# Rendering and Perception of Depth Cues on a Multi-Focal Plane Stereo Display

Nicholas G. Timmons
Downing College

*A dissertation submitted to the University of Cambridge in partial fulfilment of the requirements for the degree of Master of Philosophy in Advanced Computer Science*

University of Cambridge
Computer Laboratory
William Gates Building
15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom

Email: ngt26@cl.cam.ac.uk

June 7, 2016

# Declaration

I Nicholas G. Timmons of Downing College, being a candidate for the M.Phil in Advanced Computer Science, hereby declare that this report and the work described in it are my own work, unaided except as may be specified below, and that the report does not contain material that has already been used to any substantial extent for a comparable purpose.

Total word count: —-

**Signed**:

**Date**:

# Abstract

This is the abstract. Write a summary of the whole thing. Make sure it fits in one page.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

This research has been performed to fulfil the requirements of the MPhil Advanced Computer Science course.

It is research into the effects of the introducing focal depth cues to rendered scenes through the presence of multiple focal planes on the accuracy of depth perception.

This is attempted with the construction of a custom stereoscopic multi-focal 3D display and development of software to configure and drive it.

One of the aims of this display is to counter the vergence-accommodation problemto reduce its impact on the perceived depth of the images being shown.

This is of particular interest to the study of VR(virtual reality) and 3D vision which in current commercial implementations present the user with a single focal depth which are in conflict with the vergence and stereo disparity cues of the image being shown.

Add some of the results and interesting things of the work here

This report includes a background of current techniques being used for VR and a short overview of similar methods using multi-focal planes.

There is then a breakdown of the requirements and techniques implemented in software and hardware for this multi-focal display with details of how they have been configured to achieve our results.

We also detail our experimental methods and results before summarising our findings and discuss future work which could take our research further.

# Chapter 2

# Background

Current trends in VR/AR Brief overview

Stereo displays such as those used for showing 3D movies and in HMD (head mounted displays), are in use in many commercial products and have seen significant improvement over the past decade ref but still have some large technical and usability problems related to rendering techniques which could restrict the wide-scale appeal of the products [17] and do not yet successfully mimic all the visual cues that the visual cortex processes when viewing objects in the real world.

When considering the comfort of the user the results of many psycho-physical and usability studies have suggested that the current solutions can lead to various problems for users of these systems. These symptoms include distortion of perceived depth [9], visual fatigue [3], diplopic vision ( commonly known as double vision) [5] and temporary degradation in the oculomotor response (as measured by slight movements within the eye) [6].

There are many factors contributing to cause these conditions varying from low quality images, such as the high persistence in the early Oculus Rift displays [24], incorrect interocular distances or the inability to allow the eyes to rest. A major cause that is often mentioned is the discrepancy between accommodation and vergence when using these displays with a fixed real focal

distance. In this context accommodation refers to the focusing of the eye on objects at different distances and vergence is the motion of eyes rotating to bring the convergence point of the optical axis to intersect at the distance of the object. These two oculomotor actions are coupled when looking at an object in the real world but cannot function correctly when decoupled due to being shown objects with stereo correspondence at one depth and vergence correspondence at another - which is the general case for objects shown on standard stereo head mounted displays (See Figure. 2.3).

Past research has suggested [9] that the breaking of the link between accommodation and vergence cues can lead to a decreased perception of depth, which will effect how the user understands the space they are looking at and could have an adverse effect on the perceived realism of the space as the scale may appear inconsistent with what the user is accustomed to. This is of particular importance to artificial reality environments where the virtual environment is mixed with the real world and depth cues from the display would have to both be correct to maintain consist with the real world.

## 2.1 What is realism in a 3D Display?

In the context of this research we are considering realism to consist of many separate visual cues. In a perfectly realistic scene the user would not be able to tell the difference between looking a scene in the real world and one that is rendered as the cues would all match.

To achieve this we would want a high quality rasterisation of a scene with correct lighting and reflectance within the full colour and intensity range that the human eye can perceive as well as being seen as 3D to the user through correct visual and depth cues.

Within the limited scope of this project we will not be able to develop all of those features but will be focused on improving the realism of a rendered scene through the use of more sophisticated depth cues to measure whether that improves how real the scene appears to the user when compared to a

scene which is lacking such cues.

