# OCULAR VERGENCE RESPONSE OVER ANAGLYPHIC STEREOSCOPIC VIDEOS

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Computer Science

by
Brian C. Daugherty
May 2009

Accepted by:
Dr. Andrew T. Duchowski, Committee Chair
Dr. Donald H. House
Dr. Roy P. Pargas

## 0.1 Abstract

The effect of anaglyphic stereographic stimuli on ocular vergence response is examined. An experiment is performed comparing ocular vergence response induced by anaglyphic stereographic display versus standard monoscopic display. Two visualization tools, synchronized three-dimensional scanpath playback and real-time dynamic heatmap generation, are developed and used to subjectively support the quantitative analysis of ocular disparity. The results of a one-way ANOVA indicate that there is a highly significant effect of anaglyphic stereoscopic display on ocular vergence for a majority of subjects although consistency of vergence response is difficult to predict.

## 0.2 Acknowledgements

Special thanks go to my advisor, Dr. Andrew T. Duchowski for supplying the equipment, technical expertise, and statistical guidance. Other special thanks go to Celambarasan Ramasamy for designing, modeling, animating, and producing the anaglyphic videos.

# Table of Contents

# List of Figures

# Chapter 1

# INTRODUCTION

As Hollywood constantly looks for new and interesting ways to present movies, recent developments and trends have turned directors' attention toward stereoscopic films. Stereoscopic images, or images that give an illusion of depth as if they were three-dimensional, were introduced in 1838 by Sir Charles Wheatstone [Lipton 1982]. Though there are many different techniques to create and view stereoscopic images, one method, construction of anaglyphic images, provides a format that can easily be used in movie theaters.

Anaglyphic stereoscopic images are created using two different cameras spaced slightly apart in order to simulate binocular human vision. The two images are composited together into a single image composed of the image seen by the left eye in one color and the image seen by the right eye in a contrasting color. For the purpose of this thesis, the image seen by the left eye is shown in red and the image seen by the right eye is shown in cyan. By wearing a pair of cheap, easy to produce glasses with a red filter for the left lens and a cyan filter for the right lens, most people with normal binocular vision can see the three-dimensional illusion.

As with any form of media presentation, the effectiveness of the medium needs to be evaluated to see if the information is effectively conveyed to the viewer. In this case, the effectiveness of anaglyphic stereoscopic images to convey the illusion of depth to the viewing audience needs to be evaluated in some form of quantitative analysis. If there is a lack of evidence that anaglyphic stereoscopic images convey an illusion of depth, then there is no reason to use this presentation technique over the standard, monocular camera approach.

When a human with normal binocular vision is presented with stereoscopic imagery, the human eyes naturally and simultaneously rotate inwards, or converge, in order to fuse the two images into a single three-dimensional representation of the objects in the scene. By using an eye tracker capable of tracking both eyes simultaneously, we will analyze the effects of the vergence response when presented with anaglyphic stereoscopic stimuli in order to determine if there is a significant difference between the depth perception of an anaglyphic stereoscopic scene versus a monocular presentation.

# Chapter 2

# BACKGROUND

Since human eyes are approximately two and a half inches apart, each eye views an image from a slightly different angle. The two eyes simultaneously rotate inward toward each other, or converge, in order to allow the brain to form a single three-dimensional representation of the objects in the image. This fusional convergence of the two images in the brain is the primary characteristic of normal binocular vision [Shakhnovich 1977]. On the other hand, accommodative vergence is the convergence of the eyes due to retinal blur [Büttner-Ennever 1988].

An anaglyphic stereogram is a composite picture consisting of two colors and two slightly different perspectives that produces a three-dimensional image when viewed through two corresponding colored filters. For our purposes, an anaglyphic stereogram is created using a red image for the left eye, and a cyan image for the right eye. Likewise, viewers of these images wear glasses with a red lens on the left side, and a cyan lens on the right side. Since the red lenses filter out the cyan light, only the red image is seen in the left eye. Likewise, only the cyan image is seen in the right eye. The distance between the red and cyan images creates the illusion of depth when they are fused together in the brain.

Many other forms of stereograms exist, including the random-dot stereogram. Random-dot stereograms are created by taking one image and shifting parts of it to the left or right. Any gaps in the image are filled in using randomly colored dots. A viewer can see the three-dimensional effect by diverging his or her eyes until the shifted parts of the image merge together.

Another form of stereogram is created with the help of shutter glasses. In this form, a special display screen rapidly alternates between two different images. Using shutters for lenses, the glasses rapidly alternate which lens is currently open at the same rate that the display screen. The effect is that each eye can see only one of the two images that the brain uses to fuse together into a three-dimensional representation.

The main focus of convergence and stereogram research focuses on display techniques or ways to measure convergence. At the Schepens Eye Research Institute at Harvard Medical School, a stereo display method was developed that uses CrystalEyes LCD shutter glasses and a dynamic convergence point in order to display computer generated stereograms. An eye-tracker was used to change the focal point of the scene based on what object the user was focusing on. The rest of the scene was adjusted so that objects closer than the focal point appeared to pop out of the screen, while objects further away sank into the screen [Peli et al. 2001].

Another stereo display was developed by Akeley et al. that provides multiple focal distances for stereograms. Similar to autostereoscopic volumetric displays, this display generates near-correct focus cues due to light sources at various focal distances. However, this display correctly handles view-dependent lighting effects such as occlusion, specularity, and reflection, unlike autostereoscopic volumetric displays [Akeley et al. 2004].

The Neuroinformatics Group at Bielefeld University studied the effect that granularity of a random-dot stereogram had on the speed that a viewer's eyes would diverge in order to see the three-dimensional image. They found that divergence was achieved faster for coarse than for fine granularities [Essig et al. 2004].

