

The Effects of Fast Disparity Adjustment in Gaze-Controlled Stereoscopic Applications

Matthias Bernhard*
Vienna University of Technology

Efstathios Stavrakis
University of Cyprus

Camillo Dell'mour
Vienna University of Technology

Michael Wimmer
Vienna University of Technology

Michael Hecher
Vienna University of Technology

Abstract

With the emergence of affordable 3D displays, stereoscopy is becoming a commodity. However, often users report discomfort even after brief exposures to stereo content. One of the main reasons is the conflict between vergence and accommodation that is caused by 3D displays. We investigate dynamic adjustment of stereo parameters in a scene using gaze data in order to reduce discomfort. In a user study, we measured stereo fusion times after abrupt manipulation of disparities using gaze data. We found that gaze-controlled manipulation of disparities can lower fusion times for large disparities. In addition we found that gaze-controlled disparity adjustment should be applied in a personalized manner and ideally performed only at the extremities or outside the comfort zone of subjects. These results provide important insight on the problems associated with fast disparity manipulation and are essential for developing appealing gaze-contingent and gaze-controlled applications.

CR Categories: I.3.7 [Computer Graphics]: Picture/Image Generation—Display algorithms; I.3.7 [Computer Graphics]: Methodology and Techniques—Ergonomics

Keywords: stereoscopic rendering, comfort models, fusion time, eye tracking

1 Introduction

With the emergence of affordable 3D displays, stereoscopy is becoming a commodity in the film and games industries, virtual reality, visualization, etc. However, despite the technological advancements in 3D display technologies, users often report discomfort and fatigue even after brief exposures to stereoscopic content. One of the main reasons is known to be the conflict between vergence and accommodation that is caused by 3D displays [Lambooi et al. 2007; Howarth 2011], since the eyes naturally turn (vergence) at the distance the virtual stimuli is presented at, but the eyes' focus mechanism (accommodation) remain on the surface of the display.

The most effective way to increase stereo comfort in stereo displays is to control binocular disparities such that they fall in a disparity range where stereo viewing is comfortable. This can be achieved by either controlling stereo camera parameters (i.e., focal distance or interaxial separation) [Heinzle et al. 2011; Oskam et al. 2011; Tseng et al. 2012], or by manipulating binocular disparities in stereo images [Pritch et al. 2000; Wang and Sawchuk 2008; Lang

et al. 2010]. However, manipulating disparities often leads to a significant reduction of depth quality, for example in the form of depth flattening, known also as the cardboard effect. Another way to decrease discomfort is to use depth-of-field (DOF) blurring [Talmi and Liu 1999; Blohm et al. 1997; Blum et al. 2010; Leroy et al. 2012], which simulates the foveal sharpness of the image at the point of focus and blurring observed in peripheral vision. However, applying DOF blurring to accurately simulate the perceived retinal image is difficult to achieve in highly dynamic applications and often causes loss of visual detail.

A promising solution is offered by the use of eye-tracking data, which is used to dynamically adjust the stereo focal plane to coincide with the user's intended focal depth. This is achieved by mapping gaze deployed on a display to the underlying geometric objects of a 3D scene [Tanriverdi and Jacob 2000; Duchowski et al. 2001]. Fisker et al. [2013] performed an informal experiment to evaluate the level of comfort these methods offer to the users. Their preliminary results indicate they can have a positive effect, but a formal experiment and validation is required.

In this work, we performed such a formal experiment in which we investigate the possibility of mitigating visual discomfort by dynamically adjusting the stereo parameters of a scene using gaze data. We use an eye tracker to determine the depth at which the user's eyes converge and manipulate the stereo parameters to bring it into focus by setting the plane of zero disparities at that depth. To measure the level of discomfort, we measured binocular fusion times, which is an *objective* metric. In particular, we measure fusion times of random dot stereograms in a scene with a three-dimensional arrangement of objects, which resembles a natural scene with multiple depth layers where a user can freely direct gaze from one object to another. We utilize a QUEST [Watson and Pelli 1983] procedure to determine the amount of time the user would require to fuse a random dot stereogram (RDS) showing one of two possible orientations of a wave pattern, shown in Figure 1. After correct responses the QUEST reduce the display time of the RDS, whereas wrong responses lead to an increase of the time the RDS is shown. Effectively, the QUEST is searching for the minimum amount of time a user requires to fuse the stereogram, which we refer to as *fusion time*. We found that gaze-controlled manipulation of the plane of zero parallax can lower fusion times for large disparities. In addition, a small overhead in fusion time was observed after adjusting the plane of zero parallax to coincide with disparities that initially fell within the stereo viewing comfort zone of subjects. Since the comfort zone differs among subjects, this means that gaze-controlled disparity adjustment should be applied in a personalized manner and ideally performed only at the extremities or outside the comfort zone of subjects.

The contributions of this work are: (a) an *objective* methodology to evaluate the benefits of visual comfort-optimizing strategies for 3D stereo displays, and (b) experimental evidence that gaze-controlled dynamic adjustment of stereo disparities improves stereo fusion of nearby objects (large disparities), while it is acceptable for objects at other distances.

*e-mail: matthias.bernhard@cg.tuwien.ac.at

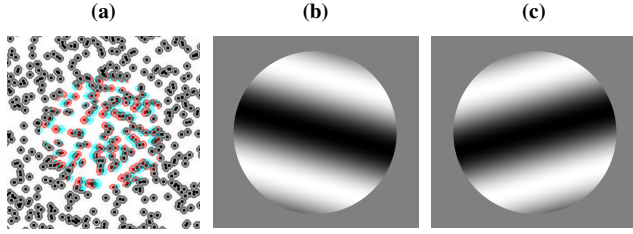


Figure 1: The random dot stereograms shown by the QUESTs for a controlled time: (a) shows an anaglyph rendering. (b) and (c) show a visualization of the two different sine wave patterns participants had to distinguish.

