecture. The eye tracker was used to determine POR in the 3D scene based on the location of gaze intersection with the display. This point allowed the system to render DOF about an appropriate focal distance for the right and left eyes.

input acquisition including but not limited to 3D mouse handling. In addition, we developed separate scripts that supported data acquisition and handling from an eye-tracker. These scripts also supported a calibration interface. Furthermore, our virtual scenes always included several camera objects that used shaders and scripts responsible for final image processing operations. This permitted us to implement a range of different visual simulations such as visual distortions and defocus blur.

For the purpose of this paper, we will describe the algorithm for the gaze-contingent DOF simulation as follows:

1. In real-time, we collected the user's gaze location in screen coordinates. This was used for determining the focal distance in 3D space. Estimation of the focal distance was achieved by determining the POR, which is the 3D point, where the user's gaze intersects with a virtual object (Vinnikov et al., 2008). We utilized Unity3D ray projection methods to accomplish this task. This helped us to establish the focal-depth and the region free of blur, in other words, the DOF around that point.
2. We simulated DOF and defocus blur based on an average pupil size of 3.5 mm associated with a display with luminance of 220 cd m^{-2} (Winn et al., 1994). This aperture size determined the extent of blur at different distances from the focal plane. To impose blur, we used a post-processing effect that permitted us to alter each pixel value based on its depth relative to the POR. We implemented this post processing with shader scripts that work directly with the graphical processing unit (GPU) to maintain low latency and not to impact frame rate.
3. We extended our system to support stereoscopic rendering. We achieved this by determining a common POR based on which we then rendered a DOF for each eye independently. It is important to emphasize that in conditions without simulated DOF (monocular cues and stereoscopic cues), we simulated an infinitely large DOF. In other words, the images were sharp at all distances in the 3D space and the left and right images contained binocular disparity. By monocular, we refer to the depth cues simulated; viewing was always binocular. Thus, to implement the monocular condition, we simulated a non-stereoscopic display by presenting identical (zero-disparity) images to both eyes; thus, all processing and rendering steps were equivalent in the monocular-cue and stereoscopic-cue cases. We should also point out that our system works well with any virtual model as far as it can be correctly scaled to correspond to a physical distance from the model to the user.

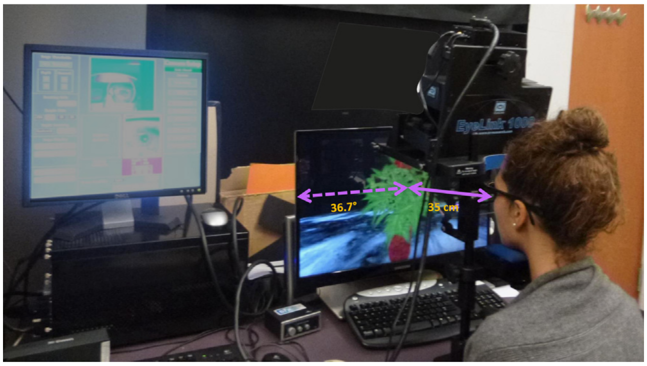


Fig. 4. Experimental setup. The participant sat in front of the 3D display (partially visible beyond the tower mount) with her head fixed by the tower mount provided with the EyeLink 1000. The viewing distance was about 35 cm and half of the field of view was 36.7°.

2.1. Apparatus

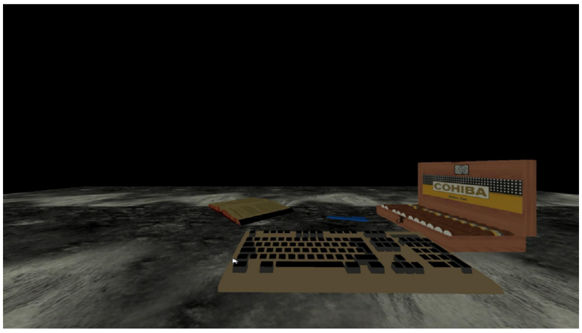
The visual scenes were generated on a desktop computer with the following specifications: AMD FirePro W9000 FireGL, Windows 7 Enterprise, Intel® Core™ i7-3770k CPU, 3.50 GHz, 3.50 GB RAM. The stimulus was presented on a Samsung 950 Series 3D monitor (27 in, with a pixel resolution of 1920 H × 1080 V and a refresh rate of 120 Hz or 60 Hz per eye). On each frame, a stereoscopic pair of images was presented sequentially to each eye via active shutter glasses. The display was used for both experiments but was not auto-stereoscopic, and therefore, the eye-wear was needed to present separate images to the two eyes (it is difficult to get high quality auto-stereoscopic displays with current technology). We carefully verified that stereoscopic shuttering of the active glasses did not interfere with the eye movement estimates. We measured the end-to-end latency at 18.33 ± 5.5 ms using the technique described in Vinnikov and Allison (2013).

During the experiment, the screen was viewed binocularly at a distance of 0.35 m, and the stimulus subtended a horizontal visual angle of 73.5° (Fig. 4). A real-time gaze-contingent system was built by incorporating an EyeLink 1000 (SR Research Ltd) eye-tracker. We used the tracker's tower mount setup, for the purpose of achieving a reliable tracking at the close viewing distance. A chinrest was used to stabilize the head and to maintain the viewing distance, with the midpoint between the right and left eyes, centred in relation to the screen. The experiment was conducted in a darkened room with only the stereoscopic display illuminated.

2.2. Calibration

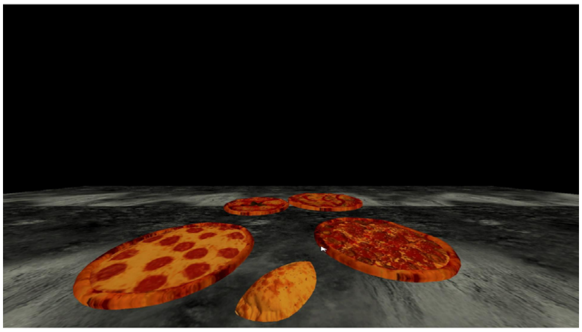
An EyeLink 1000 built-in calibration procedure was always performed before each experimental block. During longer blocks recalibration was performed after a predefined interval (see 4.0.2). The calibration involved sequentially fixating nine points displayed on the screen in a pseudo-random order. Each calibration procedure was followed by the EyeLink 1000 built-in validation procedure, where the participant successively fixated a series of additional nine targets. The error was calculated as the difference between target and estimated eye position and at the end of the validation procedure and an accuracy parameter was displayed. This parameter was used to determine whether the calibration was successful or not. In the case that the tracking error was larger than 0.5° in the display area then the calibration was attempted again. If the required calibration accuracy could not be obtained the participant was dismissed with credit for participation in the experiment. On average the unsigned tracking error for all participants was 0.39 ± 0.24. Approximately 8 participants eliminated based on this procedure for all experiments described in this paper.

a



Office Scene.

b



Pizza Scene.

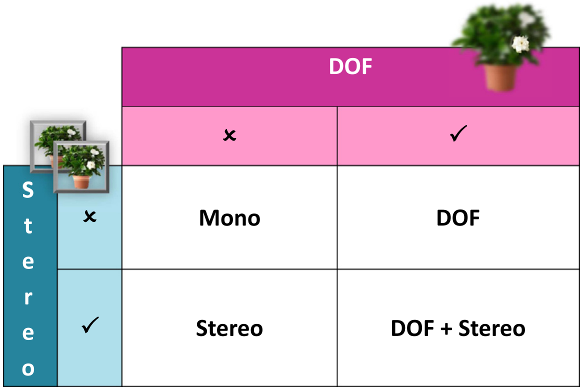
c

Flower Scene.

Fig. 5. The three scene variations.

3. Depth impressions, visual comfort and image quality

In the first series of experiments, we explored the impact of different depth cues, alone or combined, on short-term visual comfort. We were also interested in determining whether short-term visual comfort was affected by age. Numerous studies have shown that the ability to accommodate at different distances changes with age (Koretz et al., 1989, 2002; Spear, 1993). Accommodation is highly plastic in youth, but it dramatically declines by middle age (Koretz et al., 1989, 2002; Spear, 1993). Several hypothesises were outlined, which we aimed to support or disprove. First, it was hypothesized that simulated DOF would be significantly more beneficial for people who have reduced accommodative ability than people with a large range of accommodation. We also were interested to determine if a combination of depth cues in VR settings would provide more compelling depth perception than scenes rendered with fewer depth cues. On the other hand, while we expected that DOF would improve visual comfort, we hypothesised that it might degrade perceived image quality. Finally, we thought that scene content would influence user preference for all three parameters. Specifically, we hypothesized that people would enjoy the presence of DOF in scenes with fine features at different distances.



		DOF ✗	DOF ✓
Stereo	✗	**Mono**	**DOF**
	✓	**Stereo**	**DOF + Stereo**

Fig. 6. Depth cues for each condition.

3.1. Stimuli

The virtual scenes consisted of several 3D models of everyday objects, such as books. In all cases, the objects were placed on a horizontal plane so that the participant could interpret the objects as lying on a table-top. The scenes were designed to resemble those encountered in everyday interaction with objects and common surfaces within a realistic living space. Fig. 5 shows the three scenes presented in the experiment. The scene with the flowers was chosen because it had many vertical elements; conversely, the scene with miscellaneous items had many details distributed across the horizontal plane, such as the keyboard and cigars. Fine detail was also present on the pages of the book with fine print for the second scene. The scene that displayed the pizzas was chosen due to its high level of detail and texture; in this way, the effect of blur on textures would be noticeable. Each of these scenes was presented with all combinations of provided depth cues (conditions).