# Chapter 3

# TECHNICAL DEVELOPMENT

## 3.1 Eye Tracker Calibration

As the distance between the eyes is different for every human, the eye tracker must be calibrated for each participant. A standard calibration approach was used consisting of nine focus points. A focus point is simply a coordinate on the screen where the participant would fixate while the eye tracker recorded gaze samples. Each of the nine focus points was visited in turn by a small white circle that grew larger at each focus point. A small dot was placed in the center of the circle to provide the participant with a small target to fixate on. Calculations used for the calibration after all focus points had been visited were performed by software provided with the eye tracker.

As very few people have had prior experience with eye trackers, a feedback system was implemented to show the participants when and where the eye tracker could locate his or her eyes. Two large white circles were drawn on the screen indicating where the eye tracker thought the participants eyes were. This allowed the participant to move his or her head so that the two white circles were roughly in the center of the screen.

When calibration is completed, the software reports only a simple yes or no as to whether or not it is a good calibration. However, a successful calibration's quality can vary significantly. In order to test the quality of a particular calibration, another feedback system was implemented. This system simply displayed two dots on the screen where the eye tracker reported that the participant was looking. Since the eye tracker reports gaze data as a coordinate for each eye, these dots represented where the participant's left and right eyes were looking respectively.

## 3.2 3D Calibration

One goal is to not only collect gaze data over anaglyphic stereoscopic movies, but to be able to play it back as well. During this playback, it is interesting to visually depict how much a participant is converging or diverging. One problem that arises is the difference in eye convergence between two people staring at a blank screen. There are many factors that can account for this difference such as the distance between a participant's eyes. To provide a normalized metric, there is a need for a three-dimensional calibration.

For the three-dimensional calibration, a video roughly 40 seconds long was created. This video depicts a roving white circle in three-dimensional space within a cube. A checkerboard texture is applied to every face on the inside of the cube to provide depth queues for the viewer. The circle starts off against the back wall, and appears to be sinking into the screen. After stopping at all four corners and the center, the circle moves closer to the viewer. The circle continues to visit the four corners and the center, but this time it neither appears to sink into the screen nor pop out. Finally, the circle repeats the five positions, but popping out of the screen.

Gaze data is collected over this video and then analyzed. First, the gaze data is sorted into three categories, one for each visual plane. The gaze data is run through a fixation filter, and only gaze coordinates that occurred during fixations are kept. The disparity $D$ between the left and right eye is calculated as:

$$D = X_R - X_L \tag{3.1}$$

The gaze coordinates are sorted in ascending order, based on the disparity, within each group. Next, outliers are identified using the interquartile range method and removed from the groups in order to minimize the effects that eye tracker errors have on the three-dimensional calibration.

Due to the way that the disparity is calculated, a larger value indicates divergence and a smaller (negative) value indicates convergence. When the circle is in the back plane, and therefore sinking into the screen, the participant's eyes diverge in order to fuse the two images together. For this reason, we save the maximum disparity value within the back visual plane group as part of the calibration. Similarly, we save the minimum disparity value within the front visual plane group. The third value saved for the calibration is the linear average of the disparities in the middle visual plane group. A description of how these three values are used during the playback of a scan path is given below.

## 3.3 Playback

One issue that arises is the need to playback a scan path in order to evaluate where a participant looked. Since the video library was written to provide media player style functionality opposed to video processing, the library skips frames while playing a video in order to maintain the desired playing speed. The videos were all created to run at 25 frames per second while the eye tracker recorded data at 50 Hz. If no frames were lost, the solution would be easy. Two pieces of gaze data would be extracted for every frame. However, this is not the case, and there is need for synchronization.

In order to achieve synchronization, a timer object was set to timeout every twenty milliseconds to simulate the speed of the 50 Hz eye tracker. All of the scan path data is read in and the coordinates are placed in a queue. Every time the timer times out, the current system time is acquired. The difference between the current time and the time when the video started is calculated and used to compare against the timestamps of the gaze data in the queue. As long as the timestamp of the piece of gaze data at the front of the queue is smaller than the current elapsed time, the gaze data is popped off of the queue.

A second problem that is encountered when dealing with scanpaths is the natural eye jitter which leads to noisy data. One solution to this problem is to use a moving window to average gaze data. The coordinate used for playback is the linear average of the moving window. The larger the window is, the smoother the average coordinate will be over time. A moving window of size ten appears to be sufficiently large enough to produce the desired smoothing effect.

When a moving window is introduced, the problem of lag is introduced as well. If the linear average of the last ten gaze data points is used, and the difference in timestamps between any successive pair of gaze data is 20 milliseconds, then the average gaze coordinate from the window will be 100 milliseconds behind the most recent timestamp. To eliminate the lag of 100 milliseconds, the window is shifted forward by half its length. That is to say, the moving window always holds ten pieces of gaze data; five of which are from the past, and the other five are from the future.

Sometimes, the eye tracker loses track of where the eyes are looking, and the data is flagged as invalid. This situation can be caused by simple things such as when the participant blinks, or when the participant looks too far off-screen, or even moves his or her head out of the volume in which the eye tracker operates. The invalid data will skew the average gaze coordinate. In our case, invalid data is not only flagged, but the coordinate $(-1.0, -1.0)$ is returned, which pulls the average towards the top left corner of the screen.

One approach to solving this problem is to never load invalid data into the queue when the scan path is loaded. The problem with this approach is that if there is a long series of invalid data points in a row, then the moving window will only contain the last ten valid gaze coordinates. This would make the average coordinate of the moving window remain static until more valid data is found. If one were to display this coordinate on the screen, it would give the illusion that the participant looked at a particular location, or even dwelled on it, when in reality the participants eyes could not be found.