## 2.2 Depth Cues

The human visual system has many cues for determining depth. Some that are visible in two dimensional displays (Table. 2.1) and some that are only possible with stereo or multi-focal displays (Table. 2.2). To attain maximum realism in the images shown the visuals would have to be delivering all of these cues [25].

| | |
|---|---|
| **Perspective** | Objects further away from the eye appear smaller. |
| **Known Sizes** | We can judge relative depths of objects from known sizes. For example, if an image shows a football and a house as the same size we assume the house is further away. |
| **High Frequency Detail** | We assume when we can see more small details that the object is closer. |
| **Occlusion** | An object that occludes another is closer than the object it is occluding. |
| **Lighting** | The human visual system is accustomed to seeing objects under different lighting conditions and can there reason the position of objects in a scene by the lighting and shadowing between them. It is also able to judge distance by the slight dimming of objects in the distance due to atmospheric scattering. |
| **Relative motion** | An object closer to the camera appears to move across the view faster than one further away. |

Table 2.1: Depth cues in two dimensional images.

For most people 'Binocular Disparity' is a very important depth cue which is relatively easy to replicate with two screens and as such is the cue that is most well represented by consumer 3D products at the moment, however conflicts

| | |
|---|---|
| **Binocular disparity** | The difference in the apparent position of objects under projection caused by the interocular distance between the two eyes. *(Stereo Only)* |
| **Accommodation** | The changing of the focal length of the lens in the eye to focus the view on an object at a particular depth.*(Muti-Focal Only)* |
| **Convergence** | Bringing the two images from each eye to overlap more coherently and face a specific point at a particular depth through rotation towards and away from the interocular midpoint. |

Table 2.2: Depth cues in stereoscopic and/or multi-focal images.

with other cues can cause detrimental effects to the visual cortex's ability to form a single image and extract the depth like it would when viewing the real world. Patterns such as repeating identical vertical bars are particularly problematic for finding stereoscopic matches without other cues and this can lead to visual strain and fatigue.

Weighting of depth cues REF VISUAL FATIGUE PAPER for Equations of measuring strength of depth cues???

## 2.3 Standard Stereo VR Implementations

The current standard for VR in a number of commercial hardware implementations consists of a head mounted display with the screen split vertically to display a separate image of the scene for each eye and a lens to distort the image to increase the amount of the screen that is visible and to reduce the "Screen Door Effect" [19] caused by the low pixel density and visible lines between them. The software is then implemented using interocular distances of the screen in the display with an appropriate field of view and a standard Off-Axis parallel projection for each eye (See Section 5.1.1).

The correctly calibrated projection of the same scene for the two separate

views gives a feeling of depth and “realism” through the user picking up on stereo-correspondences from the disparity in the two images and perceiving them as a single object at an expected distance.

Since the screen that is being used is a fixed short distance from the users eyes at all times the user is always focusing at a fixed focal distance which is different than the identified distance for the motion and stereo correspondences that are being shown in the scene.

This triggers the Vergence-Accommodation conflict from a conflict in visual depth cues. There are examples of trying to use simulated Depth of Field blur in early VR and large screen 3D for very near objects to try and simulate one of the missing visual cues but this was found to induce ”Simulation Sickness” and is it now strongly advised against [26].

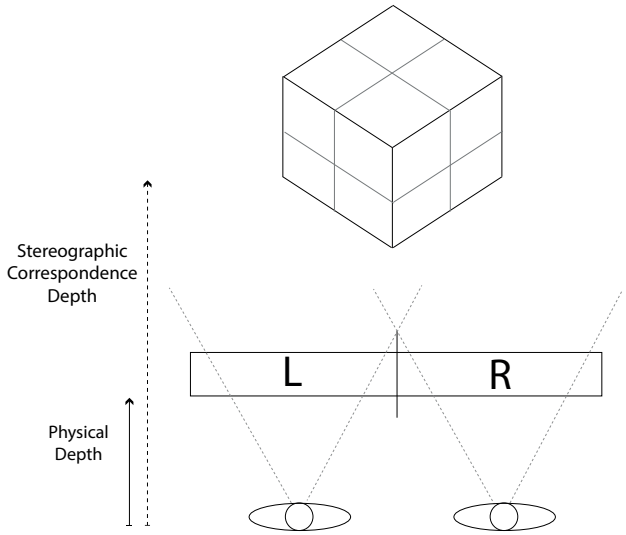



Figure 2.1: Basic comparison of the stereo and accommodation depths in a mismatched situation, such as a head mounted display.