Each trial consisted of two images of a given scene, presented sequentially each with different depth cues: monocular cues (Monoc), monocular cues combined with DOF (DOF), monocular and stereoscopic cues without DOF (Stereo), and monocular together with stereoscopic cues combined with DOF cues (DOF + Stereo) (Fig. 6).

By permutation there were 12 (4 × 3) pairings of the different cue conditions, three different scenes and four repetitions for each, providing 144 trials in total presented in a different random order to each participant. Each scene was presented to the user for 8.0 s (4.0 s per condition). The trials were divided into four blocks and the calibration procedure was performed prior to each block.

3.2. Task

Participants answered two questions following each trial. It was impractical to have participants come for two consecutive sessions and hence we conducted a between subjects experiment. As a result, half of the participants compared trials based on the following two questions:

1. Which scene had the most compelling depth? (Depth impression)
2. Which scene had the better image quality? (Image quality)

The second half of participants were asked the following two questions:

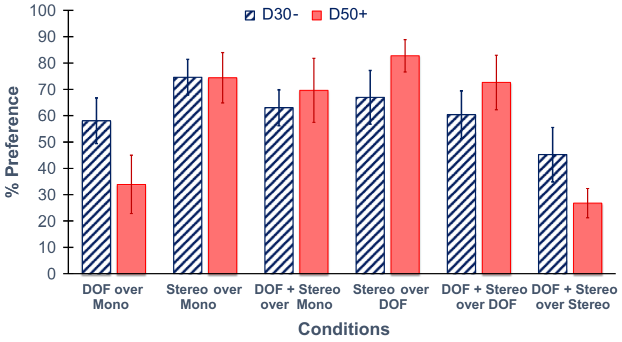


Fig. 7. The mean depth preference averaged across observers for 30–(D30–) and 50+(D50+) groups. The error bars represent the standard error of the mean for each condition.

1. Which scene was most comfortable to view? (Viewing comfort)
2. Which scene had the better image quality? (Image quality)

Participants were required to choose either the first or the second interval, even if there was little apparent difference (a forced-choice procedure). Participants were also instructed to inspect the scenes thoroughly so that fair comparisons could be made between the image pairs. To ensure that the participants were paying attention to the task and actively visually exploring the images, the experimenter monitored participants' eye movements on-line during the trial, to ensure participants were indeed actively scanning the images and that the eye-tracker was properly recording at all times.

To control for a possible Hawthorne effect or other demand effects, the study had both experimental and control conditions and was conducted by three different experimenters, two of who were unaware of the experimental hypothesis. All three followed a strict scripted experimental procedure that intentionally did not reveal the primary experimental hypothesis. Additionally, the experiments were conducted in the dark and participants could not see the experimenters; they interacted directly with the system, rather than with the experimenters.

3.3. Participants

Participants were split into two age groups – the younger adults (under age 30) and older adults (over age 50). Four groups were randomly selected from the recruited participants: (a) younger adults comparing the scenes based on depth impression and image quality (D30–); (b) younger adults comparing the scenes based on viewing comfort and image quality (C30–); (c) older adults comparing the scenes based on depth impression and image quality (D50+); (d) older adults comparing the scenes based on viewing comfort and image quality (C50+);

Adults (30–): The D30– group consisted of nine university students (6 females and 3 males, ranging in age from 22 to 29, average age 24) and the C30– group consisted of 9 university students (6 females and 3 male, ranging in age from 19 to 24, average age 21).

Older adults (50+): The D50+ group consisted of seven participates (3 females and 4 males, ranging in age from 50 to 70, average age 56.57) and the second group C50+ consisted of 7 participants (5 females and 2 male, ranging in age from 56 to 73, average age 60.57).

All participants could clearly see at the viewing distance without the use of eye-glasses. Prior to the experiment, participants were screened for stereopsis, using the RANDOT stereo-test [Chicago, USA]. All participants were able to perform at 50 s of arc or better at distance of 40.6 cm. Written informed consent was obtained from all participants in accordance with a protocol approved by the York University Ethics Board.

Table 1
The chart shows estimated weights of depth cues in determining depth preferences for both 30– and 50+ groups. Each row shows the contribution of each cue (columns) to preference relative to the base case (1st column). A positive estimated weight implies that the case is more likely to be preferred.

Depth Cues	**Mono**	**DOF**	**Stereo**	**DOF+Stereo**
	D30–			
Mono	0	0.12**	0.49***	0.33***
DOF		0	0.37***	0.21***
Stereo			0	–0.16***
	D50+			
Mono	0	–0.25***	0.62***	0.26***
DOF		0	0.88***	0.52***
Stereo			0	–0.36***

Significance of each contribution is shown by stars (.p < 0.1, *p < 0.05).

** $p < 0.01$.

*** $p < 0.001$.

3.4. Results

3.4.1. Depth perception

For both the 30– and 50+ age groups, a Cochran Q test revealed that there were significant differences in depth rendering preference between conditions ($\chi^2(5) = 35.31$, $p < .001$ and $\chi^2(5)= 260.4779$, $p < .001$, respectively). The mean percentage of trials that one condition in a pair was preferred over the other for both age groups is shown in Fig. 7.

In order to better understand the relationship between comparisons, a paired comparison analysis was conducted. Specifically, Turner and Firth model (Turner and Firth, 2007, 2012) software implementation for the Bradley and Terry (1952) models was used, as it fits generalised non-linear models (gnm) using an over-parameterised representation. Table 1 shows the estimated weights from the fitted model. These estimated weights show the contribution of the cues to preference relative to the base case (in separate rows); a positive estimated weight implies that the case is more likely to be preferred in comparisons. It is interesting to note the preference similarities and differences between the two age groups. Overall, in both age groups, stereoscopic displays gave stronger depth impressions than non-stereoscopic displays (Mono) and Stereo estimated weights were significantly more positive than the estimated weights for the other cases. The main difference between the two groups was in the effect of the DOF cue, which was positive for the 30– group, indicating a relative preference in comparison to the Mono base case, but was negative for the 50+ group. This can be seen in Fig. 7; for example, in comparisons of the DOF and Mono conditions, the 30– group choose DOF over Mono ($M = 58.05\%$, $z = 2.65$, $p = 0.008$), while the 50+ group preferred Mono ($M = 33.92\%$, $z = 4.24$, $p < 0.001$). While, the significantly positive estimated weights for DOF + Stereo indicated that the combined depth cues were preferred relative to the Mono base case, the estimated weights were significantly smaller than for the Stereo case. This suggests that adding the DOF cue to a stereo display reduced, rather than enhanced, depth. Consistent with this finding from the model, both groups preferred depth from DOF + Stereo less than from Stereo cues ($30-$: $M = 45.19\%$, $z = 3.50$, $p = 0.0005$; $50+$: $M = 36.78\%$, $z = 5.92$, $p < 0.001$).

In terms of depth perception preferences for different scenes based on age, younger participants showed that they had stronger preference for conditions with added DOF when they viewed the

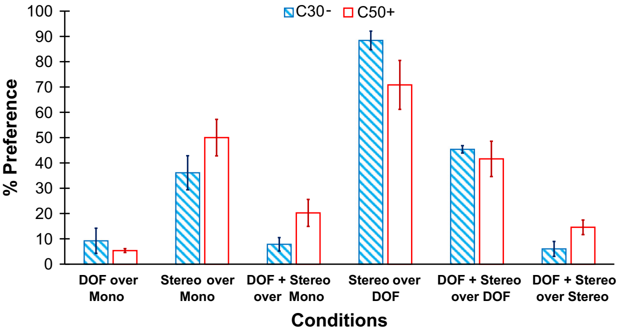


Fig. 8. The mean viewing comfort preference averaged across observers for 30− (C30−) and 50+(C50+) groups. The error bars represent the standard error of the mean for each condition.

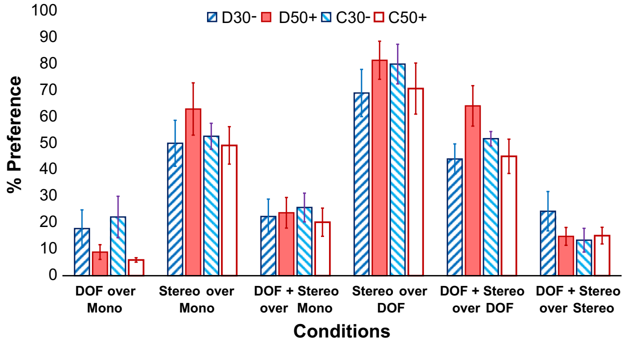


Fig. 9. The mean image quality preference averaged across observers for 30− (D30−,C30−) and 50+(D50+,C50+) groups. The error bars represent the standard error of the mean for each condition.

Table 2
The chart shows estimated weights of various depth cues in determining comfort preferences for both 30− and 50+ groups. Each row shows the contribution of each cue (columns) to preference relative to the base case (1st column). A positive estimated weight implies that the case is more likely to be preferred.

Depth Cues	Mono	DOF	Stereo	DOF+Stereo
	D30−			
Mono	0	−1.24***	−0.22***	−1.36***
DOF		0	1.02***	−0.12*
Stereo			0	−1.14***
	D50+			
Mono	0	−0.79***	−0.12.	−0.90***
DOF		0	0.67***	−0.11.
Stereo			0	−0.78***

Significance of each contribution is shown by stars (** $p < 0.01$).

· $p < 0.1$.
* $p < 0.05$.
*** $p < 0.001$.

flower scene. In particular, they preferred added depth of field for this scene (comparisons DOF over Mono, Stereo over Mono, DOF + Stereo over Mono, and DOF + Stereo over DOF) but not for other scenes. In contrast, the preferences of 50+ group for DOF cues did not depend significantly on the particular scene presented.