A better approach is to allow the invalid gaze data to flow into the moving window. This way, when the invalid data enters the queue, the oldest piece of gaze data leaves the queue, and the average coordinate in the window changes. So that the invalid data does not skew the average coordinate, all invalid data is omitted when calculating the linear average of the window. That is, the average is calculated only with the valid data.

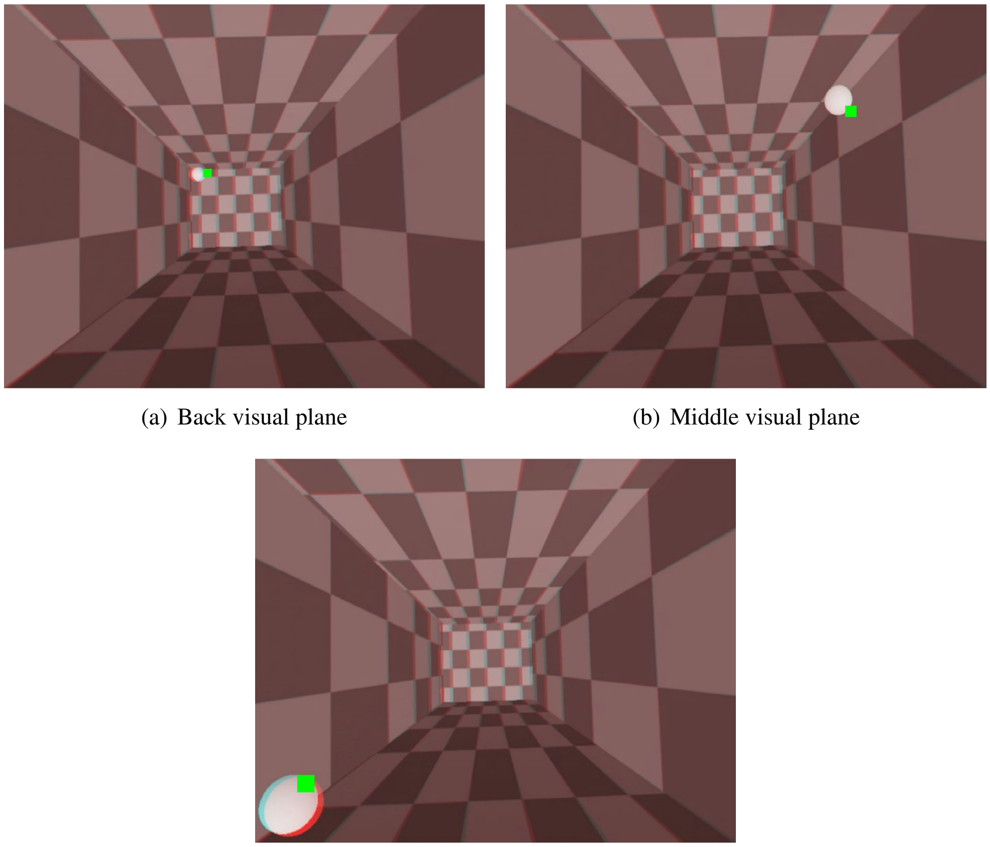
When the rolling window is comprised almost entirely of invalid data, only a few pieces of valid data are used to compute the average coordinate. The problem with this is that the average coordinate itself becomes noisy, especially if it is the average of only one piece of valid data. In the very worst case, the entire window could be invalid data, yielding no average coordinate. In order to solve this problem, a threshold of 0.8 was used. If more than 80 percent of the moving window was comprised of invalid data, then the average coordinate was flagged as invalid.

Now that the average coordinate of the rolling window is obtained, we need a way to visually display it. The simplest form is to just draw a dot on top of the video to visualize the scan path. The problem is how to display whether or not the participant's eyes are converging or diverging at that gaze coordinate. One way is to simply increase the size of the dot whenever the participant is converging, and decrease the size of the dot whenever the participant's eyes are diverging. This gives the effect that the dot is getting larger, or coming closer when the participant's eyes converge meaning that he or she is looking at an object popping out of the screen. Since the disparity is attached to every piece of gaze data, the linear average of the valid coordinates in the rolling window will also yield an average amount of disparity. We can use this disparity, along with the results from the three-dimensional calibration to calculate the radius $R$ of the dot as:

$$R = \begin{cases} R_0 + k * (1 - \frac{D - Min_F}{Avg_M - Min_F}) & D < Avg_M \\ R_0 - k * \frac{D - Avg_M}{Max_B - Avg_M} & D \geq Avg_M \end{cases} \tag{3.2}$$

where $D$ is the average disparity of the valid data in the moving window, $Min_F$ is the minimum disparity in the front visual plane, $Avg_M$ is the average disparity over the middle visual plane, $Max_B$ is the maximum disparity over the back visual plane, $R_0$ is some constant for the size of the dot at the average disparity, and $k$ is some arbitrary scalar. The effect is that the size of the dot is the result of linearly interpolating the disparity between $Min_F$, $Avg_M$, and $Max_B$.

The result of the calibration and calculation of the dot is illustrated in Figure 3.1:



(a) Back visual plane

(b) Middle visual plane

(c) Front visual plane

**Figure 3.1:** During scanpath playback, the size of the playback dot is adjusted to visually depict the disparity of the participant's eyes. Smaller dots, such as the one in Figure 3.1(a) represent divergence, while larger dots, such as the one in Figure 3.1(c) represent convergence.