There is currently a push in 3D technologies to try and increase the frame rate of applications that are using these techniques beyond the normal high bar of 60 frames per second. This has been seen as a way to reduce the discomfort some users feel from uneven or clunky motion caused by low or

varying frame-rates [22]. As resolutions and visual quality increases, making it harder to achieve, this is becoming a particularly important area to consider when developing for these technologies.

## 2.4 Vergence-Accommodation Conflict

The Vergence-Accommodation conflict is a source of discomfort and disorientation for a lot of users of stereo displays [11]. It is caused by a mismatch of the focal depth and vergence depth cues. It is a problem which originates from trying to simulate depths of objects which have a different focal depth in software leading to correct stereoscopic disparity between eyes but an incorrect focal depth for those objects. This is particularly a problem for standard stereo displays due to the use of a single focal depth because of the fixed display panel position.

This conflict can cause eye strain, headaches and dizziness in some users and is strongly associated with "simulation sickness" [26].

While this conflict does have some effect on the comfort of the user [17], in the area we are looking at we are more interested in the effect this has on how the users perceives the distances to objects in the world when this conflict is reduced.

It appears that when these visual cues are mismatched the users ability to judge the distance to objects becomes limited, vision can become less clear and the speed at which stereoscopic correspondences are matched is decreased. These symptoms can all interfere with the stereoscopic technique as it reduces the ability of the user to create an internal model of the scene correctly.

If the user is unable to judge the depth and position of objects clearly in the scene then it would not be comparable to the real-world and therefore, not realistic.

One of the main aims of using multi-focal cues is that it will reduce the

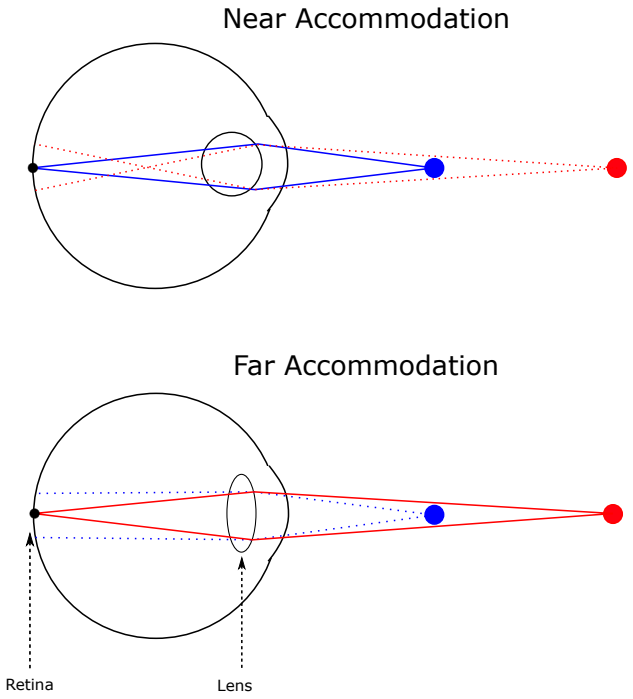



Figure 2.2: This diagram shows the basic physics behind the effect accommodation has to bring near and far objects in and our of focus by converging light towards the retina.

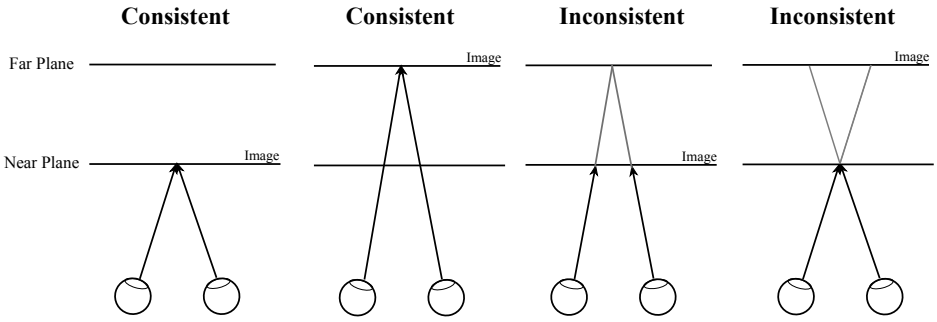



Figure 2.3: This is an illustration showing the different configurations which can cause a vergence-accommodation conflict. Any of the combinations which show a mismatch betwen the image depth and the convergence point can cause discomfort for the user.

confusion of depths and allow it to be properly processed the same as any real-world scene.