3.4.2. Viewing comfort

For both the 30− and 50+ age groups, a Cochran Q test revealed that there were significant differences in comfort preferences ($\chi^2(5) = 555.162, p < .001$ and $\chi^2(5) = 99.321, p < .001$, respectively). The means and standard error for the comfort comparisons for both age groups are shown in Fig. 8, and the estimated weights for both age groups are presented in Table 2.

From the table one can observe that for both groups, the pattern of viewing comfort preference was opposite to depth preferences and much stronger. In other words, conditions with DOF were typically considered less comfortable, while the Mono condition was considered to be the most comfortable condition. This was true even when combined with stereoscopic rendering and stereoscopic rendering was judged more comfortable than DOF + Stereo. It is likely that the DOF simulation determined these ratings since weightings of DOF and DOF + Stereo were similar and much larger than Stereo. Adding the Stereo cue to DOF had only a minor negative impact on comfort (DOF + Stereo was chosen slightly less often than DOF in direct comparisons $M = 45.37\%, z = 2.07, p < 0.04$ for 30− group and $M = 41.60\%, z = 1.85, p = 0.06$ for the 50+ group), but adding DOF to a Stereo display had a large impact (DOF + Stereo was chosen over Stereo only $M = 6.02, z = 11.69, p < 0.001$ for 30− group and $M = 14.56, z = 15.06, p < 0.001$ for the 50+ group). Furthermore, it is interesting to note that the bias against Stereo was smaller for the 50+ group (−0.12) than for the 30− group (−0.22). In particular, when Stereo was compared with Mono, the younger group felt more comfortable with the Mono condition ($M = 36.11\%, z = 1.88, p < 0.06$) while the 50+ group had no preference ($M = 50\%, z = 3.61, p < 0.0003$). However, the 30− group did have a stronger bias for stereo in Stereo versus DOF and DOF + Stereo versus Stereo comparisons.

Generally, the comfort preferences were similar across all scenes with a few exceptions. Data for the 50+ group, suggests that the older group found it more comfortable to view the flower scene with the Stereo condition over the Mono condition ($M = 66.07\%$), but had no preference between these two conditions for other scenes. On the other hand, the young adults preferred Mono over Stereo for the office scene (Stereo versus Mono: $M = 26.38\%$) and had no preference for Stereo compared for the other scenes. Another interesting pattern was observed for the DOF + Stereo versus DOF conditions, where the 30− had no significant preference for all scenes. In contrast, while the 50+ similarly had no preference for the flower ($M = 55.36\%$) and pizza ($M = 44.25\%$) scenes, they preferred DOF over DOF + Stereo 75% of the time for the office scene.

3.4.3. Image quality

Image quality preference data are presented in Fig. 9. A Cochran Q test revealed that there was a significant difference in image quality between conditions for both the 30− group ($\chi^2(5) = 282.22, p < .001$ for C30− group; $\chi^2(5) = 182.69, p < .001$ for the D30− group) and for the 50+ group ($\chi^2(5) = 235.78, p < .001$ for the D50+ group; $\chi^2(5) = 86.08, p < .001$ for the C50+ group). Overall, we observed that the pattern of image quality ratings closely resembled the pattern of comfort ratings. The Turner and Firth Model gnm estimates for both age groups are presented in Table 3. From this data, one can see that there were differences in preferences for both age groups when image quality question was coupled with depth question as opposed to when it was coupled with the viewing comfort question. The data showed interesting trends across different comparisons. For example, when participants were asked to compare Stereo against Mono, only the D50+ and C30− groups preferred Stereo (0.17 and 0.16, accordingly), while C50+ had an opposite preference (−0.13) and D30− had no significant preference (−0.05) at all. A similar pattern was observed with the DOF + Stereo versus DOF condition, where only C50+ preferred DOF + Stereo over DOF (0.26) and the rest of the groups did not have significant preference for one condition over the other.

The image quality preferences for the three scenes is presented in Fig. 10. From the figure one can see that participants preferred Stereo over Mono condition during the viewing of a flower scene,

Table 3
The chart shows estimated weights of image cues in determining image quality preferences for both 30− and 50+ groups. Each row shows the contribution of each cue (columns) to preference relative to the base case (as indicated in subtitles); a positive estimated weight implies that the case is more likely to be preferred.

Group	Mono	DOF	Stereo	DOF + Stereo
	Mono			
D30−	0	−0.61***	−0.05	−0.64***
D50+	0	−0.87***	0.17**	−0.62***
C30−	0	−0.67***	0.16**	−0.62***
C50+	0	−0.81***	−0.13*	−0.87***
	DOF			
D30−		0	0.56***	−0.03
D50+		0	1.04***	0.26***
C30−		0	0.83***	0.05
C50+		0	0.68***	−0.06
	Stereo			
D30−			0	−0.59***
D50+			0	−0.78***
C30−			0	−0.78***
C50+			0	−0.74***

Significance of each contribution is shown by stars (. $p < 0.1$).

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

but they disliked Stereo condition for the other two scenes. A similar pattern was observed for the DOF + Stereo and Stereo comparison, where, once again, during the flower scene people preferred DOF + Stereo, but disliked it when had to observe the office scene. It is interesting to note that although participants did not like DOF + Stereo over Mono for all scenes, they were less critical in the flower scene. Thus, in all cases, the positive influence of stereo cues was larger in the flower scene. For the rest of the comparisons the results were very similar across the scenes.

3.5. Discussion

In terms of depth preferences, we observed that in some cases DOF contributes to more compelling depth perception, specifically, when added to non-stereoscopic displays. The relatively weak influence of DOF on depth can be possibly explained by the fact that focal blur, in general, is not a very strong depth cue. One reason is that static focal blur itself does not provide the sign or direction of depth relative to the focal point. Hence, users need additional depth cues to fill in the missing information. Many of these cues, such as chromatic aberration and accommodative microfluctuations (Nguyen et al., 2005), are not provided by gaze-contingent DOF. In many cases, this sign ambiguity can be resolved by other depth cues such as stereopsis (Grossmann, 1987; Pentland, 1987; Marshall et al., 1996; Nguyen et al., 2005). Our findings are also consistent with Duchowski et al. (2014b), who also observed that people do not like gaze-contingent blurring effects. Their findings highlight the need to improve spatio-temporal fidelity, which includes, but is not limited to improving the accuracy of eye-tracking and reducing temporal lag.

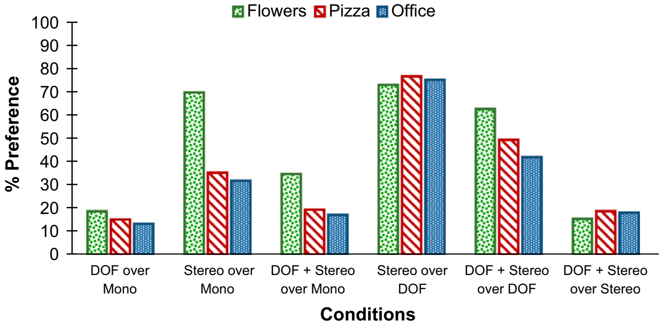


Fig. 10. Mean image quality preference for the three scenes collapsed across groups.

In terms of short-term viewing comfort, we found that both viewing comfort and image quality were reduced, when DOF was added as a depth cue for both age groups. This can be explained by the fact that people might notice the blurriness that results from DOF simulation. Furthermore, we observed very similar preference patterns between image quality and viewing comfort, but patterns were different for image quality and depth. This suggests that image quality, in particular, image blur is an important driver of comfort judgements in our displays. This is supported by medical studies of subjective *depth-of-focus*[1] (Mordi and Ciuffreda, 1998; Atchison and Woods, 1997). These studies suggest that people's tolerance of blur due to defocus can change with age independently from their actual ability to accommodate at different depths and from their objectively measured DOF. The subjective impression of increased DOF with presbyopia is a potentially important factor in the applicability of these techniques to older users.

The age of the participants also had an impact on participants' preference for depth cues. As people age, they slowly lose their ability to accommodate to near objects (Ramsdale and Charman, 1989; Koretz et al., 1989, 2002; Spear, 1993; Hayashi et al., 2003) and possibly slowly stop using DOF as a depth cue at all. One might predict that restoration of this depth cue for presbyopes who lost this ability might be more beneficial than doing the same for the younger users. However, we did not find any supporting evidence for this effect. This may be explained by our finding that the older-adults group was more tolerant to the presence of DOF in terms of viewing comfort. This is possibly due to the fact that, over the years, they have become less sensitive to the presence of blur than the younger population (Mordi and Ciuffreda, 1998; Wang and Ciuffreda, 2006). For example, in their study Mordi and Ciuffreda have shown that subjective tolerance to DOF substantially increases with age, implying that image quality degradation that would be tolerated by a typical older observer would bother the typical younger observer (Mordi and Ciuffreda, 1998). A relative insensitivity to blur (increased subjective DOF) may explain why DOF was a less effective depth cue in the older observers.

When designing a gaze-contingent DOF, one has to also understand the scene and what features will attract a participant's attention. For example, people often direct their gaze to the edge of an object, where a tracking error or depth estimation error can lead to incorrect location of simulated DOF (van der Linde, 2004). Binocular eye-tracking can provide a convergence angle signal to fixation depth that is independent of the actual geometry and screen parameters of the virtual scene. For example, Duchowski et al. (2014b) built such a system. However, vergence estimation for binocular eye-tracking is very sensitive to noise. Furthermore, observers often have fixation disparity, and this might be more pronounced in virtual environments. For instance, Duchowski et al. (2014a) compared depth estimation in a virtual environment with a physical environment. They found that vergence error varied with target distances. Moreover, this error was larger when participants looked behind or in front of the screen. Thus, depth

---

[1] Atchison and Woods (Atchison and Woods, 1997) defined subjective depth-of-focus as “the range of focusing errors for which the image of the target appears to have the same clarity, contrast and form as the optimal in-focus image”. That is not to be confused with an objective DOF that measures the actual change in blur with depth.

estimation from vergence can lead to inaccuracies in depth judgements. In this study, the most accurate eye-tracking was achieved by using a monocular tracking configuration. Moreover, out of the three scenes, people favoured the presence of DOF in the flower scene, as this was the scene with the highest number of edges and most variable depth. Finally, it is important to mention that although we tried to make our scenes realistic as possible, we had to simplify our scene to ensure reliable real-time, low latency rendering. We believe that it is possible that our current findings could have been stronger in more realistic scenes. In such scenes, we believe, correct lighting and shadowing that would be altered as a result of DOF simulation would possibly provide a more compelling impression of depth but would need to be rendered with low latency.