## 3.4 Dynamic Heatmap

While the scanpath playback for a single participant is as interesting as it is informative, there is a need to view multiple scanpaths simultaneously. A standard approach to visually depicting aggregate gaze data compiled from multiple participants is to create a heatmap. While a static heatmap has many useful applications for static stimuli, it is not sufficient for a dynamic video. One of the biggest problems for using a static heatmap over a video is that the scene changes whenever the camera moves. Objects that were once in the center of the screen can now be on the edges of the screen or even completely off of the screen. Scene changes only further complicate the issue by potentially introducing an entirely different environment where no objects from the previous scene are present in the new scene.

In this regard, there is a need for a dynamic heatmap. While the dynamic heatmap can be generated off-line ahead of time as a series of frames compiled into a video, a dynamic heatmap was developed that is generated in real-time. One limitation to this is the volume of scanpaths that can be displayed at the same time. However, with only twelve participants and twelve scanpaths, this real-time approach was feasible.

The heatmap consists of a simple accumulation buffer that is initialized to zero. Every time a video frame is rendered on the screen, the heatmap is updated. A Gaussian bubble is added to the heatmap for the linear average coordinate from the moving window of each of the twelve scanpaths. The intensity and radius of these bubbles are just some arbitrary constants chosen to provide a good heatmap. A radius of 20 pixels with an intensity of 60 was used.

As the heatmap is always changing, there is a need to reduce the intensity of the heatmap where participants are no longer looking. One approach is to zero out the accumulation buffer before every frame. While this is effective at removing residual heat, it produces an effect that is not very smooth. Also, as soon as a participant looks at a given coordinate, the heat bubble at the spot would instantly be displayed at full intensity. This is true whether the participant was focusing on that object, or if it was just part of a saccade. In this regard, some form of motion blur is needed so that the heat builds up as a participant focuses on something interesting and cools down when he or she looks away. Also, by making the heat build up to a maximum amount over several frames rather than one, a saccade looks like a streak of faint heat rather than a series of bright spots.

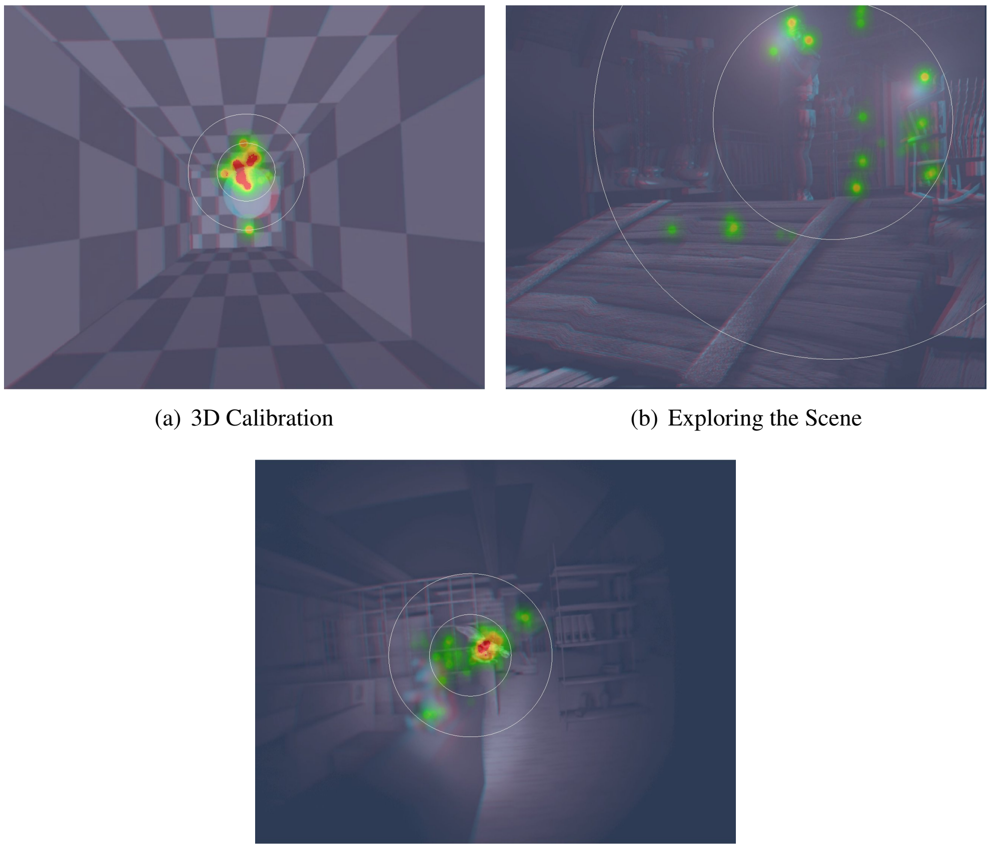
To achieve motion blur, the heat $H$ for each pixel becomes a function of the previous frame and the current frame calculated as in (3.3) where the fade factor $F$ is some arbitrary constant $F \in [0, 1]$, $H_P$ is the heat from the previous frame, and $H_N$ is the heat being added by the new frame.

$$H = (1 - F) * H_P + F * H_N \tag{3.3}$$

In practice, a value $F = 0.4$ works well.

After the accumulation buffer is calculated for a given frame, the accumulation buffer is converted into RGB pixel values. Several arbitrary values were picked as thresholds for a given color. If the value in the accumulation buffer falls between two thresholds, the resulting color is calculated by linear interpolation. For the dynamic heatmaps, a slate blue was chosen to represent where no one was looking. As more and more participants looked at a given coordinate, the color intensifies to green, then yellow, and finally red.