2 Related Work

2.1 Stereo Viewing Comfort

The main factors [Lambooy et al. 2007; Patterson 2009; Howarth 2011; Tam et al. 2011] causing viewing discomfort in stereoscopic displays are (a) *dichoptic errors*, due to stereoscopic distortions, mismatch of the stereoscopic windows or crosstalk, (b) *fast object motion*, due to difficulties of the human visual system (HVS) to adapt to changing viewing conditions, and (c) uncomfortable disparity ranges due to excessive binocular parallax, or due to a mismatch of vergence and accommodation (V/A-conflict). While factors of category (a) are a matter of hardware, processing or bandwidth limitations, the other sources of discomfort depend on *content and rendering* and could possibly be mitigated by adjusting stereo configuration parameters. To find the best parameters, a function which predicts viewing comfort is required, which is obtained by taking measurements in a perceptual study. Discomfort is usually predicted as a function of disparity, or as a bivariate function of vergence and focal distance. Measurements can be taken with *objective and subjective* observation methods. Assuming correlations between performance in binocular perception and comfort, objective measurements are taken by determining speed or acuity in the detection of binocular corrugations (e.g. inward versus outward curvature) in random dot stereo-grams [Hoffman et al. 2008; Kim et al. 2011; Kim et al. 2012]. For subjective measurements, users are usually asked to rate the discomfort they experienced on a Likert scale [Kim et al. 2011]. Due to methodological limitations and simplicity, comfort-optimizing stereo displays are usually evaluated with subjective methods only. Responses can be collected either by *rating* methods, where a user issues a score on a Likert scale [Blohm et al. 1997; Sun and Holliman 2009; Blum et al. 2010; Koppal et al. 2011; Liu et al. 2011; Ju Jung et al. 2012], or with the method of pairwise comparison [Koppal et al. 2011; Leroy et al. 2012]. An *objective* way to observe discomfort was used by [Cho and Kang 2012], who counted the eye-blink rate of users.

In this work we will predominantly focus on an *objective* evaluation. Hence we will measure performance in the fusion of RDS, similar to [Hoffman et al. 2008; Kim et al. 2011]. However, in contrast to previous work which measured fusion times for a stimulus showing a single disk with a RDS, we will place the RDS on objects in a scene arranged in three dimensions. In addition to that, we handed out a rating scale where participants cast a subjective judgment on stereo comfort and stereo depth.

2.2 Eye Tracking in 3D

Hillaire et al. [2008], and very recently Mantiuk et al. [2013], proposed to use an eye-tracker to control DOF simulations in in-

teractive applications. Eye tracking was also used in applications attempting to reduce stereo viewing discomfort by DOF-blurring [Talmi and Liu 1999; Blum et al. 2010; Leroy et al. 2012]. In all cases, gaze data was mapped to scene content (e.g., objects [Mantiuk et al. 2013]) in order to determine the focal distance. Alternatively, for applications which do not have access to a 3D representation of the stimulus (e.g., movies), [Duchowski et al. 2011] and [Pfeiffer 2012] proposed to utilize binocular gaze data to determine a depth coordinate for each gaze point.

However, a reliable and accurate determination of a user’s gaze depth is still an open problem. To avoid this problem in our experiment we used a simplified scene which was inspired by [Duchowski et al. 2011]. In this scene we placed 9 squares such that the current gaze depth can be robustly determined from the depth buffer also in the presence of eye tracker errors and a participant’s inaccurate gaze orienting behavior.

3 Experiment

We performed an experiment to evaluate the effects of dynamic adjustment of disparities in 3D applications using eye-tracking data. In particular, we consider binocular fusion times as an objective indicator of visual comfort and asked participants to focus at objects lying at different depths while stereo viewing a 3D scene. We measured fusion times of stereoscopic stimuli, where binocular disparities were manipulated dynamically, using gaze data recorded in real-time. Disparity was automatically adjusted by modifying stereo parameters in order to bring the currently attended object into focus, which equates to setting zero parallax between the two stereo views at that point. We denote this condition as *dynamic stereo (DS)* and compare fusion times to a control condition in which predefined stereo disparities remain constant, which we refer to as *static stereo (SS)*.

3.1 Participants

We recruited 38 subjects (15 female) with normal or corrected to normal vision, aged between 18 and 40 years (mean=26.3, stdev=5.4), all of which stated that they were stereoacute and not color blind. Each participant was pre-screened during a training block of trials to objectively measure their ability to achieve stereopsis. Through this procedure, we identified 2 stereo-blind participants, who were excluded.

3.2 Stimulus

To sample a variety of factors that might have an influence on fusion time, such as *position* or *disparity*, the 3D scene we used for the experiment consisted of 9 square planes, placed in three rows and columns (3×3) on a black background, as shown in Figure 2. Each row of squares was positioned at a different distance/depth, with the bottom row being the closest (**NEAR**), the middle row being at an intermediate distance (**MID**), and the top row being the farthest (**FAR**). In the static stereo condition (**SS**), the **MID** row was at the same distance as that of the focal plane (i.e. the plane of zero parallax). Consequently, the **FAR** row had positive disparities ($d = 0.7^\circ$) and appeared to be 14cm behind the focal plane, while the **NEAR** row had negative disparities of $d = -1.2^\circ$ and appeared 14cm in front of the focal plane.

In their initial state, all squares were colored red and were textured with a pattern reminiscent of those used in random dot stereograms. However, since the same pattern was rendered for both stereo views, it provided no depth cues. The color of one of the squares was dynamically modified to attract attention during the experiment, while

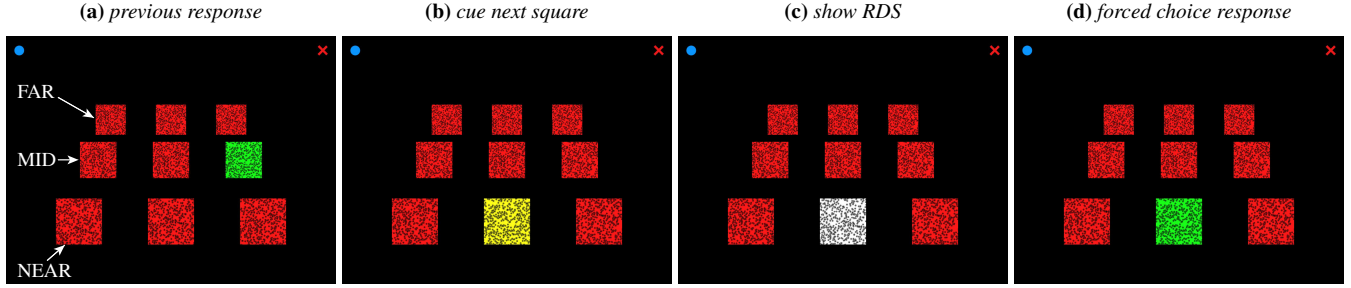


Figure 2: One QUEST trial: (a) after responding in the previous trial, (b) the next square is cued yellow, (c) as soon the eye tracker senses a user’s gaze near the cued square the RDS is shown on white for a controlled time after which (d) the square turns green with random dots.