# Chapter 3

# Related Work

## 3.1 Multi-focal viewing

Multi-focal displays aim to simulate real world light ray angles to provide correct depth cues to the user. It is an area that has been investigated for over a decade but has recently became much more popular with the rise of VR as a viable platform due to the problems the single focal displays which they use cause.

The idea for multi-focal viewing is to use multiple displays, or sections of displays [12] and a series of lenses, mirrors and beam-splitters to combine images at different depths into a single image for the user to see. The different distances to the screens will then give the user different focal points to focus on that will behave like real screens at those distances.

Through clipping the viewed images to have only the sections at the correct depths displayed on each screen you can effectively create scenes with discrete visible focal depths.

These screens were used to investigate the comfort and physical reactions of the human eye when viewing simple scenes [7] and found to increase the comfort of the user and reduce the negative effects of single focal plane stereo

rendering solutions.

A downside to this method is that it is very sensitive to calibration and the images need to match perfectly to create the illusion of a single image, see Figure.2.1. This means that displays have to be invariant to user motion and match the physical position and separation of the users eyes at all time to avoid unwanted parallax between the two planes.

Alternative more complex methods have been suggested such a lenticular display setup [18], however this method suffers from crosstalk and, like many complex 3D viewing displays, requires eye tracking. Other alternative designs include methods which make use of the active stereo techniques to limit the number of screens [13] but this has its own implications on the frame rate and how active a scene can be, as well as restricting the user to wearing special glasses which can be a problem for users who already wear corrective lenses.

For less complex display models it is possible to avoid the eye tracking by using fixed depths and designing to accommodate different eye positions - which can be quite varied among a population.

There was a study carried out by the US Army which investigated the interocular distance of its members and gives some very good data on the topic [2]. They showed a mean distance of 6.47cm for men and 6.23cm for women, see Table 3.1. Using this data work has been carried out to support many users with varying interocular distances with adjustable hardware [15]. There was a more in depth analysis of the mean interocular distances which took a more widespread look at the general population, including children [8] which are not represented in the US Army study. The general results within adults fell in line with what we see from the US Army work. As we are testing on adults we will be using the data for adults only to configure our display.

The downside to this type of method is that the user is required to maintain a calibrated head position through out to keep the effect working. To prevent the movement of the user causing problems some very *interesting* solutions have been used in research such as bite-bars to bite down on to prevent the

users head from moving [7]. However, these are a little impractical for a study covering multiple users which will need to support multiple distances and interocular distances.

| | Male | Female |
|---|---|---|
| Mean | 6.47cm | 6.23cm |
| Min | 5.20cm | 5.20cm |
| Max | 7.80cm | 7.60cm |

Table 3.1: Interocular distances from the US Army data [2]

## 3.2 Depth Perception

There have been a number of papers investigating the effect of stereo displays on perceived depth. One such study performed an investigation into the effect of different fields of view when using a stereo setup and how that changes our interpretation of the depth which points to the conclusion that relative scale and rate of change of scale is important in our perception of depth.

In one experiment [21] were able to show that a user in the correct configuration could correctly navigate to a specific position in the scene, where as a user who was not calibrated would have difficulty due to "minification".

This shows the ability to accurately determine the depth and position of objects is affected by depth cues, whether they are from stereo correspondences, projection or otherwise. Which gives us reason to believe that the accuracy and depth perception can be improved by the introduction of further depth cues. Particularly, focal length was shown as a way to determine the depth of objects from separate images static scenes under varying levels of focus [1].

# Chapter 4

# Hardware

The goal of this implementation is to create a display which we will be able to use to investigate the relative perceived realism of the scene through the effectiveness of depth perception when given the extra focal depth cue compared to standard stereoscopic rendering and allow it to be easily configurable for more specific tests in the future.

## 4.1 Display Requirements

The display is designed to allow us to investigate perceived realism for this work but also to be configurable for to support future work beyond this project.

This project requires that the display supports four high resolution good quality displays which are able to be viewed at configurable focal distances in a stereo setup that allows the user to view only two screens per eye and support a wide range of users. This is to allow the testing of combinations of zero disparity, stereoscopic and multi-focal scenes.

To support male and female users it would be ideal that the displays were configurable to support both adult male and female average interocular distances.