In order to give another form of visual measurement, the mean and standard deviation of the aggregate gaze coordinates of the twelve participants was calculated and displayed for each frame. The visualization consists of two white circles centered at the mean coordinate with radii of one and two standard deviations respectively. The smaller the circles are, the more concentrated the group focus is. Certain scenes and camera tricks in the video directed the group's gaze toward a particular object or two, while other scenes allowed the participants to gaze freely around the scene. The result of the dynamic heatmap construction and visualization is illustrated in Figure 3.2.



(a) 3D Calibration

(b) Exploring the Scene

(c) Concentrated Focus

**Figure 3.2:** The dynamic heatmap offers a clear picture of where the participants are looking as a group. During task-oriented exercises like staring at a 3D calibration circle, as in Figure 3.2(a), the effectiveness of how well the group performed can be quickly and subjectively evaluated. While watching the anaglyphic stereoscopic video, certain camera tricks, such as those in Figure 3.2(c), lend themselves to directing the group's focus to a particular object or two in the scene. Other scenes, such as the one in Figure 3.2(b), allow the participants to gaze freely about the scene.

# Chapter 4

# EXPERIMENTAL DESIGN

A within-subjects experiment was designed and implemented. Two versions of a computer generated video approximately three minutes in length were created to serve as the independent variable. The only difference between the two versions is that one was rendered in a standard two-dimensional format while the other was rendered in a red-cyan anaglyphic stereographic format. The ocular vergence response (disparity) while watching the two videos serves as the dependent variable. Twelve participants were shown the videos with six viewing the red-cyan anaglyphic stereographic version first and six viewing the two-dimensional version first. The vergence disparity while watching each movie was recorded over time and compared against the disparity while watching the other movie in order to discover whether a significant difference exists between the two.

## 4.1 Hypothesis

The null hypothesis assumed no significant difference in the amount of disparity between watching a two-dimensional video and watching an anaglyphic video.

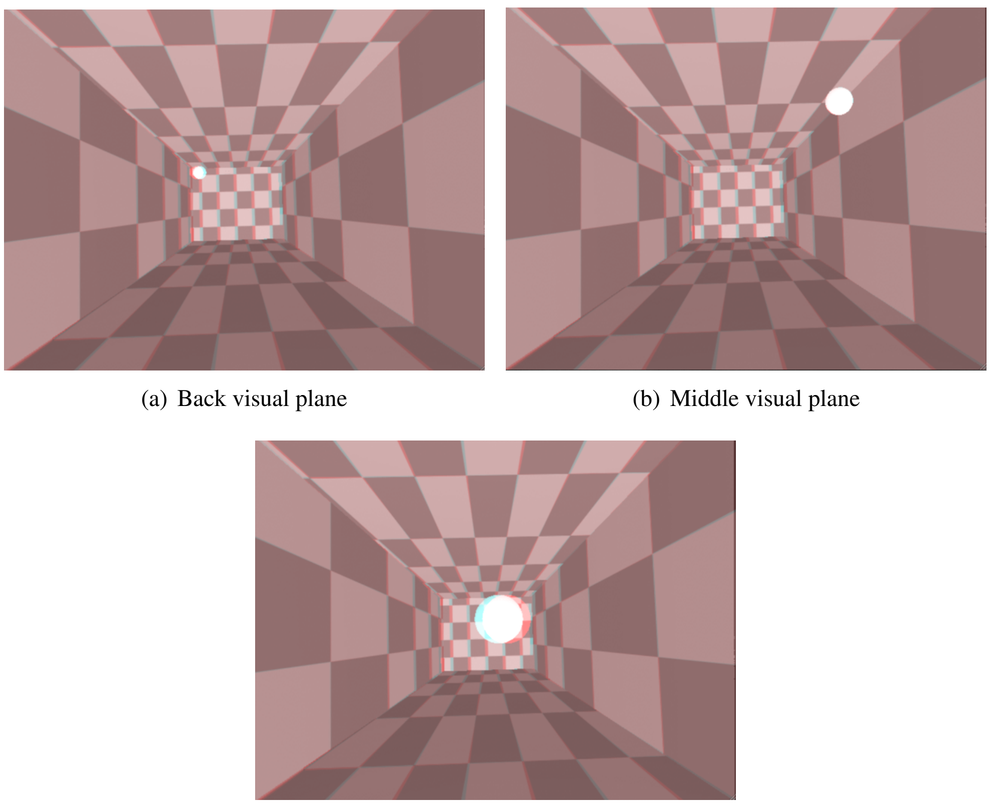
While Holliman et al. studied depth perception on desktop 3D displays [Holliman et al. 2007], they did not verify eye convergence using eye movements. Contrary to Holliman et al.'s conclusion that depth judgment cannot always be predicted from display geometry alone, we conjecture that ocular vergence could provide evidence for depth judgment.

## 4.2 Participants

Twelve college students, nine men and three women, ranging from the ages of twenty-two to twenty-seven participated in this experiment. They were recruited verbally on a volunteer basis.

## 4.3 Stimulus

Eight anaglyphic movies, and one two-dimensional movie were shown to participants. All nine movies were computer generated. One of these anaglyphic movies was a circle roving in three-dimensional space. The purpose of this movie was to try to elicit divergence as the circle sunk into the screen, and elicit convergence as it popped out of the screen. The participants were instructed to keep looking at the center of the circle as it moved around.



(a) Back visual plane

(b) Middle visual plane

(c) Front visual plane

**Figure 4.1:** During the 3D calibration video, a white circle visits the four corners and the center of each visual plane. In Figure 4.1(a), the circle appears to sink into the screen, while in Figure 4.1(b) it appears to rest on the screen. In Figure 4.1(c) the circle appears to pop out of the screen.