as soon a fixation was detected near the cued target the texture of that square was replaced by another texture containing a circular random dot stereogram (RDS) as shown in Figure 1. This RDS that would elicit binocular perception of a sine wave with a wavelength of 2° and an amplitude of 0.65° which was either tilted by $+15^\circ$ or -15° from the horizontal axis (see Figure 1). We also ensured that the rendered stereograms appeared with the same size ($radius = 1.9^\circ$), the same resolution and the same disparity amplitude on all squares by avoiding distortions of the stereogram patterns due to perspective projection.

3.3 Task

Participants were instructed to focus their gaze on one dynamically chosen square with a yellow color cue. As soon as the eye tracker could detect that a participant initiated a fixation (with two gaze point samples recorded at 50Hz) on the cued square, its color changed to white, and simultaneously the texture was replaced with a random dot stereogram. In stereo, this stereogram induced the perception of a 3D sine wave with a left- or right-winded slope on the square. The participant was exposed to the random dot stereogram (where the random dots changed in each QUEST) for a short period of time, which was controlled by a QUEST procedure, and then switched back to a different random dot pattern that induced no depth perception. Using the left and right cursor keys of a keyboard, participants had to respond whether the sine wave had left- or right-winded tilt. After the response, the next square cued the participant’s attention by switching to yellow. The experiment ended when all of the running QUEST procedures were completed with a minimum of 3 turns.

3.4 QUEST Conditions

We defined 12 QUEST procedures to determine the effects of a variety of factors that may play a role in fusion times, such as position, depth, and eye movements. For each of the three rows (**N**, **M** and **F**) of squares, we make a distinction between squares located in the center (Central) and the squares on the left- and right-hand side of the row (Lateral). We differentiate eye movements between two consecutively attended squares. In particular, we distinguish short binocular saccades between two consecutive rows and long binocular saccades between the last and first row. We also distinguish backward saccades from far to near squares, and forward saccades from near to far squares. For each possible combination of these categories, we use a different QUEST instance to measure the respective fusion time. Note that 50% of the 24 theoretically possible combinations are not possible in our scene, e.g., a *backward* saccade to the *near* row is an invalid case. Thus there are 12 QUESTs in total (**NCsb**, **NClb**, **NLsb**, **NLib**, **MCsf**, **MCsb**, **MLsf**, **MLsb**,

FClf, **FCsf**, **FLsf** and **FLlf**). All these QUESTs were run in parallel and scheduled in randomized order. In each trial, a QUEST was started after sensing two gaze points near the cued square plane. Since the mechanism to launch the QUEST was equal in both of the **SS** and **DS** conditions, the latency required to detect a fixation should not affect the measurement process itself.

We used the Matlab Psycho Toolbox to control the QUESTs. The threshold guess was set to $t_G = 1000ms$ (which is used as initial test time), the standard deviation guess to $t_{Gsd} = 1000ms$, and the probability threshold to $p_T = 82\%$. The gamma parameter was set to $\gamma = 50\%$ and the delta parameter was $\delta = 0.01$. As beta parameter we used $\beta = 5.4$, which was optimized using data obtained by one of the authors performing a beta analysis over 800 trials of the experiment run in condition **SS**.

3.5 Procedure

Each participant was assigned to perform two blocks of multiple trials for each of the two conditions (**SS** and **DS**). The experimental condition of each block was randomly selected to counter balance potential effects of ordering. The duration of each block was 20 minutes on average since it depended on the convergence time of all QUESTs, contributing to a total experiment duration of 40 minutes in average for both blocks. Each of the 12 QUESTs converged, in average, after 65 trials summing up to approximately 800 trials in total for each block and participant. To prevent eye-strain and fatigue the procedure was paused automatically every 5 minutes for a fixed 1-minute break before continuing the block. Prior to the two main blocks, a short block of 27 trials was used as a training block for the participants to familiarize themselves with the task. During training, a participant viewed the RDS on each of the 9 squares in the scene for one second and had to identify the orientation of the sine-wave tilt. This procedure was repeated three times in the training block.

3.6 Subjective Measurements

Besides the fusion times measured with the QUEST procedures, we also took *subjective* measurements according to the ITU-Recommendation for subjective methods for the assessment of stereoscopic 3DTV systems [International Telecommunication Union 2012]. This recommends a 5-level rating scale where the user scores his experience with respect to *picture quality*, *depth quality* and *visual comfort*. We told participants that the levels serve an orientation and the ratings can be done on a continuous scale. After each block, we asked participants to rate depth quality and visual comfort only, since our scene had no characteristics that could be assessed for a reliable picture quality rating.

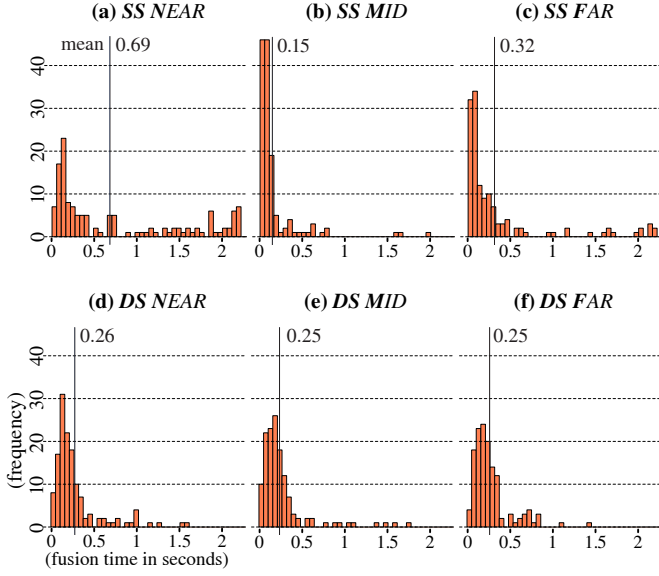


Figure 3: Distribution of fusion times sorted by stereo condition (rows) and depth plane (columns). For each histogram the abscissa denotes fusion time in seconds and the ordinate shows the occurrence frequency. Black lines mark the means.