## 4.2 Display Design

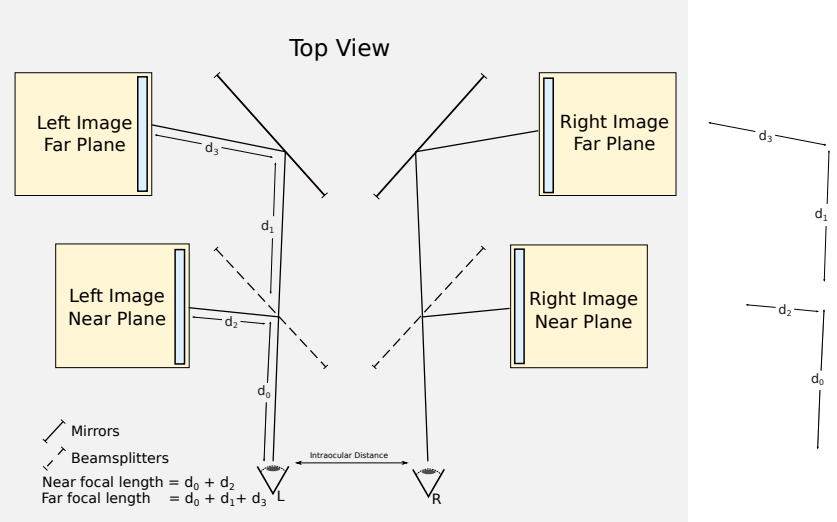



Figure 4.1: A high level view of the concept layout to achieve the stereo multi-focal images.

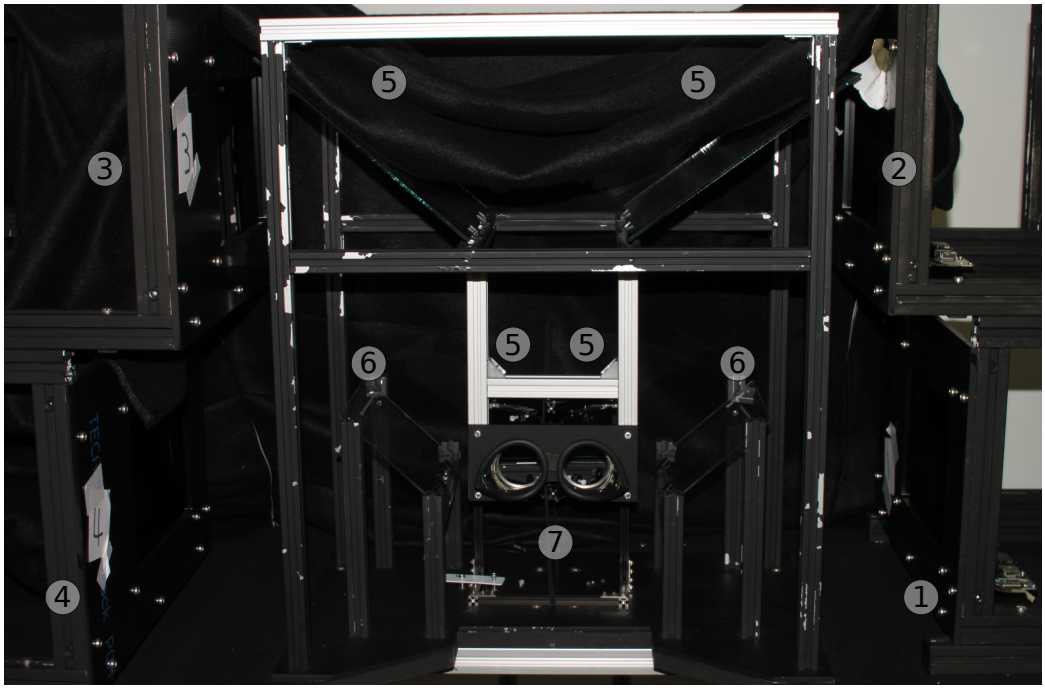
To achieve the requirements of the display we will be using four high resolution displays to get maximum quality in rendering per view and reduce visible problems from low pixel density. Full details of the panels in Table. 4.1.

These will be configured with two screens per eye. One at the distance of the required near plane and one at the distance of the required far plane. For each side of the display the images are seen through a combinations of mirrors and a 50/50 beam-splitter which merges the two images into one and then another mirror directs the combined image to the users eye. See Figure. 4.1 for a more detailed plan of the physical layout.