Six of the other anaglyphic movies were no longer than thirty seconds long. These movies were used by another researcher for a purpose unrelated to the present study.

The last two movies, one anaglyphic and one two-dimensional, were of the same three minute movie rendered using different techniques. The only difference was that one appears to be three-dimensional while the other is a conventional two-dimensional movie. In regards to vergence, the two-dimensional movie serves as a control stimulus compared to its anaglyphic counterpart.

(a) Two-dimensional version (b) Red-Cyan Anaglyphic Stereoscopic version

**Figure 4.2:** Two versions of the computer generated video were rendered. All participants were shown both videos with half seeing the red-cyan anaglyphic stereoscopic version first.

## 4.4 Apparatus

A Tobii ET-1750 eye tracker, accurate to one degree of visual angle, operating at 50 Hz with a 1280x1024 TFT flat panel display was used to measure eye movements and vergence.

## 4.5 Procedure

Participants first verbally agreed to participate in the experiment. Demographic information consisting of the age and gender of each participant was collected. Each participant filled out a short pre-experiment questionnaire regarding his or her familiarity with anaglyphic stereographs. A quick eye-tracker calibration was performed by having the participants stare at nine different locations on the screen. After calibrating the eye-tracker, the participants were presented with the stimuli, starting with the movie of a circle roving in a three-dimensional area. Next, the participants were shown three short movies, one of the long movies, three more short movies, and the second long movie. The order that the short movies were shown was rotated among all six short movies. The order that the long movies were shown was flipped every time. This provided for six possible orderings of the short and long movies, and each ordering was shown to exactly two participants.

# Chapter 5

# EXPERIMENTAL VALIDATION

## 5.1 Preprocessing

The first step in evaluating the gaze data for each subject was to extract the gaze data points that occurred during a fixation. Without fixating on a given coordinate, the calculated disparity between the left and right eyes is meaningless.

The disparity between the left and right gaze coordinates, as defined in (3.1), was calculated for every gaze point that occurred during a fixation. The disparity serves as the dependent variable for the analysis of the data. Furthermore, the disparity serves as a measurable effect of ocular vergence.

## 5.2 Analysis

Averging across displays, repeated-measures one-way ANOVA indicates that the display type (stereo or mono) has a marginally significant effect on vergence response ($F(1,10) = 5.85$, $p < 0.05$), with disparity averaging closer to 0 (-1.51e-04) under the stereoscopic condition than under the monoscopic condition (-1.28e-02).

No significant difference is observed when averaging across subjects ($F(1,11) = 0.02$, $p = 0.89$ n.s.), meaning that, on average, vergence response was about the same for all viewers. This analysis, however, only indicates that no particular individual exhibited a large relative difference in disparity between conditions. However, the analysis obscures the ‘polarity’ of the vergence response under stereoscopic and monoscopic conditions. To illustrate, if one individual responded with an average of 0 vergence response over time when viewing in stereo and with an average of -1 over time when viewing monoscopically, the overall average vergence response would be -.5. Similarly, if another individual responded equally but inversely (0 vergence response when viewing monoscopically and -1 when viewing in stereo), their average response would also be -.5. ANOVA would report no signifcant difference between these two individuals' responses.

An examination of individual gaze data reveals a stronger effect caused by the anaglyphic stereoscopic video on the ocular vergence response. A one-way ANOVA was performed on both smoothed and non-smoothed fixation points over the entire duration of the anaglyphic stereoscopic and two-dimensional videos. The non-smoothed disparity data for Subject 11 for both the anaglyphic stereoscopic and two-dimensional videos is shown in Figure 5.1. The mean disparity, or vergence response, along with the standard error for both videos for Subject 11 is shown in Figure 5.2.

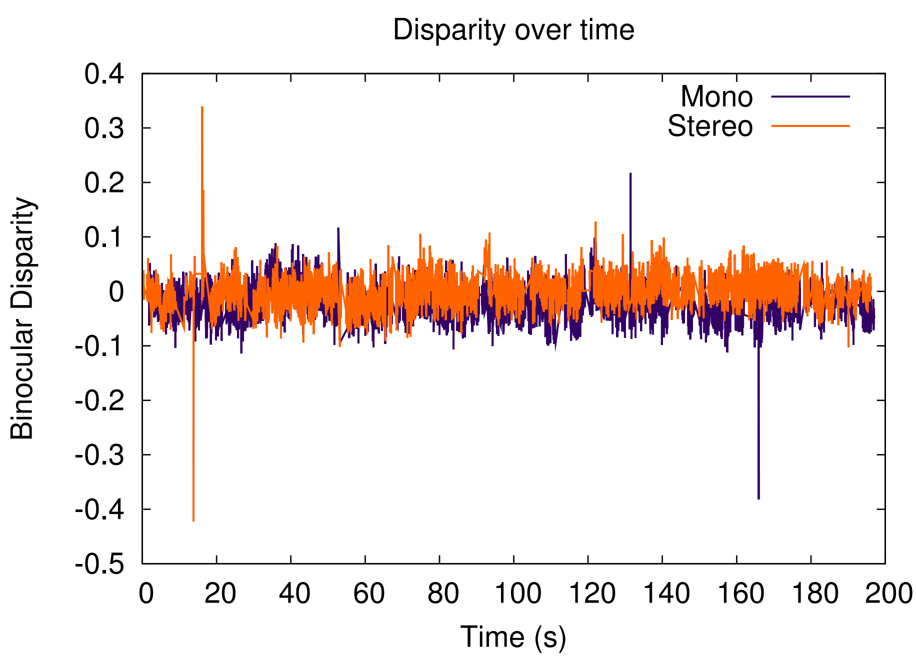



**Figure 5.1:** Non-smoothed disparity data for Subject 11 over both videos. By just looking at the graph, it is hard to tell whether there is any possible significance due to noise in the data caused by natural eye jitter.

