Measuring Gaze Depth with an Eye Tracker During Stereoscopic Display

Andrew T. Duchowski, Brandon Pelfrey, Donald H. House, Rui Wang∗

School of Computing, Clemson University

Abstract

While determining 2D gaze position using eye tracking is common practice, the efficacy of using eye tracking to measure 3D gaze point in a stereoscopic display has not been carefully studied. In this paper we explore this issue using a custom Wheatstone stereoscope augmented with an eye tracker. In a pilot study, we showed that there is strong evidence that eye vergence measurements do, in fact, track the depth component of the 3D stereo gaze point. In a subsequent full study, we compare depth estimation for a scene viewed in four different ways: without either stereo or motion parallax, without stereo but with motion parallax, with stereo but without motion parallax, and with both stereo and motion parallax. We show a significant effect related to depth from eye vergence for both of the stereo cases, but no effect for the monoscopic cases. Since depth from vergence is very noisy, we use a Butterworth filter to smooth the data and show that this greatly improves the stability of the depth reading without significant lag or loss in accuracy. We also demonstrate that using quadratic regression to perform a depth calibration can greatly improve the accuracy of the depth measurement.


Keywords: eye tracking, stereoscope, stereoscopic rendering

1 Introduction

Since their 1838 introduction by Sir Charles Wheatstone [Wheatstone 1838; Lipton 1982], stereoscopic images have appeared in a variety of forms, including dichoptic stereo pairs (different image to each eye), random-dot stereograms [Julesz 1964], autostereograms (e.g., the popular Magic Eye and Magic Eye II images), and anaglyphic images and movies, with the latter currently resurging in popularity in American cinema (e.g., Avatar, Twentieth Century Fox Film Corporation).

Eye trackers, while still expensive compared to the cost of computers, are readily available, reliable, accurate and relatively easy to use. Perhaps less dramatically, there is a developing perceptual science behind the design of visualizations for optimal data or information display. Most notably, the texts by Ware [2004; 2008] focus on perceptual issues in visualization design. While there is a natural and obvious tie between eye tracking and the science behind the perceptual optimization of visualization design, there have been as yet no such studies in the important area of stereoscopic visualization.

This paper is directed at beginning to build a solid understanding of how to track gaze for stereo displays, with the ultimate goal of gaining new insights into visualization design principles for stereo displays. It describes a study that exploits binocular eye tracking in a simple perceptual experiment designed to elicit vergence eye movement. Specifically, a remote eye tracker is used to measure depth from ocular vergence when viewing stimuli presented using a Wheatstone stereoscope. We show that the measured depth is noisy, but clearly responds to the depth component of a 3D stereo target.

The paper presents techniques for smoothing the noisy signal as well as for fitting it to targets at known depth. The combined approach points the way to a sound new method for pre-calibrating and utilizing a viewer’s gaze depth in real-time under stereoscopic conditions.

2 Background

The convergent movement of the eyes (binocular convergence), i.e., their simultaneous inward rotation toward each other (cf. divergence which denotes the outward rotation), ensures that the projection of images on the retina of both eyes are in registration with each other, allowing the brain to fuse the images into a single percept. This fused percept provides stereoscopic vision of three-dimensional space. Normal binocular vision is primarily characterized by this type of fusional vergence of the disparate retinal images [Shakhnovich 1977]. Vergence driven by retinal blur is distinguished as accommodative vergence [Büttnер-Ennever 1988]. Fusional vergence and accommodative vergence are known to be tightly coupled in the human visual system [Fincham and Walton 1957].

Dichoptic presentation of stereo displays typically decouples the link between vergence and accommodation—the eye’s focusing ability—by changing the curvature of the eye’s crystalline lens. Stereoscopic decoupling occurs when the viewer’s accommodation-vergence reflex verges the eyes at a fixation point at the screen surface while the binocular disparity component of the depth signal can move freely in front of or behind the screen surface. This decoupling is considered to be the source of eye strain and fatigue leading to discomfort with the use of such systems [Howard and Rogers 2002].