As the image is merged through the beam-splitter it is essentially additively combining the colour values of the two screens. To ensure a clear image of both views we need to isolate the beam-splitter, mirrors and all screens from external light to prevent obscuring or offsetting colour as it is merged towards the eye. To attain this we will be shielding the constructed display with matte black boards to prevent any light entering the display from outside and to

reduce internal reflectance.

An extra consideration for combining the images and maintaining stereoscopic correspondences is that each of the images being displayed on each screen must be matching in colour range and intensity when they reach the eye. This will mean that the displays will have to be calibrated for any differences in the displays as well as error from light absorption from the mirrors or beam-splitter. Additionally this will also help mitigate the error for any external light or internal reflection we are not able to remove.




1 Screen 1 (Right near plane) 3 Screen 3 (Left far plane) 5 Mirror

2 Screen 2 (Right far plane) 4 Screen 4 (Left near plane) 6 Beamsplitter

7 View goggles

Figure 4.2: The front view of the display with the mirrors, beam-splitters and screens labelled.

### 4.2.1 Display Panels

The display panels used in this display are replacement screens for the iPad 3. These have been chosen due to the high pixel density and good colour reproduction for the price point.

These displays are also a good size for our test setup. Any larger screen would have made the desk prohibitively large and impacted other features we would like to look into in the future, such as HDR support.

| | |
|---|---|
| Resolution | 2048 x 1536 |
| Width | 169.5 mm |
| Height | 240.0 mm |
| Type | LED-backlit IPS LCD |
| Colours | 16 million |
| Model Number | LP097QX1 6091L-1579A |

Table 4.1: iPad Display Panel Specification

The panels are mounted into the display and fixed in place on the near and far planes. They are powered and controlled through a 3rd party control board and each screen is connected to the PC using *mini display port* with the two left screens being connected to one GPU and the two right screens to the other.

### 4.2.2 Display Configuration

The display has ten configurable components to allow support for varying distance and angle from the screens to the eye.

The screens themselves are on fixed beams that allow the screens to slide to be nearer or further from the mirrors so the total distance to each screen can be easily calibrated.

There are two mirrors and a beam-splitter for each side of the display. All three components can be rotated and skewed to reach alignment.

Additionally the mirrors that are placed directly in-front of the user can also be translated to account for offsets in eye position

ref diagram / new component diagram ?.

### 4.2.3 Machine Specification

Although this display will work on any machine supporting 4 *mini display ports* it is worth mentioning the setup we are using and why for easier replication.

| | |
|---|---|
| CPU | Intel(R) Core(TM) i7-4790K CPU @ 4.00GHz |
| GPU | 2 x NVidia 970 |
| Memory | 16GB |
| OS | Ubuntu 14.04 LTS |

Table 4.2: Controlling Machine Specification.

**CPU** The CPU in our test machine is a little over specification for what is required as we will only be running simple scenes, however, we want to ensure that any analysis or logging applications we are using alongside the main multi-focal program will not interfere with the performance of the display. It also allows for more other research which may involve more intensive video processing that may not be suitable for the GPU.

**GPU** This machine has two GPU's to allow for a lot of monitors to be attached. In this case we have four display panels for the multi-focal display and one monitor attached to control the testing. In the future we would like to expand this to allow for up to another four outputs to possibly allow the iPad panels to be used as High Dynamic Range displays using additional projectors [10].

We have quite powerful GPU's due to the need to render to quite a high resolution which requires a lot of memory and to process potentially very complex scenes.

**OS** We have chosen to use Ubuntu on this machine to maximise the support for research projects in the academic community and maintain compatibility with other projects in the Cambridge University Computer Laboratory.

### 4.2.4 Known limitations

**Limited Field of View.** As our screens will only cover a limited field of view when compared to the full coverage of HMDs, so we are limited in our ability to give as full an immersive effect, rather for full scenes it will look like peering through a window into another room than the full VR experience people may be accustomed to. This could be improved in a less experimental setup through the use of lenses or different focal lengths.