The difference between age groups and their scene-dependent preferences was also evident in the office scene. In this scene, several big objects contained many small objects and were rich with edges and small cavities. For instance, the keyboard had many small keys and several cigars were lined up in the cigar box. As a result, this stimulus provided participants' with a strong monocular perspective depth cue. Indeed, this was evident, when participants compared (a) the Stereo condition with the DOF condition and (b) the DOF + Stereo condition with the DOF condition. In both cases, although, participants preferred the stereo conditions, their preferences were weaker for the office scene than for the other scenes, on all three questions (depth impression, visual comfort and image quality). Similarly for the image quality question, participants preferred the Mono condition over the Stereo condition in the office scene. This suggests that when dealing with a detailed scene with fewer depth cues (monocular versus stereo) users preferred to see the details as clearly as possible.

The preference for Stereo in the office scene could possibly be explained by the peripheral layout of the scene. When we designed the office scene, we tried to minimize occlusion of small objects and hence we spaced out the large objects throughout the virtual scene. Therefore, most of the big objects ended up either in the bottom of the scene or at the far right edge of the scene. Yet, DOF can only be an effective cue to depth or influence image quality, if the blur introduced can be detected by the visual system. In the office scene, when participants fixated a particular object, the rest of the objects in the scene fell in the peripheral region of the observer's visual field. However, the blur of the object on the retina is due to two factors: defocus and peripheral blur. Perhaps, people could not detect the defocus blur in the periphery or attributed blur in their periphery not as defocus blur but rather as peripheral blur. Unfortunately, there is very little research that addresses the combination of defocus blur and peripheral blur. It is possible that simulated blur due to defocus has to be stronger in the periphery in order to provide less ambiguous depth cues to the user. van der Linde (2004) was one of the few people who incorporated both types of blur into a gaze-contingent system, but he did not include any empirical evaluations.

4. Long term comfort

In the previous section, short-term exposure to different depth cues and observers' preferences was examined. In the study presented below, we were interested in examining observers' comfort ratings, when they were presented with different depth cues for a prolonged periods of time.

Our main objective in conducting this experiment was to evaluate and compare a user's visual fatigue with stereoscopic and gaze-contingent DOF displays. Our main hypothesis was that by combining DOF with stereo, it is possible to alleviate some of the negative effects associated with stereoscopic displays in general and specifically, with accommodation-vergence conflict (or more precisely with disparity-blur conflict).

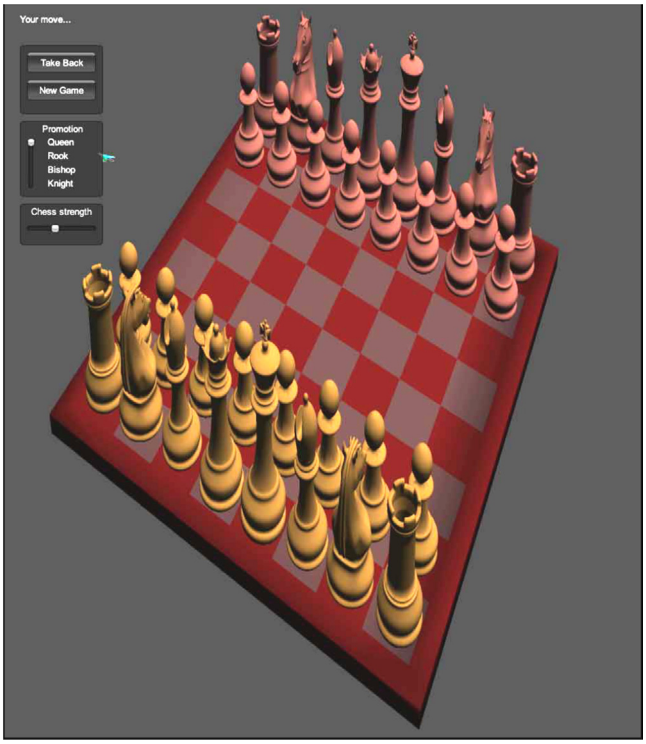


Fig. 11. Stereoscopic chess board.

Our intent was to assign a representative and realistic task to a user for a prolonged period of time (such as object selection in 3D space). We also wanted to select a task that would require participants to interact with the virtual environment (VE) – reaching and shifting their gaze to different regions of 3D space. The game of chess was very appropriate and met our requirements, as the game requires a detailed scan of all the pieces on the board. The rules of the game also require players to perform piece selection and transfer to a new location. A between-subjects experimental design was used with half the participants playing the game with stereoscopic images (Stereo condition) and infinite DOF and half playing with stereoscopic images and realistic DOF (DOF + Stereo).

4.1. Stimuli

We adapted and incorporated a freely-available Unity3D chess project into our gaze-contingent system (Chessforeva, 2010). Consequently, the virtual scene consisted of a chess board and chess pieces (Fig. 11). We controlled the camera convergence (horizontal image translation) so that the chessboard partially protruded from the screen. Thus, part of the board on the player's side was presented stereoscopically out of the screen plane, while the rest of the board that belonged to the computerized player was set further into the screen. By performing these manipulations, we created a focal range of 0.45 m. Specifically, the participant sat at a distance of 0.35 m away from the screen, the near part of the board was at about 0.27 m from the participant and the furthest point was about 0.72 m from the participant. We rotated and fixed the chessboard in such a way that the user could easily select any square in 3D space. We also adjusted the colour scheme of the board and chess pieces to minimize the ghosting commonly associated with stereo displays (Tsirlin et al., 2011) so that it would not be perceptible by the user.

4.2. Procedure

The experimental session proceeded in three stages. The first stage was a pre-test stage. During that period of time, participants underwent a series of visual tests and viewed a short video about how to play chess. They also had a short non-stereoscopic demonstration on how to play the chess game. At the end of this stage, participants had to fill out a fatigue questionnaire that served as a baseline for comparison. The second stage was the game stage, where participants played the chess game for 30 min. If a game was completed (by checkmate or draw), then another game immediately followed until the 30-min interval was completed. It was very important during this stage that the eye-tracker was recording participants' gaze location accurately. Therefore, participants went through a calibration procedure before beginning the game. Special care was taken to limit the overall calibration time to no more than five minutes (up to three setup adjustments and calibration/validation attempts). To minimize visual fatigue due to calibration, the experimental protocol required dismissing participants who required longer calibration time (however, there were no cases that required prolonged calibration). Every five minutes throughout the game, calibration was evaluated and a drift correction was applied, if required. If the drift correction result exceeded a threshold of 0.5°, the eye-tracker was calibrated again. Although the drift correction check was performed approximately five times during the game, care was taken to limit calibration time, including any re-calibrations, to five minutes total as discussed above. This was done, again, in order to minimize visual fatigue. The third and final stage was a post-test stage. During that time participants repeated the fatigue questionnaire and visual tests performed in the pre-test stage. Each session concluded with a debriefing.

4.3. Participants

Each of the two groups of participants consisted of 11 university students (5 females and 6 males in each group, ranging in age from 18 to 29, average age 21.23). All participants had uncorrected distance visual acuity of 20/30 or better and could see clearly at the viewing distance without glasses. Written informed consent was obtained from all participants in accordance with a protocol approved by the York University Ethics Board.

As part of the pre-test, participants were screened for stereopsis (see Section 4.4) and were excluded if stereo acuity was worse than 50 s of arc at a distance of 40.6 cm.

4.4. Evaluation

4.4.1. Subjective measures

One of the most common measures of visual fatigue involves questionnaires to probe the subjective experience. Unfortunately, a fatigue questionnaire that is sensitive, valid, reliable and robust has yet to be established (Lambooij et al., 2009). Some researchers have asked only one general question. For example, Yano et al. (2004) asked participants to rate how tired they were after an exposure. They used a five-point Likert category scale ("not tired", "sense a little tired", "little tired", "tired", "very tired"). Similarly, Kooi and Toet (2004) also asked their participants to rate viewing comfort with a five-point scale. On the other hand, other researchers used longer questionnaires that were modified from other standards and requirements like ISO (1992). Peli (1998) asked participants to rate aspects of visual discomfort such as eye dryness, irritation in eyes, difficulty in focusing, postural discomfort and headache on a scale from none to severe. Kuze and Ukai (2008) conducted a subjective evaluation of visual fatigue caused by moving images using a more extensive list of symptoms querying different eye sensations and postural pains. Similarly Duchowski at el. used questions such as "How tired are your eyes?", "How clear is your vision?", "How tired and sore are your neck and back?", "How do your eyes feel?", "How does your head feel?". Interestingly, for their main study they choose a 0–100 rating over the typical Likert scale claiming it provided a more detailed report of participants' experiences. For our study, we adopted the questionnaire of Sheedy and Bergstrom (2002), Sheedy et al. (2003) that was developed based on clinical experiments and observations. This questionnaire identifies different symptoms based on sensation type and location, as well as internal and/or external symptoms. Specifically, our questionnaire can be divided into three categories. The first category concerns eye symptoms, and it included questions on tired eyes, irritated eyes, dry eyes, watery eyes, pulling feeling on eyes, burning sensation in eyes, ache behind or in the eyes, eye strain, uncomfortable vision, blurry vision, difficulty focusing and headache. The second category is associated with motion, and is related to simulator sickness symptoms in virtual environments. It included questions on nausea and difficulty in concentrating. The third category was a musculo-skeletal category that included questions on neck pain, postural discomfort and muscular stress. In their review paper, Lambooij (2009) recommended to not only use Sheedy's questionnaire, but also to extend the questions to query additional background information such as previous experience. As a result, we added two more questions intended to reflect the general physical state of a participant. Specifically, we asked them to rate tiredness and sleepiness. Finally, we also included a demographic questionnaire that inquired about participants' previous experiences with DOF + Stereo displays.