While the dissociative problems of accommodation and vergence have been studied over the last few years, measurement techniques, especially of vergence, have not improved much over this time. In particular, no eye-tracking system exists that could measure vergence in real-time with no restriction of the field of view or head movement. An infra-red limbus tracker has been used to record vergence eye movements during the use of a Helmet-Mounted Display, or HMD [Wann et al. 1995]. Since then, the vergence issue has mainly been investigated indirectly via systematic manipulation of easily adjustable parameters, namely those of head-mounted optics and screen disparity.

To gauge the effect of stereo display on ocular vergence, it is sufficient to measure the disparity between the left and right horizontal gaze coordinates, e.g., \(x\_r - x\_l\) given the left and right gaze points, \((x\_l, y\_l), (x\_r, y\_r)\) as delivered by current binocular eye trackers. Thus far, to our knowledge, the only such vergence measurements...
to have been carried out have been performed over random dot stereograms or anaglyphic stereoscopic video [Essig et al. 2004; Daugherty et al. 2010].

In this paper, vergence measurements are made with a commercially available eye tracker when viewing the Wheatstone stereoscope’s fused display. Although depth perception has been studied on desktop 3D displays [Holliman et al. 2007], eye movements were not used to verify vergence. Holliman et al. conclude that depth judgment cannot always be predicted from display geometry alone. The only other similar effort we are aware of is measurement of interocular distance on a stereo display during rendering of a stereo image at five different depths [Kwon and Shul 2006]. Interocular distance was seen to range by about 10 pixels across three participants. In this paper, we report observations on how gaze depth responds to targets at different stereo depths when shown monoscopically or stereoscopically.

3 Methodology

Our study was designed to gauge the relationship between gaze depth as determined by eye vergence response and presentation depth in a stereoscopic display. To do so we used an eye tracker to infer 3D gaze position from the gaze directions of the two eyes, while the viewer observed a set of stimuli arranged in space and displayed in stereo. All subjects were tested for the ability to correctly understand depth relationships in a random dot stereogram, and subjective impressions of the displays were obtained using a questionnaire.

3.1 Apparatus

A custom-built, high-resolution, Wheatstone-style stereoscope, whose original design was first reported by Bair et al. [2006], was used for our experiments. Figure 1 shows this stereoscope in use. It was constructed by fixing two IBM T221 “Big Bertha” LCD displays on a track, on opposite sides of two small mirrors angled at 45° from the medial axis. The screens are 48 cm wide and 30 cm high (16:10), with a screen resolution of 3840 × 2400, or 9.2 million pixels. The screens were set at a viewing distance of 86 cm from the nominal eye point. This yielded a visual angle per pixel of approximately 30 seconds of arc. This is comparable to the size of the receptors in the fovea and is sufficient to display the finest grating pattern that can be resolved by the human eye—about 60 cycles per degree, corresponding to 1 cycle per minute of visual angle [Campbell and Green 1965]. According to the manufacturer, the screen contrast ratio is 400:1, and maximum brightness is 235 cd/m². Images were rendered assuming an eye separation of 6.3 cm. The monitors and mirrors were carefully aligned so that when looking into the mirrors, both monitors are fused into a single virtual image, as illustrated in Figure 2. By displaying a view rendered from the left eye position on the left monitor, and the right eye position on the right monitor, a very high quality, high luminance stereoscopic image is formed.

Eye tracking cameras from LC Technologies were mounted beneath each monitor. The cameras were part of LC Technologies’ Eyegaze System that is used to image the viewer’s eyes as seen by the cameras in the mirrors. Each of the cameras houses a small, low power, infrared light emitting diode (LED) located at the center of the camera lens illuminating the eye [Hutchinson 1989]. The Eyegaze system uses the Pupil-Corneal Reflection (P-CR) method to measure the direction of the eye’s gaze, sampling at 60 Hz with Root Mean Squared (RMS) tracking error of 0.635 cm or less [LC Technologies, Inc. 2010]. A proprietary algorithm is used to provide gaze depth (z), along with the xy coordinates of the gaze points of the two eyes.