3.7 Setup and Configuration

The hardware setup in use was a See Real Technology Cn 201.05 Autostereoscopic 20" monitor which uses a vertical interlace technique to display stereo image pairs. The resolution of each image was 800×1200 pixels, but due to the interlace technique they were shown with a 4 : 3 ratio. To avoid reflections in the display, the experiment was carried out in a room without daylight under dim light conditions. We used a Tobii x50 eye-tracker, which was placed in front of the display. Due to the display's refresh rate of 60Hz, the eye tracker's sampling rate of 50Hz and the eye tracker's latency of 35ms, stereo disparity adjustments were performed with a delay¹ ranging, depending on the sampling phase of the eye-tracker, between a minimum of 72ms (best case) and a maximum of 92ms (worst case). Configuration and a 5-point calibration was carried out for each participant before the training block using the software provided by the manufacturer. In addition, we used a chin-rest placed 67cm away from the display's center. Besides allowing faster and more accurate gaze sampling, the chin-rest's main purpose was to prevent participants from moving their eyes away from the auto-stereo display's optimal 3D view point.

4 Results

4.1 Distribution of Fusion Times

We use descriptive statistics to analyze the main effect of the two conditions. We computed the fusion times' histograms shown in Figure 3. Since the effects vary considerably with the distance of the focused row of squares, we split the data for all rows. Nevertheless, the distribution of the fusion time in the DS condition is very similar for each row and also consistently of lower variance. To analyze the overall effect of condition and its significance, we performed a paired Wilcoxon test, a non-parametric alternative to

¹best case: 35ms(latency)+1 × 20ms(sampling)+17ms(display)
worst case: 35ms(latency)+2 × 20ms(sampling)+17ms(display)

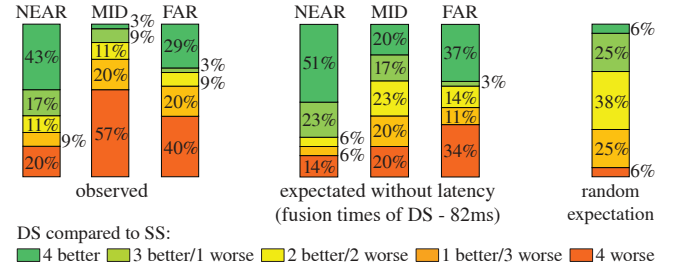


Figure 4: Visualization of the percentage of participants who had a benefit from dynamic stereo. We grouped users according to the number of QUESTs having a shorter fusion time in DS compared to SS. In the second group, we added results we expect from an ideal disparity adjustment without a technical latency. The last chart shows the expected result of a random process.

the Student's *t*-test, which could not be used for our data due to violating the assumptions of normality and equality of variance. The test results are shown in Table 1.

Concerning the effect of condition on fusion time, we found a highly significant positive effect for DS only in the NEAR row. The overall effect for the MID row was, however, a significant increase of fusion times due to dynamic switching of the focal plane. Comparing the results of both conditions in the MID row, where we had in both conditions zero disparity, reveals that focal plane switching causes an overhead in fusion time which was in average 94ms (mean difference between DS and SS), what lies close to the average delay (82ms) in disparity adjustments caused by gaze sensing and the display refresh rate. To some extent, the overhead explains why the effect of DS was more on the negative side for the NEAR row, where many users had relatively short fusion times in the SS condition. However, the benefits of DS seem to vary considerably among participants, as discussed next.

4.2 User Group Analysis

One important result of this experiment is that fusion times vary considerably with users. This suggests that stereoscopic fusion which occurs within the stereo viewing comfort zone are idiosyncratic characteristics of each participant's visual system. To investigate on how many users DS is beneficial, we counted for each row the amount of corresponding QUESTs which had a better fusion time in DS rather than in SS. Since each participant performed 4 QUESTs in each row, we visualized the result with pie charts with five sectors, each corresponding to one of the possible outcomes (1, 2, 3 or 4 better in DS, or all being worse) for each participant. We show the results in Figure 4 for each row of squares. In addition, we include a fourth pie chart which shows the expected result for a random process. The charts illustrate that a majority of 60% of users benefit from DS in the NEAR row (3/4 or 4/4 of the fusion times better in DS than SS) which is more than double the percentage (31%) expected by a random process. For the MID row, however, there is a significant majority of 77% which has a disadvantage (0/4 or 1/4) from DS and a significantly smaller group having an advantage of DS than expected by chance. In the FAR row the re-

	SS vs. DS	Effect
NEAR	$p < .0001^{****}$	DS < SS
MID	$p < .0001^{****}$	SS < DS
FAR	$p = .1353$	-

Table 1: The effect of SS versus DS for NEAR, MID, and FAR row.

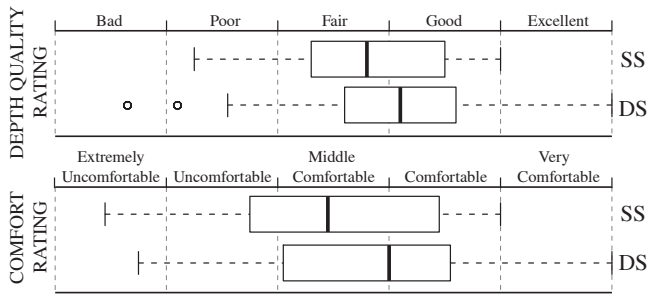


Figure 5: Distribution of subjective comfort and depth quality ratings visualized as Boxplots.

sults are more ambivalent. The percentage of users (32%) having an advantage with **DS** lies within the range of a random process (31%), while the fraction of those benefiting from **SS** (63%) is significantly higher than expected by chance. However, looking at the group of 29% of all participants having a maximum benefit (4/4), we also see a significant deviation from a random expectation (6%). Overall, these results suggest that the user’s comfort zone is personalized and is related to the positive or negative impact that dynamic focal plane manipulation may have on him.