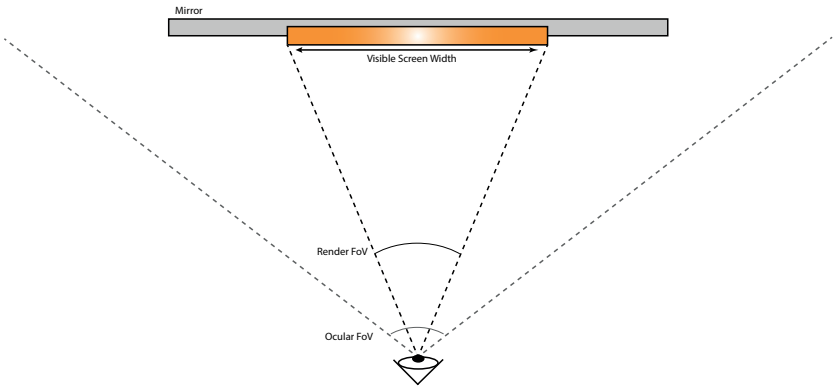



Figure 4.3: This image demonstrated the limited field of view when using this method to display multi-focal stereo.

**iPad Displays** While working with the displays we have found that the glossy coating on the screen increases light pollution within the display and from some angles within the display there is a clear reflection of another screen. This can be mitigated with more obstructing panels to prevent it interfering with any testing but it raises the potential improvement from using matte display panels in the future.

**Relative Head Position** As this display is not physically attached to the user the user is able to slightly move the head which will cause parallax from misalignment with the screens. This is especially true when using all four displays as small head motions will result in a large change of position of the near screen in the view direction. To alleviate this, extra consideration will have to be made in software design and user testing to ensure correct calibration throughout the use of the display.

**Large display.** The users view eye coverage is limited by the size of the screen at the furthest distance. This means that if we want to have a display which covers all of the user's view then the screen would have to be of the correct size to cover the full view at the given focal distance. Because we are using iPad panels in our display we are limited to quite a small view of the scene due to the limited size. With a more complex setup using custom lenses it would be possible to reduce this problem but it would be more expensive and less easily modified and is therefore out of the scope of this project.

**Benefits of head mounted displays** As this display is desk mounted instead of head mounted we lose the ability to do head tracking and rotation which greatly help the user by giving subtle relative motion depth cues. The fixed view point means we need to use motion of the scene rather than motion of the user which is less intuitive. If user movement was possible it would have been particularly useful for measuring how small motions may help determine depth.

**Varying resolutions** Due to the cameras being at physically different distances and no use of lenses to increase the size of the displayed screens we have a reversal of the ideal resolutions. In the best case we would want the objects which are appearing closest to the user to have the highest pixel density and the objects further away to have the lowest. In our display the image which is closest is scaled down to match the distant screen and as such is only using a small portion of the possible resolution.

**Head motion** As this is a desk mounted display the head position in the scene should be fixed. Motion of the camera in VR without actual head motion can cause motion sickness from mismatching motion cues from the ears when compared to the what the scene shows [22]. So in order to give the user cues of depth from relative motion we are relying on objects in the scene moving relative to the camera and avoiding any scenes which could feel like the user is moving. This may mean relying on static object to be points to ground the user perception. Correct depth in the scene is only maintained when looking at the centre of the view so to avoid the user from moving focus around the displayed image we will want objects to try and only be in motion in a single position unless we are measuring how distance tracking is changed when they are in motion.

# Chapter 5

# Software

## 5.1 Software Requirements

Configuration: Four render outputs - One per screen Full standard rendering support - To be able to support objects to test depth + textures and what have you. Configurable positions for each screen - For alignment. Adjustable scene depth planes - to allow testing of mismatch.

In addition to the blending we also need to ensure that the field of view being rendered in the scene is the same as the real field of view of the human eye. Investigations have shown [21] how decreasing the field of view causes the user to misjudge distances. This could effect the perceived distance to the near plane and as such in our software the field of view will have to be correct to maintain that depth cue accurately.

**Rendering** As this is implementation will require the use of multiple outputs and be displaying real-time 3D it will also need to support 3D mesh, texture and shader loading as well as lighting control. For this we will be using OpenGL as the graphics API and building support onto that.

**Testing** The software will need to be able to support real-time configuration of the screen position within each display panel to support multiple users during testing. This will mean accounting for error from head position, interocular distance and slight motion during the test.

It will also be required to support and switch between different rendering modes without significant delay to allow for comparison of different techniques.

### 5.1.1 Projection Modes

When rendering stereo pairs there are a number of techniques available that each have unique features. The aim of the stereo pair is to simulate the viewing conditions of the user in the real world and rely on having the plane of projection at a fixed convergence distance and at that distance there should be zero vertical or horizontal parallax between the left and right eye.