The Eyegaze System server ran on a Mac Mini running the Windows XP operating system. The client display application ran on an Intel Xeon E5502 Quadcore PC equipped with two NVIDIA Quadro FX 5800 graphics cards and running the CentOS Linux operating system. One graphics card drives the left half of each of the Big Bertha monitors using clone mode stereo, and the other card drives the right half. The client/server PCs were connected via 1 Gb Ethernet (connected via a switch on the same subnet).

3.2 Stimulus

To elicit the perception of depth, a 5 × 5 grid of cubes was rendered with the closest row of cubes shown at 30 cm in front of the screen with each of the four remaining rows 12 cm farther from the viewer than the last (see Figure 3). Each cube measured 2 × 2 × 2 cm. A right-handed coordinate system was used so that the screen plane was aligned at z = 0; three cube rows were made to project in front of the screen at depths z = 30, z = 18, and z = 6, and two rows receded into the screen at depths z = −6 and z = −18.

In order to provide depth cues from motion, the cube grid was designed so that it can rock around its vertical center by a total angle of 80° (40° to the left and right) within a 5-second cycle. Note that this is equivalent to rotating the camera around the vertical center of the scene, and induces motion parallax. In order to draw the viewer’s attention to the cube that they were to attend to, individual cubes could also rock. Individual cubes rock 7° (3.5° to the left and right) over each 5-second viewing cycle.
3.3 Experimental Design

The experiment consisted of a within-subjects 2 (display) × 2 (motion parallax) × 5 (depth) design. The display was either rendered monoscopically or stereoscopically, motion parallax was induced by either rocking the cube grid or holding it steady (fronto-parallel to the image plane), and depth was indicated to the viewer by rocking the individual cube that the viewer was asked to fixate.

The primary task assigned to participants was to visually fixate whichever single cube was rocking. Individual cube rocking order always proceeded along either the left, center, or right column of the cube grid. The column selected went from left to right or right to left, and the starting row either began at the front (protruding out of the screen) or at the back (receding toward the screen interior).

Familiarity and fatigue effects were mitigated by counterbalancing both the 2 × 2 stereo and motion combinations, and by alternating depth order. Treating stereo and motion as four conditions, \{A, B, C, D\}, referring to the tensor product combinations of \{monoscopic, stereoscopic\} ⊗ \{static, rocking\}, respectively, viewing trials were presented in rotated Latin square order. Depth was varied so that all even-number participants were asked to fixate cubes at a depth starting in the back and progressing to the front, and odd-number participants from front to back.

3.4 Participants

Twenty college students (18 M, 2 F; ages 16-34) participated in the study, recruited verbally on a volunteer basis. Two of the participants wore contact lenses, the rest wore no type of corrective lenses. All had a minimum of six years experience with computers and all spent at least two hours per day at a computer. Only three participants had previously seen a stereoscopic film.

Participants were not pre-screened for color blindness, but were pre-screened for depth perception by showing them the random-dot stereogram display of two tiles as depicted with left and right eye views superimposed in Figure 4. For persons with normal stereo vision, the left tile’s disparity placed it behind the screen, and the right tile’s disparity placed it in front of the screen. All but one participant was able to see this depth difference correctly, with this participant seeing the two tiles at the same depth.

3.5 Procedure

Following introductions and reading of the Institutional Review Board’s approved informational letter, demographic information was collected consisting of the participant’s age, gender, usage of corrected lenses and computer experience. Each participant filled out a short pre-test questionnaire regarding their familiarity with stereoscopic displays. The stereogram pair was then presented and the participant was asked whether either of the stereograms protruded out of the screen, receded into the screen, or was aligned flat with the screen plane (the right stereogram protruded out of the screen in all cases).

A quick (two-dimensional) calibration of the eye tracker was then performed by having participants visually follow a roving dot between nine different locations on the screen. After presentation of the initial nine targets, the calibration procedure can be accepted and terminated by the experimenter, or it can be allowed to proceed until the average calibration error is reduced to an acceptable level. A pilot study conducted prior to the experiment indicated that gaze depth estimation was heavily dependent on calibration accuracy. Thus, we chose to continue calibration until the average accuracy was reduced below 1.3 cm accuracy. An example of the server screen displaying calibration results is shown in Figure 5. The calibration points are shown as round dots, and the corresponding average gaze points are shown as crosses. The two image inserts show the two eye tracker camera views, and the selected eye regions of these views.