To gain insight on what can be expected from an optimal disparity adjustment performed with zero latency, we simulated a ideal case by subtracting the average latency of 82ms from the fusion times observed in the **DS** condition. However, while the amount of users having a benefit from **DS** increases considerably in all three rows, there is still a significant percentage of users who saw the maximum disadvantage from **DS** in the **MID** and **FAR** rows. Particularly for the **FAR** row there appears to still be some overhead probably resulting from other factors other than delayed disparity adjustments.

4.3 Subjective Measurements

The analysis of user ratings is performed using descriptive statistics. Results are shown in Figure 5 where we visualized the distribution of the subjective comfort and quality ratings. For the comfort ratings we obtained a positive effect of using **DS** compared to **SS**, albeit with a weak significance (paired two-tailed Wilcoxon T test, $V = 174$, $p = 0.09$). Overall 50% of all participants rated **DS** to be “comfortable” at least, while the median for **SS** is at the “middle comfortable” level. A slight positive shift from **SS** to **DS** was also observed in the quality ratings, but the effect is not significant (Wilcoxon T test, $V = 187.5$, $p = 0.16$). Since **DS** actually distorts the relation of disparities within a scene of objects, no decrease in perceived quality can be also interpreted as a result favoring the application of **DS** if viewing comfort can be increased.

4.4 Other Factors Influencing Fusion Times

Besides disparity, we contemplated that fusion times may also be affected by the position of the square or the type of saccade prior to the fusion process. We performed a factorial analysis to determine the effects of depth (**NEAR** vs **MID** vs **FAR**), saccade length (long vs short) and direction (forward vs backward), and azimuth (central vs lateral). We first analyzed the effect of depth using a Friedman test (the non-parametric equivalent to a 2-way ANOVA) which allows replicates (i.e., 4). However, the Friedman test is not applicable for a multi-factorial analysis. Assuming that depth is the most dominant factor, we split the data for each depth (**N,M,F**). Since there are only two independent variables per row which have only two categories, we could use a Wilcoxon test. In particular, we

		SS	Effect	DS	Effect
	Depth	$p=10^{-26}***$	$M < F < N$	$p=.069^*$	$(F < N < M)$
FAR	s.length	$p=.0013***$	$s < l$	$p=.48$	–
	azimuth	$p=.20$	–	$p=.95$	–
MID	s.dir.	$p=.04^{**}$	$f < b$	$p=.65$	–
	azimuth	$p=.014^{**}$	$c < l$	$p=.37$	–
NEAR	s.length	$p=.77$	–	$p=.067^*$	$(s < l)$
	azimuth	$p=.17$	–	$p=.056^*$	$(c < l)$

Table 2: Factorial analysis of rank-transformed fusion times.

performed a paired Wilcoxon test on the single factors, i.e., saccade length, saccade direction and azimuth. To compensate potential effects resulting from interactions of the two factors present in each row, we sorted in the evaluation of one factor the data such that the Wilcoxon test always compared pairs corresponding to the same category of the other factor.

The results which are listed in Table 2 show that the lower variance of fusion times in the **DS** condition is due to a clear reduction of effects resulting from object distance, saccade length and direction. These factors are more pronounced in the **SS** condition. There we found that fusion times are significantly lower when a user performs a “short” saccade from the **MID** to the **FAR** row compared to a “long” saccade from the **NEAR** to the **FAR** row, though this does not apply for saccades in the backward direction. Furthermore, we found a significant fusion time advantage at the object located in the center of the **MID** row and for forward saccades compared to backward saccades. Overall, these results suggest that fusion times are also affected by prior gaze movement. Other effects of preceding saccades become apparent in the gaze analysis which we discuss next.

4.5 Gaze Analysis

By recording binocular gaze data throughout the experiment, we can also analyze the eye vergence behavior. To this end, we extracted the gaze signal of the first fixation each time the fusion task started and computed the screen distance between left and right gaze point in degrees of visual angle, which we further denote as *gaze disparity*. Since this requires an additional so-called depth calibration [Duchowski et al. 2011], we computed for each participant a correcting shift such that the average gaze disparity observed when a user fixates the plane of zero parallax in the (disparity manipulation free) **SS** condition fits to zero.

We analyzed this data by plotting the average gaze disparity of over 2000 fixations recorded during the experiment for each QUEST condition as a function of time. Averaging many fixations turned out to be a good remedy against the strong noise in vergence measurements taken with a conventional eye tracker (cf. [Duchowski et al. 2011]). We temporally aligned all gaze samples such that the fixation begin time corresponds to $t = 0$ and averaged for each sampling time interval ($\Delta t = 20ms$) the gaze disparities. The results corresponding to each QUEST condition are shown in Figure 6. For the **DS** condition we split the gaze data obtained from fixations on the **NEAR** or **FAR** row in two groups: one group contains participants having a fusion time advantage in the **DS** condition (green graph), and the other group includes those who had a disadvantage (red graph). These groups correspond to the green/lime and red/orange fractions shown in the group analysis before, respectively (Figure 4). However, for the **MID** row we have consolidated all data of the **DS** condition into a single graph (blue), since we did not observe a clear by participant clustering as for the other two rows.