4.4.2. Measures of visual function

Visual fatigue is often associated with impacts on visual function, which can impact task performance. As a result, and due to the lack of standard, validated subjective measures, there has been interest in measuring oculomotor or visual correlates of visual fatigue. Two of the most commonly used measures are dynamic accommodative response and refractive error (Yano et al., 2004; Fukushima et al., 2009; Inoue and Ohzu, 1997; Okada et al., 2006; Ukai and Howarth, 2008). One of the main problems with these measures is that they require special equipment with very specific installation requirements. In addition, some devices provide very coarse results (Lambooij, 2009). Due to the nature of our physical setup, this type of measure could not be incorporated.

On the other hand, changes in pupil diameter are associated with accommodation - vergence eye movements and hence have been proposed as a correlate of visual discomfort (Ukai and Kato, 2002; Ukai and Howarth, 2008). It has been observed that the pupil constricts with near viewing to provide a large DOF and dilates at far viewing distances (Howard, 2002). Pupil diameter was chosen over dynamic accommodation since that data was already available from an eye-tracker, while the latter required specialized equipment to measure. We expected to see an ongoing change in pupil size with change in participants' vergence. We also expected to see less variation in pupil diameter over time in cases of people experiencing visual fatigue.

Among many other possible physiological parameters, blink rate has been associated with fatigue and consequently, we wanted to examine if visual fatigue in stereoscopic displays is associated with a higher rate of blinking as found in other studies of visual fatigue (Stern et al., 1994; Kim et al., 2011). As such, we collected the number of blinks along with other eye-data from the eye-tracker.

Another approach that is commonly used for fatigue examination is a clinical optometric measurement. In many cases these measurements are preferred, as they are easily accessible,

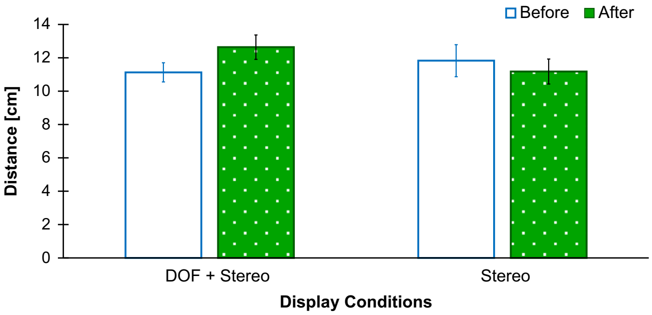


Fig. 12. Mean before and after results for near point vision test for the two display conditions. The error bars indicate the standard errors of the means for each condition.

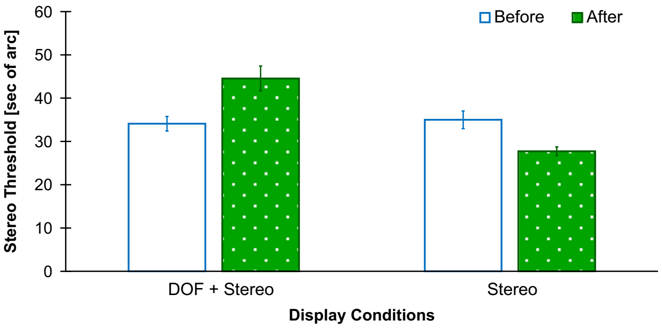


Fig. 13. Mean before and after results for Randot stereo test for the two display conditions. The error bars indicate the standard errors of the means for each condition.

noninvasive and can be used consistently with a large group of participants. Usually, these measures are taken before and after visual exposure to a stimulus. They are also helpful in the prescreening process. We chose measures that were similar to Lambooij (2009), who recommended performing the following tests: the visual acuity test, stereo acuity test and convergence test. Therefore, we also included phoria[2] (horizontal and vertical) tests to be consistent with other studies that looked at visual fatigue with stereo displays (Mon-Williams et al., 1993). All of the above measures were important as they allowed us to ensure that participants were successful at the seeing the stereo content that was presented to them during the experiment. We also wanted to ensure that the tests we used were validated as standard clinical tests and provided important information about binocular visual performance.

Consequently, we had three types of evaluation. First, we evaluated visual and oculomotor functions with the following instrumentation:

- *Stereo acuity*: RANDOT® stereo-test [Chicago, USA].
- *Visual acuity*: Freiburg Vision Test (Bach, 2007).
- *Near point visual test*: The head was stabilized on a chin-rest and participants had to fixate at the smallest line they could see on standard near-vision test card at a distance of 0.65 m away. Participants then had to a slide test card toward their head along an optical rail until they could just see the selected line in focus. To achieve maximum accuracy, participants were prompted to move the card to a position, where they detected the text to appear blurred and then move the card back until the text appeared sharp again.
- *Vertical and lateral phoria*: We used Titmus Biopter Vision Test [Petersburg, USA].

Second, we composed our own visual fatigue questionnaire as described above. Finally, we also collected and analysed eye-data, specifically blink rate and pupil size.

4.5. Results

4.5.1. Measures of visual and oculomotor function

The results for the near point of clear binocular vision measure (near point vision test) are shown in Fig. 12. With a mixed-effect ANOVA, we found a significant interaction between the testing interval (pre-test and post-test results) and the different type of displays (Stereo or DOF + Stereo) ($F(1,20)=8.248$, $p=0.009$). This interaction is the factor of principal interest, as it reflects differences in exposure effects between the two display types. Specifically, participants in DOF + Stereo condition experienced an increase in near point distance, while participants in the stereo condition had their near point decreased (i.e., improved) on average after stereoscopic presentation with infinite depth of field.

Table 4
Measures of visual and oculomotor functions: the table shows the interaction between the objective measures before and after the chess game and the display type (Stereo or DOF + Stereo).

Visual and oculomotor function	$F(1,25)$	p
Randot Score	6.183	0.0219*
Binocular Acuity	0.707	0.41
Right Eye Acuity	0.005	0.945
Left Eye Acuity	0.000	0.995
Vertical Phoria	1.250	0.277
Lateral Phoria	1.763	0.199
Near Point	8.248	0.009**

Similarly, a significant interaction was found between testing interval (before and after scores) and the two viewing conditions on the stereo acuity test ($F(1,20)=4.67$, $p=0.043$). In the DOF + Stereo condition, the stereoscopic threshold was increased on average relative to the baseline, while following stereoscopic displays without DOF simulation stereo threshold improved (declined) on average relative to baseline (results are shown in Fig. 13, note that the decrease in stereo acuity actually corresponds to an increase in threshold seconds of arc).

The rest of the effects and interactions for the other measures were not significantly different between the two viewing conditions (Table 4), except for an interaction between gender, display type and the pre- and post- scores on the lateral phoria test. This seems to be mainly due to one female participant in the stereo condition, who had a very large change in phoria measurement. It is possible this is due to gender differences in perception of 3D (Maurin et al., 2006).

4.5.2. Subjective fatigue questionnaire

A summary of Kruskal–Wallis Chi-squared tests for the difference in exposure effects between before and after responses on the questionnaires collapsed over display condition is presented in Table 5. The results show that there was a significant difference between subjective responses before and after playing the chess game. Specifically, in terms of *general experience*, there was a significant post-test versus pre-test increase in tiredness. In the *eye-symptoms* section, there were significant differences for the *tired eyes* and *uncomfortable vision* questions, where the scores grew larger in the post questionnaire. For the rest of the questions in this category, the difference between the pre and post exposure

---

[2] Phoria is a visual test that evaluates vergence position of the eyes during monocular viewing, when user's eyes accommodate at a certain distance. It considered to be a good measure of determining the most comfortable fusion distance for the user.

Table 5
Kruskal–Wallis Chi-square for difference in the mean pre-test versus mean post-test responses to the fatigue questions (collapsed over conditions).

Questions	Pre	Post	$\chi^2(2)$	p
General questions				
Q1 Tiredness	1.5	2.72	8.313	0.004**
Q2 Sleepiness	1.41	2.0	2.074	0.150
Eye-symptoms				
Q3 Tired eyes	1.24	2.55	9.398	0.002**
Q4 Irritated eyes	0.77	1.64	3.428	0.064 .
Q5 Dry eyes	0.77	1.32	2.051	0.152
Q6 Watery eyes	0.18	0.82	2.932	0.087 .
Q7 “Pulling” feeling on eyes	0.81	1.41	3.499	0.061 .
Q8 Burning sensation on eyes	0.31	0.59	3.113	0.078 .
Q9 Ache behind or in eyes	0.68	1.27	3.385	0.066 .
Q10 Uncomfortable vision	0.31	1.18	5.484	0.019 *
Q11 Blurry vision	0.45	0.59	0.012	0.912
Q12 Difficulty focusing	0.45	0.91	2.834	0.092 .
Q16 Headache	0.27	0.77	3.394	0.065 .
Simulator sickness symptoms				
Q15 Nausea	0.0	0.36	7.401	0.007 **
Q18 Difficulty concentrating	0.41	0.77	1.177	0.278
Musculo-skeletal symptoms				
Q13 Stress	0.45	1.27	0.258	0.611
Q14 Neck pain	0.73	1.72	3.713	0.054 .
Q17 Postural discomfort	0.55	1.55	3.679	0.055 .