After 2D calibration, the participant was shown the first of four variations of the cube grid. Following visual fixation of 15 cubes
shown on the left in Figure 5. Calibration results rendered by the eye tracker. (5 in each of 3 columns), the participant was presented with a question shown on the screen asking for the viewer’s impression of the stereoscopic effect. The question simply posed the statement, “The display gave me an impression of depth,” and the participant was asked to respond verbally with a numerical value given on a 7-point Likert scale (also shown on the screen) with 1 indicating strong disagreement, 4 indicating neutrality, and 7 indicating strong agreement.

3.6 Pilot Study

Following hardware setup and software development, a short pilot study was run where data from three subjects (not part of the main study) were analyzed to look for indications of a strong effect of target depth on eye vergence response as returned by the eye tracker. The pilot data were also used to develop our data processing and analysis scripts for the full study. Representative depictions of data captured from two trials (monocular and stereoscopic) are shown in Figure 6. The staircase curves represent the depth of the target cube, rocking at each depth location for a duration of 5 s. The smooth curves represent the depth as returned by the eye tracker. In all of these curves, outliers were removed by computing the standard deviation of all data points for a single depth target, and removing those points beyond 2 standard deviations.

Several important observations were made regarding the data captured during the pilot study. While the monocular data did not appear to track target depth, the stereo data indicated that there might be a strong positively correlated effect (compare the left and right columns of Figure 6). Also, it was apparent that the gaze depth (z-coordinate) delivered by the eye tracker is quite noisy, as can be seen in Figure 6(a). Filtering the raw data with a 6th order Butterworth filter allowed us to produce the smoother curves shown in Figure 6(b). Although a response to stereoscopic projection was observed, in most cases the data appeared to be significantly offset in depth, relative to the screen distance. This was particularly true under monoscopic conditions, when the depth reading was expected to average over time at the screen depth, \( z = 0 \). Shifting the data to correct the mean response led to results such as those shown on the left in Figure 6(c). An optimal root mean square (RMS) fit of the data using a quadratic regression is shown on the right of the same figure. Finally, the observed depth signal pattern under stereoscopic conditions appeared much less reliable when a sufficiently accurate eye tracker calibration could not be achieved (anecdotally we observed that data was generally unreliable when average screen-space accuracy was greater than 1.3 cm).

The pilot study thus served several purposes. First, and foremost, we could see in the data a very definite response from the eye tracker indicating that in the stereo conditions the extent of eye vergence corresponded to the depth of the target being viewed. This made it clear to us that a full study was justified. We also learned important lessons concerning the need for careful 2D calibration before attempting to measure depth, the need for a filtering mechanism to smooth the very noisy depth signal, and the need to calibrate or fit the data in 3D to improve the accuracy of depth readings.

3.6.1 Filtering

The noise in the depth data from the eye tracker made it obvious that we would have to use some filtering method to smooth the data before using it. There are many choices for a smoothing filter, including the popular approach of averaging across a sliding window of samples [Augustin 2009]. Since our end goal is to be able to use gaze depth tracking in an real-time environment, it was important to us that whatever filter we chose be causal, and also not introduce a severe lag. We decided on a lowpass Butterworth filter. They are causal, and are known to have a very flat frequency response below the cutoff frequency, no ringing in the stop band, and a steep transition from the pass band to the stop band [Oppenheim and Schafer 1989].