In order to minimize the effects of eye tracker error, or outliers, a smoothing algorithm was applied to the dataset. A moving window of size 150 was used to compute the average gaze coordinate of the previous 75 and the next 75 gaze coordinates. Figure 5.3 depicts the

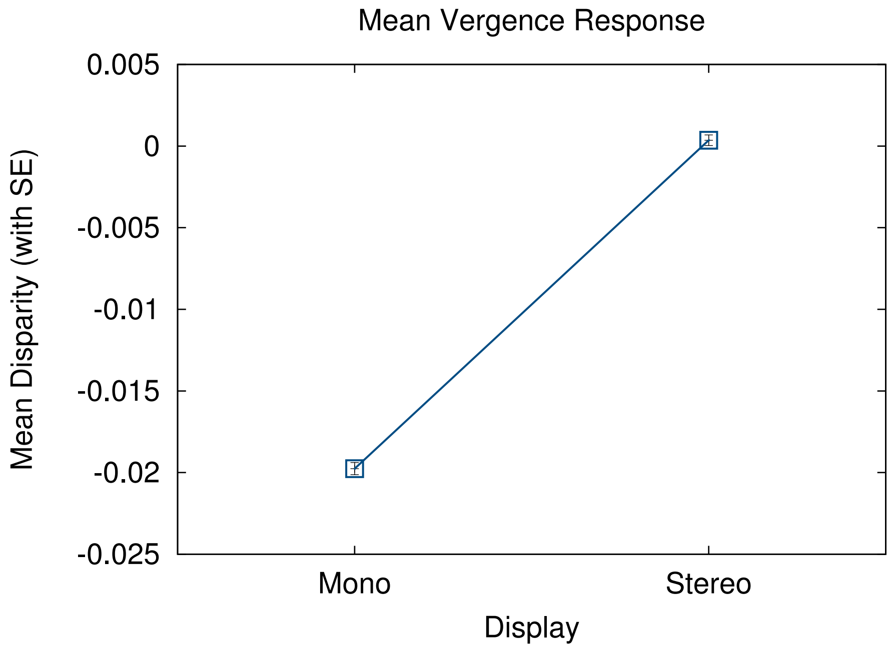



**Figure 5.2:** Aggregate statistics for the non-smoothed disparity data for Subject 11 over all time for both videos.

fixation gaze coordinates after the smoothing algorithm has been applied. Figure 5.4 shows the mean disparity along with the standard error of the smoothed dataset for both videos for Subject 11.

While the Figures 5.1, 5.2, 5.3, and 5.4 are all created from the scanpath data obtained from Subject 11, similar graphs exist for 10 of the other 11 participants.

For eleven out of the twelve participants, there was a highly significant ($p < 0.01$) effect of the anaglyphic stereoscopic video on ocular vergence. The results were inconclusive for only one subject.

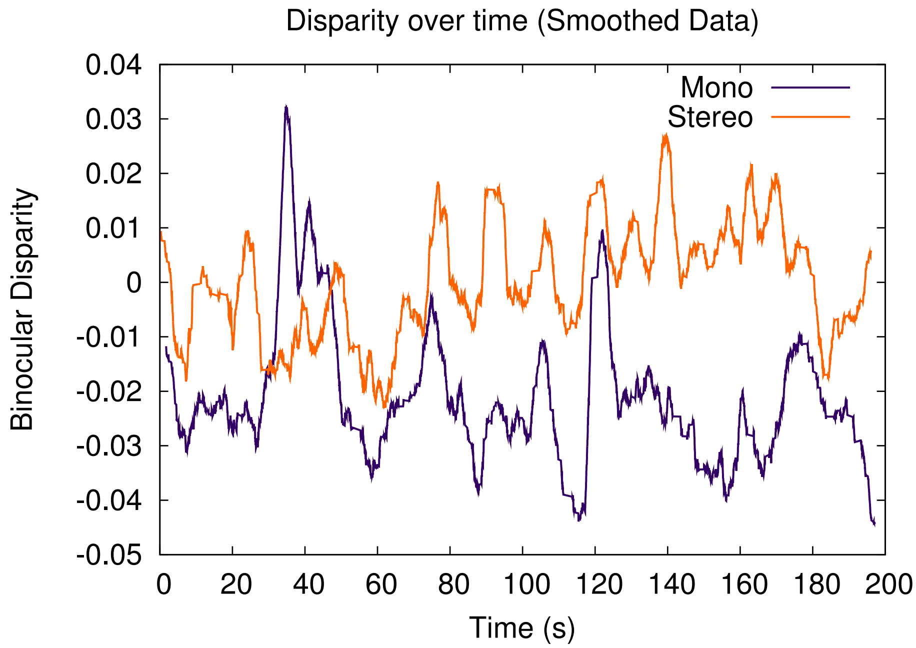



**Figure 5.3:** Smoothed disparity data for Subject 11 over both videos. By smoothing the data, and consequently reducing the noise, the graph becomes much clearer.