Table 6
Kruskal–Wallis Chi-square for differences between the Stereo and DOF + Stereo conditions in exposure effect (Post-test minus pre-test fatigue responses).

Difference between before and after	$\chi^2(2)$	p
General questions		
Q1 Tiredness	0.011	0.917
Q2 Sleepiness	1.279	0.258
Eye-symptoms		
Q3 Tired eyes	2.016	0.156
Q4 Irritated eyes	0.287	0.592
Q5 Dry eyes	0.341	0.559
Q6 Watery eyes	1.014	0.314
Q7 “Pulling” feeling on eyes	0.479	0.489
Q8 Burning sensation on eyes	2.335	0.127
Q9 Ache behind or in eyes	1.417	0.234
Q10 Uncomfortable vision	1.123	0.289
Q11 Blurry vision	0.787	0.375
Q12 Difficulty focusing	0.121	0.729
Q16 Headache	1.432	0.232
Simulator sickness symptoms		
Q15 Nausea	5.091	0.024 *
Q18 Difficulty concentrating	3.279	0.070 .
Musculo-skeletal symptoms		
Q13 Stress	0.158	0.691
Q14 Neck pain	1.43	0.231
Q17 Postural discomfort	3.286	0.070.

was marginal. Nonetheless, the mean post scores were larger. There was some gender interactions in terms of sensitivity to *eye ache* and type of displays presented. However, due to a small sample size, we cannot conclude if the effect was due to an outlier. In the *Motion* section, there was a significant difference for the *nausea* and marginal difference for the *difficulty concentrating* questions. Finally, in the *Musculo-Skeletal* section only *postural discomfort* was marginally significant.

The above findings indicate that participants showed increase in fatigue after the game. The most relevant question is whether these effects differed between the two conditions. A summary of Kruskal–Wallis Chi-squared tests for the difference in pre- versus post-test change between the DOF + Stereo and the Stereo conditions questionnaires are presented in Table 6. Significant differences between conditions were observed only for three questions. Specifically, in the *motion* (simulator sickness) category, there was a significant difference between stereo displays compared to stereo and DOF displays in terms of increase in nausea. Marginally significant differences were found in the case of difficulty in concentrating, where people experienced a harder time concentrating following the Stereo condition than the DOF + Stereo. Lastly, in the musculo-skeletal condition, we observed marginally significant differences in ratings of postural discomfort, where people experienced a larger increase in discomfort following stereo compared to DOF +Stereo. To clarify the magnitude of our results, the results for these three questions are presented in Fig. 14.

4.5.3. *Eye-data*

Pupil size and blinks were detected using EyeLink built-in algorithms. We tracked the pupil size and blinking rate for the entire 30 min of each session. In this data, we looked for trends by looking for regressions for the pupil size and blinking rate as a function of time, and we binned this data in five minutes thresholds and looked for changes across the bins. The change in pupil size or change in blinking rate did not manifest any significant trends as a function of either time or viewing condition.

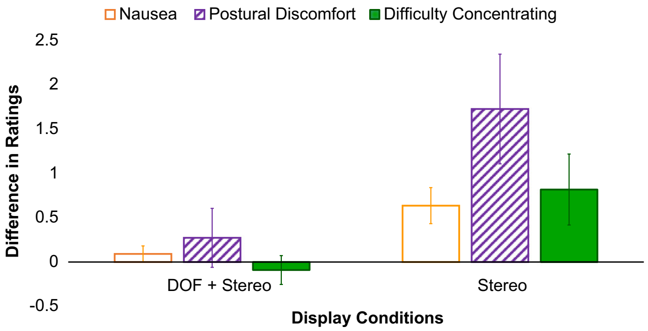


Fig. 14. Change in subjective ratings after display exposure averaged across observers. The difference was calculated by subtracting the before ratings on the Likert scale from the after ratings on the scales. Error bars show the standard error of the mean for each condition.

4.6. *Discussion*

The purpose of the study was to evaluate visual fatigue after prolonged viewing of stereoscopic displays by comparing two conditions: Stereo and DOF + Stereo. It was expected that there would be a significant alleviation of the fatigue symptoms following DOF + Stereo relative to Stereo alone. We indeed observed some mixed indications that DOF might be a beneficial addition to stereoscopic displays.

In terms of visual and oculomotor functions, stereoscopic acuity increased and the near point of convergence decreased (i.e., both improved) post-game with stereoscopic viewing, but worsened with DOF. The improvement with stereoscopic viewing is consistent with stereoscopic adaptation that has been shown in the literature following practice with stereoscopic stimuli (Long and Over, 1973; Blakemore and Julesz, 1971). It is possible that this adaptation effect was absent in the DOF + Stereo condition because of the additional depth cue provided by DOF that reduced reliance on stereopsis and hence the stimulus to adaptation. As a result, in the DOF + Stereo condition, there was no significant improvement in stereo acuity and near point. As argued by Mather and Smith (2000) the effective range of stereopsis and blur are complementary and in some cases visual blur is expected to predominate for relative depths that exceed the range of disparity mechanisms. Specifically, Mitchell (1966) showed that blur can limit the range of disparity processing and thus DOF simulation may have reduced the range of stereoscopic depth processing which, in turn, limited adaptive changes in oculomotor functions.

In terms of subjective measures, Stereo display induced more nausea relative to the DOF + Stereo display, which is consistent with a beneficial effect of DOF (Mon-Williams et al., 1993; Yang and Sheedy, 2011). Nausea is usually associated with simulator sickness. Therefore, this is a very important subjective measure and these results suggest gaze-contingent DOF might be beneficially incorporated in virtual environment applications such as simulations and games. We also found that postural discomfort and difficulty concentrating was more of a problem after viewing displays with stereo alone than when DOF was added to stereo. This could be due to a direct effect of reduced cue conflict in the DOF + Stereo reducing visual fatigue. An alternative possibility is that, due to a more realistic display in the DOF + Stereo (compared to the Stereo condition), people were more immersed and had experienced fewer physical and cognitive distractions. Mauderer et al. (2014) in their recent study have argued that gaze-contingent DOF contributes to perception of depth (see also experiment 1) and hence makes the virtual scene appear more realistic. As reported in the results section, no significant differences were observed between the two conditions on any other fatigue questions. This suggests that dynamic DOF may not be enough to alleviate all fatigue symptoms.

5. General discussion

In summary, we built a gaze-contingent DOF system simulating the defocus blur that is associated with a user's eye accommodating at different distances within virtual space. We also presented two studies to evaluate different aspects of DOF impact. Specifically, we looked at the effect of age associated with short-term exposure to DOF and we also looked at the effects of short and prolonged exposure to DOF. We hypothesized that DOF could improve the perception of depth in an older population. However, our results showed that DOF degraded depth quality for people who were not able to accommodate in real-life. On the other hand, for the young participants, we hypothesized that DOF could potentially enhance depth perception and alleviate the accommodation-convergence conflict. We demonstrated that simulated gaze-contingent DOF could indeed enhance depth perception when combined with other depth cues. Furthermore, for all age groups, DOF when combined with stereo did not seem to enhance short-term viewing comfort and has strong negative impact on perception of image quality. This implies that DOF is not a straightforward solution for accommodation-convergence conflict. Nonetheless, we hypothesized that DOF might alleviate the fatigue associated with long-duration exposure to stereo display. Our second experiment demonstrated that DOF indeed alleviates some symptoms of visual fatigue, but not all.

We believe that our studies also shed some light at why there are some inconsistencies between different studies. Based on our first study, we believe that special care should be taken when choosing participants for DOF studies/applications. One needs to be aware that a decrease in the ability of user to accommodate, will also result in a decrease in the ability of the user to enjoy the benefits associated with simulated blur. For example, a game for a young audience could choose to add DOF to a non-stereoscopic display and the users should experience stronger depth, while for an application developed for a mature crowd it is important to incorporate stereo and reduce blur as much as possible. When catering for the general public, one needs to find a balance to accommodate both age groups.

In their paper, Zhang et al. (2014) claim that DOF cannot be a primary cause of discomfort. They supported their claim by the O'Hare et al. (2013) study that looked at prolonged viewing of stereoscopic images with different levels of DOF. Zhang et al. (2014) study is similar to our long-term comfort experiment. It is important to reiterate that we have also found some evidence for a positive influence of DOF in extended viewing scenarios. Therefore, one of the contributing factors to the difference in visual comfort between studies is exposure time, which possibly activates different mechanisms during various periods of time. As we discussed in the introduction, visual discomfort can be broken into two stages: the visual stress associated with short exposure and visual fatigue that results from prolonged exposure. Another important factor that possibly attributes to the difference between studies is that the results from O'Hare et al. were not based on gaze-contingent DOF simulation. Although the task in O'Hare et al. required people to maintain fixation at a predefined distance, it has been shown that there is a significant difference between static and gaze-contingent DOF (Hillaire et al., 2008b). This could be due to the fact that eye-movement possibly contributed to depth perception and/or people do not follow the assigned task.

It is also interesting to note that there was a significant difference between experimental results and the type of displays used during experimentation. For example, in their work, Hoffman et al. (2008), looked at visual performance and visual fatigue with or without vergence-accomodation conflict. In their study, they used a volumetric display and they showed that such conflict indeed creates visual fatigue and discomfort. Duchowski et al. (2014b), on the other hand, tried to replicate Hoffman et al. (2008) and Hoffman and Banks (2010), but instead of using a fixed-viewpoint, volumetric display they used a gaze-contingent DOF stereo display, they did not find any difference in user's comfort rating.