The formulation of a discrete time 2nd order Butterworth filter is given by the difference equation

\[
\hat{x}(n) = a_0x(n) + a_1x(n-1) + a_2x(n-2) - b_1\hat{x}(n-1) - b_2\hat{x}(n-2),
\]

where \(x(n)\) is the nth raw data sample, and \(\hat{x}(n)\) is the corresponding nth filtered sample. The coefficients of Equation 1 are determined mainly by the order of the filter \(N\), and the ratio \(r\) of the filtering frequency \(f_s\) to the desired cutoff frequency \(f_c\). They are computed as follows:

\[
r = f_s/f_c, \quad \omega = \tan(\pi/r),
\]

\[
c = 1 + 2\cos(\frac{\pi}{2N}\omega + \omega^2),
\]

\[
a_0 = a_2 = \omega^2/c, \quad a_1 = 2a_0,
\]

\[
b_1 = 2(\omega^2 - 1)/c, \quad b_2 = (1 - 2\cos(\frac{\pi}{2N}\omega + \omega^2))/c.
\]

By cascading 2nd order Butterworth filters, a higher order filter can be obtained. We found that, for our data, a 6th order filter, formed by cascading three 2nd order filters, gave excellent results. Our tracker’s sampling frequency was \(f_s = 60\) Hz, and we experimentally chose a cutoff frequency \(f_c = 0.15\) Hz. Thus, our filter parameters were

\[
r = 400.0, \omega = 0.00785, c = 1.01117,
\]

\[
a_0 = a_2 = 0.00006, a_1 = 0.00012,
\]

\[
b_1 = -1.97779, b_2 = 0.97804.
\]

The impulse response of this filter has a lag from impulse to peak response of 1.15 seconds, with a gain of 1.0 (i.e., the integral of the response to a unit impulse is 1.0).
Figure 6: Visualizations of representative gaze depth in response to stimulus (rocking cube) at varying z-coordinates: (a) raw data with outliers beyond 2 SD removed; (b) data filtered with a 6th order Butterworth filter; and (c) filtered data either shifted (monocular condition) or fit via least squares minimization (stereo condition).
3.6.2 Calibration

A popular approach to 2D eye tracking calibration depends on solution of a quadratic polynomial by defining \((s_x, s_y)\) as the screen coordinates of calibration points, and \((x, y)\) as the vector from some static reference point (e.g., corneal reflection of an IR source) to the pupil center (which varies as the eye moves) [Morimoto and Mimica 2005]:

\[
\begin{align*}
    s_x &= a_0 + a_1x + a_2y + a_3xy + a_4x^2 + a_5y^2, \\
    s_y &= b_0 + b_1x + b_2y + b_3xy + b_4x^2 + b_5y^2.
\end{align*}
\]

The parameters \(a_0\)–\(a_5\) and \(b_0\)–\(b_5\) are the unknowns. Reformulating in matrix notation leads to:

\[
\begin{bmatrix}
    s_{1x} & s_{1y} \\
    s_{2x} & s_{2y} \\
    \vdots & \vdots \\
    s_{nx} & s_{ny}
\end{bmatrix} =
\begin{bmatrix}
    1 & x_1 & y_1 & x_1y_1 & x_1^2 & y_1^2 \\
    1 & x_2 & y_2 & x_2y_2 & x_2^2 & y_2^2 \\
    \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
    1 & x_n & y_n & x_ny_n & x_n^2 & y_n^2
\end{bmatrix}
\begin{bmatrix}
    a_0 \\
    a_1 \\
    a_2 \\
    a_3 \\
    a_4 \\
    a_5
\end{bmatrix} +
\begin{bmatrix}
    b_0 \\
    b_1 \\
    b_2 \\
    b_3 \\
    b_4 \\
    b_5
\end{bmatrix}
\]

which reduces to \(S = XB\) for \(n\) calibration points used. Note that typically a large number of gaze points are sampled at each calibration point. To effect correspondence between matrix dimensions, sampled gaze points can be aggregated in the average producing a centroid so that the error between each sampled gaze point centroid and corresponding calibration point is minimized. Alternatively, if the number of gaze points exceeds the number of calibration points (the usual case), each calibration point is repeated in matrix \(S\) for each sampled gaze point in the corresponding calibration point set. We used the latter approach to fitting the observed gaze depth to target cube depth (i.e., treating the cubes as calibration targets).