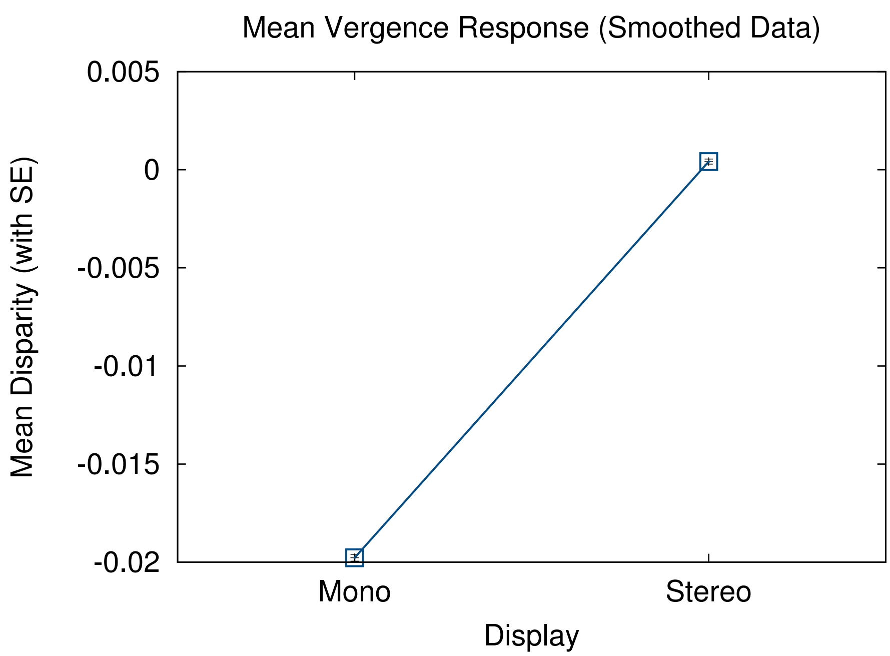



**Figure 5.4:** Aggregate statistics for the smoothed disparity data for Subject 11 over all time for both videos.

# Chapter 6

# CONCLUSION

Through experimentation, we conclude that there is a highly significant ($p < 0.01$) effect that the anaglyphic stereoscopic video had on ocular vergence when compared to a two-dimensional video for the majority of subjects. We further conclude that anaglyphic stereoscopic imagery provides an effective means of conveying the illusion of depth over a monocular camera approach for most people.

Consequently, we conclude that the Tobii ET-1750 eye tracker is capable of measuring the disparity caused as a result of the convergent and divergent responses.

## 6.1 Limitations

While the effect that the stereo video had on ocular vergence is highly significant when compared to the two-dimensional video for the majority of participants, there is no specific trend that can be determined. That is, some participants tended to diverge more while others tended to converge more. Since the stimulus was an anaglyphic stereoscopic video and no task was given to the participants, the participants were free to look at whatever part of the video that attracted the most attention. Some participants were drawn more towards the objects in the background of each scene, such as the end of a long hallway or the back wall of a large room. These types of stimuli appear to sink into the screen, thus evoking a divergent eye movement. Other participants, however, seemed to favor looking at objects that were closer to the camera, especially those that appeared to pop out of the screen. The closer stimuli that pop out of the screen would evoke a convergent eye movement. This preference, whether subconscious, intentional, or accidental is most likely the cause for a lack of a trend in the ocular response.

One possible reason that the results were inconclusive for one subject is that the subject could have impaired depth perception. Similarly, the participant may be partially color blind which would cause difficulty seeing the two separate images. Not being able to see the two separate images effectively, the human brain would have trouble fusing the two images together for a single three-dimensional representation of the scene. Since the eye convergence and divergence responses are evoked by the illusion of a three-dimensional representation of a scene, both impaired depth perception and color blindness would cause a lack of the eye convergence or divergence responses. The participants were not screened or tested in any way for any possible trouble with depth perception or color blindness.

## 6.2 Future Work

A replication of this experiment using a more controlled stimulus such as using the anaglyphic stereoscopic version of the calibration video compared to a two-dimensional version of the same video may be able to find a trend for convergence or divergence.

Perhaps replicating this experiment using the same stimuli could produce a highly significant conclusion if all of the participants were pre-screened for color blindness or impaired depth perception.

Similar techniques could be applied in an environment with an eye tracker and an autostereoscopic display in order to determine whether or not ocular vergence response occurs when viewing an image or video on the autostereoscopic display.

The data collected from the 3D calibration video could be compared against the data collected over the stereo and non-stereo versions of the three minute video. If the pixel separation of the red and cyan images were recorded for every pixel in every frame, a comparison between the eye gaze disparity and the pixel separation values could lead to a method of calculating how much the eyes converge given an object that appears to pop $Z$ units out of the screen. One of the large problems that would need to be solved is the issue of eye tracker accuracy when the participant is viewing a small object popping very far out of the screen with a large amount of eye convergence.

# Bibliography

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

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

ESSIG, K., POMPLUN, M., AND RITTER, H. 2004. Application of a Novel Neural Approach to 3D Gaze Tracking: Vergence Eye-Movements in Autostereograms. In *Proceedings of the Twenty-Sixth Annual Meeting of the Cognitive Science Society*, K. Forbus, D. Gentner, and T. Regier, Eds. Cognitive Science Society, 357–362.

HOLLIMAN, N., FRONER, B., AND LIVERSEDGE, S. 2007. An Application Driven Comparison of Depth Perception on Desktop 3D Displays. In *Stereoscopic Displays and Applications XVIII*. SPIE.

LIPTON, L. 1982. *Foundations of the Stereoscopic Cinema: A Study in Depth*. Van Nostrand Reinhold Company Inc., New York, NY. ISBN O-442-24724-9 , URL: <http://www.stereoscopic.org>.

PELI, E., REED, H., JINSHAN, T., AND DAN, L. 2001. A Binocular Stereoscopic Display System with Coupled Convergence and Accommodation Demands. In *SID International Symposium, Digest of Technical Papers*. Vol. 53.2. Society for Information Display, 1296–1299.

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