They attributed this to two flaws associated with the experimental design: the limited scale that did not provide enough room to exhibit the difference between conditions and prolonged breaks between each session and filling in of the fatigue questionnaire. Interestingly, Langer and Siciliano (2014) also tried to replicate Held et al. (2012) study by using DOF stereo displays with controlled fixation. Although the task was not related to visual comfort, but rather to a depth discrimination, both tasks require the viewer to resolve blur associated with defocus. Once again their work indicated that observers did not utilize DOF in contrast to Held et al. (2012). It is possible that there are methodological differences between studies. However, there is also a possibility that simulated DOF omits or competes with other depth cues that are naturally presented in a volumetric stereo display.

Generic simulations of natural DOF based on a thin lens plus aperture optical model have natural limitations. Although we did our best to match the blur to a realistic DOF, we could not account for individual optical differences in our users' eyes. Such differences are idiosyncratic and dependent on pupil size, chromatic content and other factors (Burge and Geisler, 2011). Eye-tracking can provide on-line estimates of pupil size to allow for modelling of simulated aperture to actual pupil size. However, the DOF simulation modelled focal blur that results from viewing through a finite sized pupil but not the effects of aberrations and micro-fluctuations of accommodation that are believed to be important to disambiguate depth from accommodative blur (Nguyen et al., 2005). Theoretically, if simulated depth of field matched a natural depth of field, the simulation should be indistinguishable from a real scene (in terms of image sharpness). Despite our best effort to minimize latency of the system, in some cases, the change in focus was still noticed by participants and it unfortunately likely reduced the comfort (Kooi and Toet, 2004; O'Hare and Hibbard, 2013). Kooi and Toet (2004) and O'Hare and Hibbard (2013) reported that noticeable imperfections in a stereoscopic image can be described as very uncomfortable for an observer. Despite the fact that we used a high quality research grade eye-tracker in its most precise tower set-up, all eye-trackers are prone to estimation error and quality of tracking varies between observers. In gaze-contingent DOF simulation this can lead to inaccuracies in estimation of the POR, and hence, focal distance. It is possible that these technical limitations on spatial and temporal precision reduced image quality from expectations of natural viewing. However, it is also possible that users actually preferred the image quality of the non-realistic scene with infinite DOF over their natural viewing. This is a reasonable supposition, since the image quality of the infinite DOF is better than expected from natural viewing in terms of image sharpness and image information. Consequently, high-frequency content in the scene is preserved everywhere in the image. Therefore, it is possible that users judged image quality as better in these 'hyper realistic' images than in more realistic, but blurry, DOF simulations.

Acknowledgements

The authors would like to thank Rida Waseem and Carly Hylton for their help in conducting some of the experimental sessions; Tonny Lay and Anna Kiseleva for their assistance in making a video demo, and Katherine Allison for the voice-over. We would like to thank NSERC Canada for the Discovery grant to Robert Allison.

References

Akeley, K., Watt, S.J., Girshick, A.R., Banks, M.S., 2004. A stereo display prototype with multiple focal distances. In: Transactions on Graphics, vol. 23. ACM Press, New York, NY, USA, pp. 804–813.

Atchison, D.A., Woods, R.L., 1997. Subjective depth-of-focus of the eye. In: Optometry and Vision Science, vol. 74. LWW, pp. 511–520.

Bach, M., 2007. The Freiburg visual acuity test-variability unchanged by post-hoc re-analysis. In: Graefe's Archive for Clinical and Experimental Ophthalmology, vol. 245, pp. 965–971.

Bernhard, M., Dell'mour, C., Hecher, M., Stavrakis, E., Wimmer, M., 2014. The effects of fast disparity adjustment in gaze-controlled stereoscopic applications. In: Symposium on Eye Tracking Research and Applications. ACM Press, New York, NY, USA, pp. 111–118.

Blakemore, C., Julesz, B., 1971. Stereoscopic depth aftereffect produced without monocular cues. In: Science, vol. 171. American Association for the Advancement of Science, New York, NY, USA, pp. 286–288.

Blum, T., Wieczorek, M., Aichert, A., Tibrewal, R., Navab, N., 2010. The effect of out-of-focus blur on visual discomfort when using stereo displays. In: International Symposium on Mixed and Augmented Reality. IEEE, Seoul, Korea, pp. 13–17.

Bradley, R.A., Terry, M.E., 1952. Rank analysis of incomplete block designs the method of paired comparisons. Biometrika 39 (3–4), 324–345.

Brooker, J.P., Sharkey, P.M., 2001. Operator performance evaluation of controlled depth of field in a stereographically displayed virtual environment. In: Woods, A.J., Bolas, M.T., Merritt, J.O., Benton, S.A. (Eds.), Stereoscopic Displays and Virtual Reality Systems VIII, vol. 4297. SPIE, San Jose, CA, pp. 408–417.

Burge, J., Geisler, W.S., 2011. Optimal defocus estimation in individual natural images. In: Proceedings of the National Academy of Sciences, vol. 108. pp. 16849–16854.

Chessforeva, 2010. Unity3D chess project. ⟨http://chessforeva.blogspot.ca/2010/10/unity3d-chess-project.html⟩.

Duchowski, A.T., House, D.H., Gestring, J., Congdon, R., Ŝwirski, L., Dodgson, N.A., Krejtz, K., Krejtz, I., 2014a. Comparing estimated gaze depth in virtual and physical environments. In: Symposium on Eye Tracking Research and Applications. ACM Press, New York, NY, USA, pp. 103–110.

Duchowski, A.T., House, D.H., Gestring, J., Wang, R.I., Krejtz, K., Krejtz, I., Mantiuk, R., Bazyluk, B., 2014b. Reducing visual discomfort of 3D stereoscopic displays with gaze-contingent depth-of-field. In: Symposium on Applied Perception. ACM Press, New York, NY, USA, pp. 39–46.

Fukushima, T., Torii, M., Ukai, K., 2009. The relationship between ca/c ratio and individual differences in dynamic accommodative responses while viewing stereoscopic images. In: Journal of Vision, vol. 9. Association for Research in Vision and Ophthalmology, pp. 1–13.

Grossmann, P., 1987. Depth from focus. In: Pattern Recognition Letters, vol. 5. Elsevier, North-Holland, pp. 63–69.

Gullstrand, A., 1910. The optical system of the eye. In: von Helmholtz, H. (Ed.), Helmholtz's Treatise on Physiological Optics, vol. 1. J.P.C. Southall, New York, NY, USA, pp. 350–358.

Hayashi, K., Hayashi, H., Nakao, F., Hayashi, F., 2003. Aging changes in apparent accommodation in eyes with a monofocal intraocular lens. In: American Journal of Ophthalmology, vol. 135. Elsevier, New York, NY, pp. 432–436.

Held, R.T., Cooper, E.A., Banks, M.S., 2012. Blur and disparity are complementary cues to depth. In: Current Biology, vol. 22. Elsevier, pp. 426–431.

Hillaire, S., Lécuyer, A., Cozot, R., Casiez, G., 2008a. Depth-of-field blur effects for first-person navigation in virtual environments. In: Computer Graphics and Applications, vol. 28. IEEE, New York, NY, pp. 47–55.

Hillaire, S., Lécuyer, A., Cozot, R., Casiez, G., 2008b. Using an eye-tracking system to improve camera motions and depth-of-field blur effects in virtual environments. In: Virtual Reality. IEEE, Reno, NE, pp. 47–50.

Hoffman, D.M., Banks, M.S., 2010. Focus information is used to interpret binocular images. In: Journal of Vision, vol. 10. Association for Research in Vision and Ophthalmology, pp. 1–17.

Hoffman, D.M., Girshick, A.R., Akeley, K., Banks, M.S., 2008. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. In: Journal of Vision, vol. 8. Association for Research in Vision and Ophthalmology, pp. 1–30.

Howard, I.P., 2002. Seeing in Depth. Basic Mechanisms vol. 1. University of Toronto Press, Toronto, Canada.

Inoue, T., Ohzu, H., 1997. Accommodative responses to stereoscopic three-dimensional display. In: Applied Optics, vol. 36. Optical Society of America, pp. 4509–4515.

ISO 9241-302, 1992. Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) – Part 3: Visual Display Requirements. Technical Report.

Kim, D., Choi, S., Park, S., Sohn, K., 2011. Stereoscopic visual fatigue measurement based on fusional response curve and eye-blinks. In: Digital Signal Processing. IEEE, Corfu, Greece, pp. 1–6.

Kooi, F.L., Toet, A., 2004. Visual comfort of binocular and 3D displays. In: Displays, vol. 25. Elsevier, Soesterberg, Netherland, pp. 99–108.

Koretz, J.F., Cook, C.A., Kaufman, P.L., 2002. Aging of the human lens: changes in lens shape upon accommodation and with accommodative loss. In: Journal of the Optical Society of America, vol. 19. pp. 144–151.

Koretz, J.F., Kaufman, P.L., Neider, M.W., Goeckner, P.A., 1989. Accommodation and presbyopia in the human eye – aging of the anterior segment. In: Vision Research, vol. 29. Elsevier, Kidlington, UK, pp. 1685–1692.

Kuze, J., Ukai, K., 2008. Subjective evaluation of visual fatigue caused by motion images. In: Displays, vol. 29. Elsevier, Amsterdam, Netherland, pp. 159–166.

Lambooij, M., 2009. Measuring visual discomfort associated with 3D displays. In: Stereoscopic Displays and Applications XX, vol. 7237. SPIE, pp. 72370K–1–72370K–12.

Lambooij, M., Fortuin, M., Heynderickx, I., IJsselsteijn, W., 2009. Visual discomfort and visual fatigue of stereoscopic displays: a review. In: Journal of Imaging Science and Technology, San Jose, CA, vol. 53. pp. 030201–1–030201–14.

Langer, M.S., Siciliano, R.A., 2014. Are blur and disparity complementary cues to depth? In: Vision Research, vol. 107. Elsevier, pp. 15–21.