The general solution for estimation of \(B\) relies on Lagrange’s method of least squares, or the multivariate multiple regression model [Lancaster and Šalkauskas 1986; Finn 1974]. The general linear model describing \(i \in [1, n]\) observations of \(p\) random variables \(s_k, k \in [1, p]\), from \(q\) predictors \(x_j, j \in [1, q]\) is specified by the following \(p\) separate univariate equations:

\[
\begin{bmatrix}
    s_{i1} \cdots s_{ip}
\end{bmatrix} =
\begin{bmatrix}
    1 & x_{i1} & \cdots & x_{iq}
\end{bmatrix}
\begin{bmatrix}
    \hat{\alpha}_1 & \hat{\alpha}_2 & \cdots & \hat{\alpha}_p \\
    \hat{\beta}_11 & \hat{\beta}_12 & \cdots & \hat{\beta}_1p \\
    \vdots & \vdots & \cdots & \vdots \\
    \hat{\beta}_q1 & \hat{\beta}_q2 & \cdots & \hat{\beta}_qp
\end{bmatrix} +
\begin{bmatrix}
    \hat{\epsilon}_{i1} \cdots \hat{\epsilon}_{ip}
\end{bmatrix}
\]

which reduces to \(S = XB + \hat{E}\), where \(\hat{E}\) is the \(n \times p\) matrix of sample residuals or errors, and \(\hat{B}\) is the \((q + 1) \times p\) matrix of partial regression coefficients for predicting each outcome measure from the \(p\) independent variables. To estimate \(B\), the squared sample residuals of \(n\) sampled observations are minimized. The sum of squared residuals for one outcome measure is one diagonal element of \(\hat{E}^T\hat{E}\), and their sum is the trace of \(\hat{E}^T\hat{E}\), which is minimized by setting the partial derivatives with respect to the elements of \(\hat{B}\) to zero and solving. The resulting normal equations are \(X^TX\hat{B} = X^TS\). The system is left-multiplied by \((X^TX)^{-1}\) to obtain the estimate of \(\hat{B} = (X^TX)^{-1}X^TS = G^{-1}X^TS\), where \(G^{-1}\) is known as the pseudo-inverse of \(X\).

For gaze depth point fitting, we formulated \(S = XB\) by using a second-order polynomial representation of the observed gaze depth to yield the calibrated depth coordinate \(s_z\):

\[
\begin{bmatrix}
    s_{iz}
\end{bmatrix} =
\begin{bmatrix}
    1 & z_i & z_i^2
\end{bmatrix}
\begin{bmatrix}
    a_0 \\
    a_1 \\
    a_2
\end{bmatrix}
\]

and solved for \(\hat{B}\) as indicated.

4 Results

Results from our full study with 20 subjects, conducted on the Wheatstone stereoscope augmented with an eye tracker, are summarized in Figure 7.

We note first that the observed average 2D screen space accuracy range for the 20 participants’ data used in the analysis of the study was [0.6, 1.1] cm, with maximum worst calibration point observed at 2.8 cm. Data from two participants were not used due to lost data over extended periods of time (e.g., the eye tracker did not provide data for a period of over 10 s—because each trial duration was 80 s, this was deemed too severe a loss of data to contribute to the analysis). Before analysis, all data were filtered using the Butterworth filter and then processed to eliminate outliers.

Analysis was carried out to examine the effects of data shifting and fitting. Shifting was accomplished by first calculating the signal mean over the duration of each trial, and then subtracting that amount from every record, effectively shifting the entire trial record to \(z = 0\), as seen in the left column of Figure 6(c). Fitting was accomplished by applying the quadratic regression method, described above, to both shift and scale the data to each depth interval. This effectively simulated depth calibration in a post-hoc fashion. Results are presented in terms of RMS error between processed gaze depth and the depth of the intended (cube) targets.

Figure 7: Root mean square error of gaze depth under the four viewing conditions.
For shifted data, a repeated-measures three-way ANOVA of gaze depth error revealed significance of the main effect of depth (F(4,76) = 59.50, p<0.01), no significance of display (mono vs. stereo) (F(1,19) = 3.29, p<0.09), and no significance of motion parallax (static vs. rocking) (F(1,19) = 1.21, p=0.29, n.s.). No interaction effects were observed (see Figure 7(a)).