In the case of **SS** condition (black), these plots make apparent that

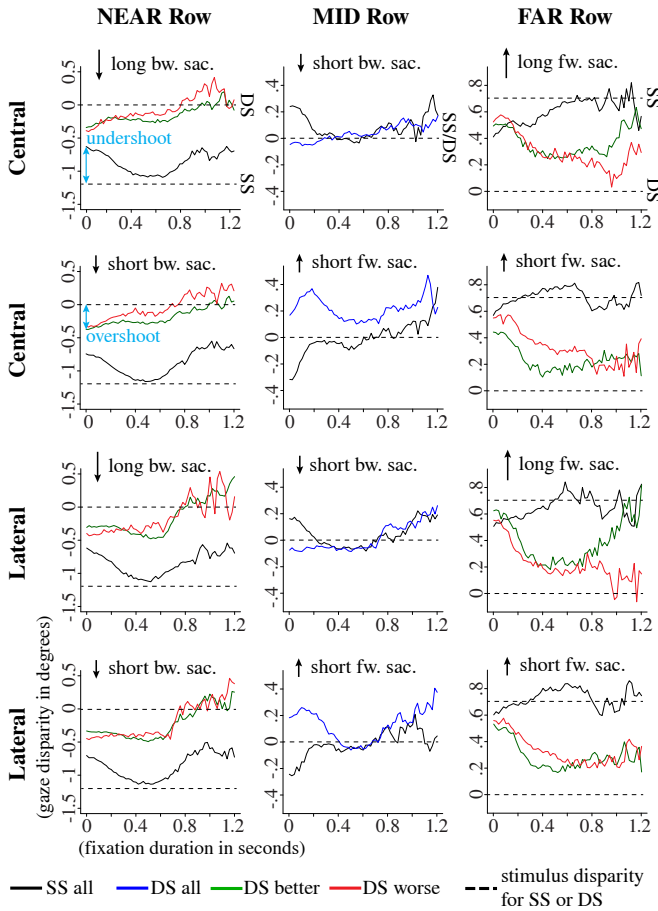


Figure 6: Temporal behavior of gaze disparities. We derive these plots by averaging all fixations corresponding to the same QUEST.

after a saccade the gaze disparity has an **undershoot** relative to the targeted stimulus disparity (dotted line). At fixation time eye rotations accommodate the stimulus disparity. In particular, saccades in forward direction are followed by a divergent eye rotation, and backward saccades by a convergent eye rotation, respectively. Moreover, we observe in the NEAR row after 600ms, that gaze disparity slowly drifts away from the stimulus disparity. A reason for this could be that it is difficult to maintain the alignment between extensive stimulus disparities and the gaze disparity for a longer time due to an **acute** emergence of visual fatigue. As a side-remark, detecting such patterns could be a new way to measure discomfort without measuring fusion time.

In the **DS** condition, the dynamic disparity modifications fundamentally change the behavior pattern observed under consistent disparity conditions. In contrast to **SS**, we observe an **overshoot** of the stimulus disparity after the saccade, which is followed by an accommodating vergence in the opposite depth-direction as the saccade. Another difference is that the stimulus disparity is not accommodated so well as observed in the **SS** condition, what is most obvious in the **FAR** row. Thus, stereo fusion in **DS** seems to often take place with higher retinal disparities requiring to fuse objects nearer to the limits of Panum's fusional area (which is usually in the range of $0^\circ \pm 1^\circ$ of retinal disparity). A reason for this could be that a user's expectations together with monocular depth cues of the scene are in conflict with binocular depth perception. Moreover, extensive peripheral disparities, which occur in the NEAR row when the plane of zero parallax is shifted to the FAR row, might impede

the accommodating vergence to disparities in the **FAR** row.

Looking at the difference between those participants having a fusion advantage in **DS** (green graphs), we can observe the most noticeable difference to those with a disadvantage (red graphs) in fixations following a short saccade to the FAR row, where the green group approaches faster the stimulus disparity. When focusing the central square this behavior is most pronounced and green vergence graphs do also have a smaller overshoot. However, the gaze disparity graphs do not explain why the green group has a fusion advantage after long saccades in the FAR row, or in all fixation on the NEAR row, respectively.

Overall, the gaze analysis results mainly reveal that users perform preparatory vergence movement already during a saccade, which is consistent with findings from vision literature [Westheimer and Mitchell 1969; Howard 1995]. While under natural viewing conditions the target disparity is undershot, **DS** yields an overshoot through manipulating the stimulus disparity.

5 Discussion

We found that gaze-controlled manipulation of the plane of zero parallax can lower fusion times for large disparities. However, there is a delay in fusion time that is attributed to the way we detect *when* to adjust the plane of zero parallax. In this work we did not assume prior knowledge of the user's future gaze target intentionally. Instead we used a strategy that would be more appropriate for replicating the behavior of such a technique in a real-time gaze-contingent application. Our strategy after cuing the next square was to collect two gaze samples in proximity to the target, which constituted the initiation of a fixation. This source of fusion time delay could be technically reduced, for instance by sampling at a higher rate or by methods which can predict future gaze points.

This technical latency does not fully explain the overhead in fusion time that was observed after adjusting the plane of zero parallax. This became particularly apparent in the **FAR** plane, where a large group of users had a longer fusion time overhead than predicted by the latency alone. We believe two reasons may be responsible for this. First, the abrupt change of disparities may require a readjustment of the visual system that introduces a delay in fusion. Second, stereo fusion may be facilitated by sensory hysteresis ([Howard 1995, p. 322]) or peripheral vision processing [Cisarik and Harwerth 2005]. Fast manipulation of the stimuli's disparities may be interfering or inhibiting the function of these processes. Moreover, the results of the gaze analysis suggest that the conflict between anticipated and adjusted disparities after a saccade may have a negative impact on the post-saccadic vergence accommodation. Under natural viewing conditions, vergence eye movements occur already at saccade time and usually undershoot the disparity of the stimulus at the saccadic landing position. This is followed by a slow and more accurate post-saccadic vergence accommodation into the same orientation as the vergence movements occurring during a saccade [Westheimer and Mitchell 1969]. Due to that, disparity manipulations cause an overshoot of the stimulus disparity. Since the vergence accommodation at fixation time has to be performed in the reverse direction as under natural conditions, we speculate that this could be another source of discomfort.

Overall, our methodology allowed an in depth investigation identifying the main problems of fast gaze-contingent disparity adjustments. The implications of our results should mainly affect the design of smart disparity adjustment strategies and the way models for stereo comfort should be designed and measured:

- **Ideal comfort models should account for gaze movement from previous to the next object**, as suggested by the results

of the factorial analysis and the gaze analysis showing that fusion times and vergence behaviors are both affected by the type of the previous saccade.