Leroy, L., Fuchs, P., Moreau, G., 2012. Real-time adaptive blur for reducing eye strain in stereoscopic displays. In: Transactions on Applied Perception, vol. 9. ACM Press, New York, NY, USA, pp. 1–18.

Long, N., Over, R., 1973. Stereoscopic depth aftereffects with random-dot patterns. In: Vision Research, vol. 13. Elsevier, pp. 1283–1287.

Luebke, D.P., 2003. Level of Detail for 3D Graphics. Morgan Kaufmann, San Francisco, CA, USA.

Mantiuk, R., Bazyluk, B., Tomaszewska, A., 2011. Gaze-dependent depth-of-field effect rendering in virtual environments. In: Ma, M., Oliveira M.F., João., M.P. (Eds.), Serious Games Development and Applications. Lecture Notes in Computer Science, vol. 6944. Springer, Berlin, Heidelberg, Germany, pp. 1–12.

Marshall, J.A., Burbeck, C.A., Ariely, D., Rolland, J.P., Martin, K.E., 1996. Occlusion edge blur: a cue to relative visual depth. In: Journal of the Optical Society of America, vol. A. pp. 681–688.

Mather, G., 1997. The use of image blur as a depth cue. In: Perception, vol. 26. Pion Ltd, pp. 1147–1158.

Mather, G., Smith, D., 2000. Depth cue integration: stereopsis and image blur. In: Vision Research, vol. 40. Elsevier, pp. 3501–3506.

Mauderer, M., Conte, S., Nacenta, M.A., Dhanraj, V., 2014. Depth perception with gaze-contingent depth of field. In: Conference on Human Factors in Computing Systems. ACM Press, New York, NY, USA, pp. 217–226.

Maurin, H., Sonnenwald, D.H., Cairns, B., Manning, J.E., Freid, E.B., Fuchs, H., 2006. Exploring gender differences in perceptions of 3d telepresence collaboration technology: an example from emergency medical care. In: Nordic Conference on Human-Computer Interaction. ACM Press, New York, New York, USA, pp. 381–384.

Mitchell, D.E., 1966. Retinal disparity and diplopia. In: Vision Research, vol. 6. Elsevier, pp. 441–451.

Mon-Williams, M., Wann, J.P., 1998. Binocular virtual reality displays: When problems do and don't occur. In: Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 40. SAGE Publications, pp. 42–49.

Mon-Williams, M., Wann, J.P., Rushton, S., 1993. Binocular vision in a virtual world: visual deficits following the wearing of a head-mounted display. In: Ophthalmic and Physiological Optics, vol. 13. Wiley Online Library, pp. 387–391.

Mordi, J.A., Ciuffreda, K.J., 1998. Static aspects of accommodation: age and presbyopia. In: Vision Research, vol. 38. Elsevier, pp. 1643–1653.

Nguyen, V.A., Howard, I.P., Allison, R.S., 2005. Detection of the depth order of defocused images. In: Vision Research, vol. 45. Elsevier, pp. 1003–1011.

O'Hare, L., Hibbard, P.B., 2013. Visual discomfort and blur. In: Journal of Vision, vol. 13. Association for Research in Vision and Ophthalmology, pp. 1–12.

O'Hare, L., Zhang, T., Nefs, H.T., Hibbard, P.B., 2013. Visual discomfort and depth-of-field. In: i-Perception, vol. 4. Pion Publications, pp. 156–169.

Okada, Y., Ukai, K., Wolffsohn, J.S., Gilmartin, B., Iijima, A., Bando, T., 2006. Target spatial frequency determines the response to conflicting defocus-and convergence-driven accommodative stimuli. In: Vision Research, vol. 46. Elsevier, pp. 475–484.

Peli, E., 1998. The visual effects of head-mounted display (hmd) are not distinguishable from those of desk-top computer display. In: Vision Research, vol. 38. Elsevier, pp. 2053–2066.

Pentland, A.P., 1987. A new sense for depth of field. In: Pattern Analysis and Machine Intelligence, vol. 4. IEEE, San Francisco, CA, USA, pp. 523–531.

Ramsdale, C., Charman, W.N., 1989. A longitudinal study of the changes in the static accommodation response. In: Ophthalmic and Physiological Optics, vol. 9. Wiley Online Library, Great Britan, pp. 255–263.

Schor, C.M., 1979. The relationship between fusional vergence eye movements and fixation disparity. In: Vision Research, vol. 19. Elsevier, Great Britan, pp. 1359–1367.

Sheedy, J., Bergstrom, N., 2002. Performance and comfort on near-eye computer displays. In: Optometry and Vision Science, vol. 79. LWW, pp. 306–312.

Sheedy, J.E., Hayes, J.N., Engle, J., 2003. Is all asthenopia the same? In: Optometry and Vision Science, vol. 80. LWW, pp. 732–739.

Spear, P.D., 1993. Neural bases of visual deficits during aging. In: Vision Research, vol. 33. Elsevier, pp. 2589–2609.

Stern, J.A., Boyer, D., Schroeder, D., 1994. Blink rate: a possible measure of fatigue. In: The Journal of the Human Factors and Ergonomics Society, vol. 36. Sage Publications, pp. 284–297.

Sucharov, L., 1998. An Investigation of New Methods of Creating Three-Dimensional Multiplanar Displays. University of Oxford.

Suyama, S., Takada, H., Uehira, K., Sakai, S., Ohtsuka, S., 2000. Novel direct-vision 3-d display using luminance-modulated two 2-d images displayed at different depths. In: SID Symposium Digest of Technical Papers, vol. 31. Wiley Online Library, pp. 1208–1211.

Thompson, W., Willemsen, P., Gooch, A., Creem-Regehr, S., Loomis, J., Beall, A., 2004. Does the quality of the computer graphics matter when judging distances in visually immersive environments? In: Presence: Teleoperators and Virtual Environments, vol. 13. MIT Press, pp. 560–571.

Tsirlin, I., Wilcox, L.M., Allison, R.S., 2011. The effect of crosstalk on the perceived depth from disparity and monocular occlusions. In: Broadcasting, vol. 57. pp. 445–453.

Turner, H., Firth, D., 2007. Generalized Nonlinear Models in R: An Overview of the gnm Package. (R Package Version 1.0-6). Technical Report.

Turner, H., Firth, D., 2012. Generalized Nonlinear Models in R: An Overview of the gnm Package. Technical Report.

Ukai, K., Howarth, P.A., 2008. Visual fatigue caused by viewing stereoscopic motion images: background, theories, and observations. In: Displays, vol. 29. Elsevier, pp. 106–116.

Ukai, K., Kato, Y., 2002. The use of video refraction to measure the dynamic properties of the near triad in observers of a 3-d display. In: Ophthalmic and Physiological Optics, vol. 22. Wiley Online Library, pp. 385–388.

van der Linde, I., 2004. Multiresolution image compression using image foveation and simulated depth of field for stereoscopic displays. In: Woods, A.J., Merritt, J. O., Benton, S.A., Bolas, M.T. (Eds.), Electronic Imaging, vol. 5291. SPIE, San Jose, CA, USA, pp. 71–80.

Villarruel, C., 2006. Computer graphics and human depth perception with gaze-contingent depth of field (Ph.D. thesis). Mount Holyoke College.

Vinnikov, M., Allison, R.S., 2013. Gaze-contingent simulations of visual defects in virtual environment: Challenges and limitations. In: CHI 2013 Workshop on Gaze Interaction in the Post-WIMP World.

Vinnikov, M., Allison, R.S., 2014. Gaze-contingent depth of field in realistic scenes: the user experience. In: Symposium on Eye Tracking Research and Applications. ACM Press, New York, NY, pp. 119–126.

Vinnikov, M., Allison, R.S., Swierad, D., 2008. Real-time simulation of visual defects with gaze-contingent display. In: Symposium on Eye Tracking Research and Applications. ACM Press, Savannah, Georgia, pp. 127–130.

Wang, B., Ciuffreda, K.J., 2006. Depth-of-focus of the human eye: theory and clinical implications. In: Survey of Ophthalmology, vol. 51. Elsevier, New York, pp. 75–85.

Wann, J.P., Rushton, S., Mon-Williams, M., 1995. Natural problems for stereoscopic depth perception in virtual environments. In: Vision Research, vol. 35. Elsevier, pp. 2731–2736.

Watt, S.J., Akeley, K., Ernst, M.O., Banks, M.S., 2005. Focus cues affect perceived depth. In: Journal of Vision, vol. 5. Association for Research in Vision and Ophthalmology, pp. 834–862.

Winn, B., Whitaker, D., Elliott, D.B., Phillips, N.J., 1994. Factors affecting light-adapted pupil size in normal human subjects. In: Investigative Ophthalmology and Visual Science, vol. 35. The Association for Research in Vision and Ophthalmology, Rockville, MD, pp. 1132–1137.

Yang, S., Sheedy, J.E., 2011. Effects of vergence and accommodative responses on viewer's comfort in viewing 3D stimuli. In: Stereoscopic Displays and Applications XXII. International Society for Optics and Photonics, San Francisco, California, pp. 78630Q–78630Q.

Yano, S., Emoto, M., Mitsuhashi, T., 2004. Two factors in visual fatigue caused by stereoscopic hdtv images. In: Displays, vol. 25. Elsevier, pp. 141–150.

Zhang, T., O'hare, L., Hibbard, P.B., Nefs, H.T., Heynderickx, I., 2014. Depth of field affects perceived depth in stereographs. In: Transactions on Applied Perception, vol. 11. ACM Press, New York, NY, pp. 1–18.

Zone, R., 2007. Stereoscopic Cinema and The Origins of 3-D Film. University Press of Kentucky, Lexington, KY.