For fit data, a repeated-measures three-way ANOVA of gaze depth error revealed significance of the main effects of depth (F(4,76) = 323.29, p<0.01) and display (F(1,19) = 126.00, p<0.01), but not of motion parallax (F(1,19) = 3.12, p = 0.09, n.s.). Interaction between depth and display was significant (F(4,76) = 62.52, p<0.01). No other significant interactions were detected (see Figure 7(b)).

Because gaze depth under monoscopic viewing conditions is not expected to vary away from $z = 0$, the screen depth, it may not make sense to fit the data to targets that vary in depth. However, it does make sense to shift the data en masse to $z = 0$. Figure 7(c) shows the comparison between shifted data captured under monoscopic conditions and fit data captured under stereoscopic conditions. For this data, a repeated-measures three-way ANOVA of gaze depth error revealed significance of the main effects of depth (F(4,76) = 127.14, p<0.01) and display (F(1,19) = 125.55, p<0.01), and marginal significance of motion parallax (F(1,19) = 5.63, p<0.05). Interaction between depth and display was significant (F(4,76) = 83.77, p<0.01). No other significant interactions were detected.

Interaction between depth and display is depicted in Figure 8. As target depth either protrudes in front of the screen plane or recedes behind, gaze depth error increases.

Pairwise t-tests of participants’ subjective evaluations of the stereoscopic effect, depicted in Figure 9, show a significant difference between the static monocular and rocking stereo conditions (p<0.01, with Bonferroni correction), with a marginally significant difference between the static monocular and static stereo conditions (p<0.05, with Bonferroni correction). No other significant differences were observed. Modal responses to the Likert questionnaire are in general agreement to computed means, with both types of stereo garnering strong agreement (7 on the Likert scale), rocking monoscopic receiving agreement (6 on the Likert scale), and static monoscopic acquiring some agreement (5 on the Likert scale).

5 Discussion and Future Work

Eye vergence movements clearly respond and match the depth component of a 3D stereo target point. However, measured vergence response is unexpectedly noisy. The source of the noise may be specific to the eye tracking equipment used in the study, further investigation is needed with other equipment to verify this. Our error study showed that depth error is minimized near the screen depth. In future studies we plan to look at this issue more carefully using a more fine grained set of stimuli around the screen depth.

The Wheatstone stereoscope requires that binocular eye tracking optics be split apart so that the tracker functions as a combination of two monocular eye trackers. However, most modern eye trackers are binocular and thus their applicability to the Wheatstone apparatus is at present unknown. We are anxious to verify how a more standard tracker using an active stereo display would function.

Hardware configuration notwithstanding, we believe that the Butterworth filter is likely to be effective for gaze depth filtering on other platforms. We also feel that the data fitting model that we developed for our experiments will be very useful in forming the basis for a pre-calibration procedure to improve depth measurement accuracy.

Although we expected vergence to respond to stereoscopic display, we were somewhat surprised by participants’ somewhat positive subjective evaluation of the monoscopic displays, particularly the display without motion parallax. The counterbalanced order of presentation of the stimuli suggests that subjective responses may be skewed depending on which display is seen first: the stereoscopic effect is quite pronounced when experienced after viewing the monoscopic display. Conversely, something of an after-effect may be present when the stereoscopic display is experienced first: the monoscopic display is strangely still considered somewhat stereoscopic. This result may not be too surprising since it is known that linear perspective provides a strong depth cue, and this is highly apparent in the arrangement of the cubes in our experimental setup (see Figure 3).

6 Conclusion

We have reported empirical measurements of gaze depth, computed from eye vergence, in relation to a target at varying depth distances shown either monoscopically or stereoscopically. As expected, gaze depth clearly responds and tends towards the target depth under stereoscopic conditions. We have also presented combined filtering and fitting techniques to simultaneously counteract the observed high level of noise in the gaze depth signal, as well as the viewer’s depth under- or over-estimation of the target. The combined use of the Butterworth filter along with least squares fitting suggests an effective means of gaze depth pre-calibration.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 0915085. Any opinions, findings, and conclusions expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF. We are grateful to NVIDIA for their generous equipment donations.

References

Figure 8: Root mean square error of gaze depth over the five viewing depth intervals.