- Fusion times should not be measured for isolated objects. There is **an effect of disparities in the context** of the currently foveated object [Howard 1995, p. 177], as the results for the FAR row indicate. Thus future work experiments should consider the properties of the context as an additional independent variable. E.g., by adding or removing other objects in a controlled manner.
- Apart from measuring fusion times, the **additional use of gaze data to analyze temporal vergence behavior provides a powerful tool giving deeper insights how discomfort may arise**. We believe that features of temporal vergence signals could be linked to discomfort and potentially allow objective comfort measurements with more natural stimuli than RDSs.
- And finally, **smart disparity adjustments should be performed such that a stimulus disparity overshoot in saccadic vergence movements is avoided**. Such a strategy would be, for instance, to shift the plane of zero parallax to the fixated stimulus distance biased by the expected overshoot.

6 Conclusion and Future Work

We consider this work as an important step towards the development of gaze-controlled disparity in stereoscopic applications. To develop such stereo controllers, it is necessary to devise a computational model that predicts the costs (i.e., expected fusion delay) and benefits (i.e., expected fusion time reduction) of a possible disparity adjustment in order to select the most appropriate configuration by means of maximizing the cost-benefit ratio. Although our results could be implemented as a preliminary instance of such a model (with 3 discrete distances), a more fine-grained model should be derived by further experiments.

For instance, our experiment may be carried out with stimuli arranged in more depth layers, which would allow obtaining samples of fusion times for more than three disparity levels. Furthermore, other disparity-adjustment strategies than fast switching need to be investigated. Such strategies may include smooth transitions and methods which can mask disparity adjustments or even avoid latency. For instance, using high frequency (e.g. 200Hz) eye-trackers with low latency may potentially allow to predict saccade-landing positions at the beginning of a saccade using a ballistic model [Komogortsev and Khan 2008]. Being able to estimate the future gaze position would then allow to perform the disparity adjustments within the time window of a saccade during which the stream of visual processing is disrupted [Paradiso et al. 2012]. However, avoiding latency is a hard problem and requires that it is being traded off against a much lower accuracy in the determination of the attended depth plane. An inaccurate identification of a future gaze location would be serious hindrance for many end-user applications. Furthermore, disparity adjustments performed during a saccade produce a conflict between the disparity a user anticipates and the disparity finally seen on the saccade landing target. This results in an unnatural experience which may cause fusion time delays and is potentially another source of discomfort.

Latency and fusion time delays, as identified in the presented experiment, are not the only problems that should be solved to advance the state-of-the-art in applications with gaze-controlled stereo disparities. For more complex scenes, that are commonplace in end-user commercial applications, it is particularly important to be able to reliably identify the feature or object currently in the user's focus, thus obtaining accurately the depth at which the user is directing

his gaze. In dynamic 3D scenes with a high density of objects and motion (e.g. 3D computer games) identifying the attended depth plane can be very challenging, because visual attention does not always coincide with the exact deployment of gaze (e.g. gaze may be deployed in close proximity to the target). Robust gaze-to-object mapping methods have to be devised for 3D applications, an area in which only very recently methods have emerged [Sundstedt et al. 2013; Mantiuk et al. 2013].

Finally, an ideal model should account for the fact that viewing comfort zones differ among users. Ideally, future end-user applications would use calibration to generate a personalized model by fitting the model's parameters to the perceptual characteristics of each user.

Acknowledgements

This work was funded by the Austrian Science Fond Project FWF P24600-N23.

References

- BLOHM, W., BELDIE, I. P., SCHENKE, K., FAZEL, K., AND PAS-TOOR, S. 1997. Stereoscopic image representation with synthetic depth of field. *Journal of The Society for Information Display* 5.
- BLUM, T., WIECZOREK, M., AICHERT, A., TIBREWAL, R., AND NAVAB, N. 2010. The effect of out-of-focus blur on visual discomfort when using stereo displays. In *2010 9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 13–17.
- CHO, S.-H., AND KANG, H.-B. 2012. An assessment of visual discomfort caused by motion-in-depth in stereoscopic 3d video. In *Proceedings of the British Machine Vision Conference, BMVA Press*, 65.1–65.10.
- CISARIK, P. M., AND HARWERTH, R. S. 2005. Stereoscopic depth magnitude estimation: effects of stimulus spatial frequency and eccentricity. *Behavioural Brain Research* 160, 1 (May), 88–98.
- DUCHOWSKI, A. T., MEDLIN, E., GRAMOPADHYE, A., MELLO, B., AND NAIR, S. 2001. Binocular eye tracking in vr for visual inspection training. In *Proceedings of the ACM symposium on Virtual reality software and technology*, ACM, New York, NY, USA, VRST '01, 1–8.
- DUCHOWSKI, A. T., PELFREY, B., HOUSE, D. H., AND WANG, R. 2011. Measuring gaze depth with an eye tracker during stereoscopic display. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*, ACM, New York, NY, USA, APGV '11, 15–22.
- FISKER, M., GRAM, K., THOMSEN, K. K., VASILAROU, D., AND KRAUS, M. 2013. Automatic Convergence Adjustment for Stereoscopy using Eye Tracking. In *EG 2013 - Posters*, 23–24.
- HEINZLE, S., GREISEN, P., GALLUP, D., CHEN, C., SANER, D., SMOLIC, A., BURG, A., MATUSIK, W., AND GROSS, M. 2011. Computational stereo camera system with programmable control loop. *ACM Trans. Graph.* 30, 4 (July), 94:1–94:10.
- HILLAIRES, S., LÉCUYER, A., COZOT, R., AND CASIEZ, G. 2008. Depth-of-field blur effects for first-person navigation in virtual environments. *IEEE Comput. Graph. Appl.* 28 (November), 47–55.

- HOFFMAN, D. M., GIRSHICK, A. R., AKELEY, K., AND BANKS, M. S. 2008. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision* 8, 3.
- HOWARD, I. P. 1995. *Binocular Vision and Stereopsis*. No. no. 29 in Oxford psychology series. Oxford University Press, New York.
- HOWARTH, P. A. 2011. Potential hazards of viewing 3-d stereoscopic television, cinema and computer games: a review. *Ophthalmic and Physiological Optics* 31, 2, 111–122.
- INTERNATIONAL TELECOMMUNICATION UNION, 2012. Recommendation ITU-R BT.2021: Subjective methods for the assessment of stereoscopic 3DTV systems.
- JU JUNG, Y., LEE, S.-I., SOHN, H., WOOK PARK, H., AND MAN RO, Y. 2012. Visual comfort assessment metric based on salient object motion information in stereoscopic video. *Journal of Electronic Imaging* 21, 1, 011008–1–011008–16.
- KIM, J., HOFFMAN, D. M., AND BANKS, M. S. 2011. The zone of comfort : Predicting visual discomfort with stereo displays takashi shibata. *Journal of Vision* 11, 8, 1–29.
- KIM, D., CHOI, S., AND SOHN, K. 2012. Effect of vergence-accommodation conflict and parallax difference on binocular fusion for random dot stereogram. *IEEE Transactions on Circuits and Systems for Video Technology* 22, 5 (May), 811–816.
- KOMOGORTSEV, O. V., AND KHAN, J. I. 2008. Eye movement prediction by kalman filter with integrated linear horizontal oculomotor plant mechanical model. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, ACM, New York, NY, USA, ETRA '08, 229–236.
- KOPPAL, S., ZITNICK, C., COHEN, M., KANG, S. B., RESSLER, B., AND COLBURN, A. 2011. A viewer-centric editor for 3d movies. *Computer Graphics & Applications, IEEE* 31, 1, 20–35.
- LAMBOOI, M. T. M., IJSSSELSTEIJN, W. A., AND HEYNDERICKX, I. 2007. Visual discomfort in stereoscopic displays: a review. *Proceedings of SPIE* 6490, 1584, 64900I–64900I–13.
- LANG, M., HORNUNG, A., WANG, O., POULAKOS, S., SMOLIC, A., AND GROSS, M. 2010. Nonlinear disparity mapping for stereoscopic 3d. *ACM Trans. on Graphics (Proc. SIGGRAPH)* 29, 3.
- LEROY, L., FUCHS, P., AND MOREAU, G. 2012. Visual fatigue reduction for immersive stereoscopic displays by disparity, content, and focus-point adapted blur. *Industrial Electronics, IEEE Transactions on* 59, 10 (oct.), 3998–4004.
- LIU, C.-W., HUANG, T.-H., CHANG, M.-H., LEE, K.-Y., LIANG, C.-K., AND CHUANG, Y.-Y. 2011. 3d cinematography principles and their applications to stereoscopic media processing. In *Proceedings of the 19th ACM international conference on Multimedia*, ACM, New York, NY, USA, MM '11, 253–262.
- MANTIUK, R., BAZYLUK, B., AND MANTIUK, R. K. 2013. Gaze-driven object tracking for real time rendering. *Computer Graphics Forum* 32, 2pt2, 163–173.
- OSKAM, T., HORNUNG, A., BOWLES, H., MITCHELL, K., AND GROSS, M. 2011. Oscan - optimized stereoscopic camera control for interactive 3d. *ACM Trans. Graph.* 30, 6 (Dec.), 189:1–189:8.
- PARADISO, M. A., MESHI, D., PISARCIK, J., AND LEVINE, S. 2012. Eye movements reset visual perception. *Journal of Vision* 12, 12 (Dec.).
- PATTERSON, R. 2009. Review paper: Human factors of stereo displays: An update. *Journal of the Society for Information Display* 17, 12, 987.
- PFEIFFER, T. 2012. Measuring and visualizing attention in space with 3d attention volumes. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ACM, New York, NY, USA, ETRA '12, 29–36.
- PRITCH, Y., BEN-EZRA, M., AND PELEG, S. 2000. Automatic disparity control in stereo panoramas (omnistereo). In *Proceedings of the IEEE Workshop on Omnidirectional Vision*, 54–61.
- SUN, G., AND HOLLIMAN, N. 2009. Evaluating methods for controlling depth perception in stereoscopic cinematography. In *Proc. SPIE 7237, Stereoscopic Displays and Applications XX*, 72370I–72370I–12.
- SUNDSTEDT, V., BERNHARD, M., STAVRAKIS, E., REINHARD, E., AND WIMMER, M. 2013. Visual attention and gaze behavior in games: An object-based approach. In *Game Analytics*, M. Seif El-Nasr, A. Drachen, and A. Canossa, Eds. Springer London, London, 543–583.
- TALMI, K., AND LIU, J. 1999. Eye and gaze tracking for visually controlled interactive stereoscopic displays. *Signal Processing: Image Communication* 14, 10, 799–810.
- TAM, W. J., SPERANZA, F., YANO, S., SHIMONO, K., AND ONO, H. 2011. Stereoscopic 3D-TV: visual comfort. *IEEE Transactions on Broadcasting* 57, 2 (June), 335–346.
- TANRIVERDI, V., AND JACOB, R. J. K. 2000. Interacting with eye movements in virtual environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, ACM, New York, NY, USA, CHI '00, 265–272.
- TSENG, K.-L., HUANG, W.-J., LUO, A.-C., HUANG, W.-H., YEH, Y.-C., AND CHEN, W.-C. 2012. Automatically optimizing stereo camera system based on 3d cinematography principles. In *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, 2012, 1–4.
- WANG, C., AND SAWCHUK, A. A. 2008. Disparity manipulation for stereo images and video. In *Proc. SPIE 6803, Stereoscopic Displays and Applications XIX*, SPIE, A. J. Woods, N. S. Holliman, and J. O. Merritt, Eds.
- WATSON, A. B., AND PELLI, D. G. 1983. Quest: A Bayesian adaptive psychometric method. *Attention Perception & Psychophysics* 33, 113–120.
- WESTHEIMER, G., AND MITCHELL, D. 1969. The sensory stimulus for disjunctive eye movements. *Vision research*.