Vertical parallax is generally avoided due to it causing discomfort and effecting stereo correspondence [4] so in our setup there should only be a very small amount that should differ per used as the eyes sit approximately on the same horizontal plane but this varies person to person. This error can break the illusion, cause diplopic vision and cause discomfort in the user. However a certain amount of horizontal parallax is required for the stereo effect to work. This comes from positive (Figure. 5.2) and negative parallax (Figure. 5.1). Positive parallax is when the point is beyond the projection plane and negative is when the point is in front of the projection plane. Generally points with positive parallax are more comfortable and we aim to avoid too much negative parallax as it can quickly become extreme or cause clipping from the screen when it moves very close to the eye position.

For early testing of the display we are going to support three main projections of the scene, 'Toe-In', 'Off-axis' and 'Oblique'.

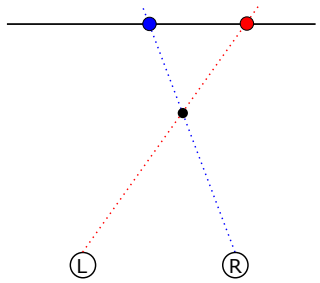



Figure 5.1: Example of negative parallax

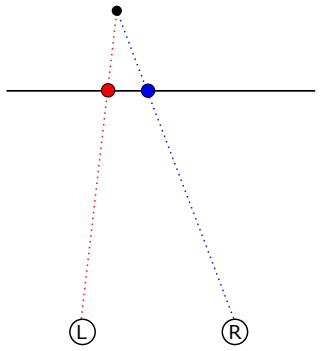



Figure 5.2: Example of positive parallax

**Toe-in** This projection mode points both cameras at a single focal point. This method produces reasonably correct stereoscopic vision on the projection plane but points in front and beyond that plane and particularly towards the left and right side of the image suffer from quite severe vertical parallax caused by the different rate of change of depth due to the non-parallel projection planes. This conflict is increased when the field of view of the viewing camera is decreased as the difference in depths between the left and right images become more prominent nearer to the centre of the image.

**Off-Axis** Off-Axis projections corrects the non-parallel projection planes of the 'Toe-In' method and forces the camera direction to be parallel. These changes remove all vertical parallax making for a more comfortable viewing experience. This method requires the creation of non-symmetric camera frustums which can look incorrect when viewed individually. An added benefit of this method which is used in 3D cinema experiences is to alter the

$$\begin{bmatrix} \frac{near}{right} & 0 & 0 & 0 \\ 0 & \frac{near}{top} & 0 & 0 \\ 0 & 0 & \frac{-(far+near)}{far-near} & \frac{-2far*near}{far-near} \\ 0 & 0 & -1 & 0 \end{bmatrix} \tag{5.1}$$

Figure 5.3: Symmetric simplifed OpenGL matrix (OpenGL depth is in the range -1.0 to 1.0) [27]

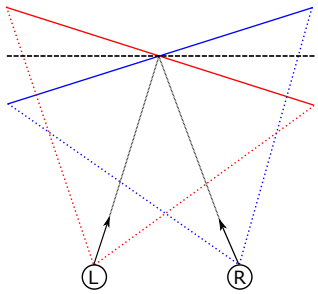



Figure 5.4: Toe-in projection layout

interocular distance to exaggerate depth.

$$\begin{bmatrix} \frac{2near}{right-left} & 0 & \frac{right+left}{right-left} & 0 \\ 0 & \frac{2near}{top-bottom} & \frac{top+bottom}{top-bottom} & 0 \\ 0 & 0 & \frac{-(far+near)}{far-near} & \frac{-2far*near}{far-near} \\ 0 & 0 & -1 & 0 \end{bmatrix} \tag{5.2}$$

Figure 5.5: Standard frustum OpenGL matrix [27]

**Oblique** Our oblique projection is very similar the Off-Axis projection in that it also removes vertical parallax on the projection plane by sharing the same plane between the two eyes, but it has the added benefit of better

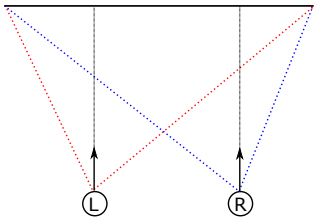



Figure 5.6: Off-Axis projection layout

modelling vergence for the central position on that plane?. It also keeps the benefits of the Off-Axis projection while emphasising depth due to the converging eye direction. This leads to increased disparity between the left and right image as objects move towards and away from the projection plane. In our implementation the oblique projection is an accurate model of vergence at the projection plane as an alternative to having the angle of the display panels able to rotate.

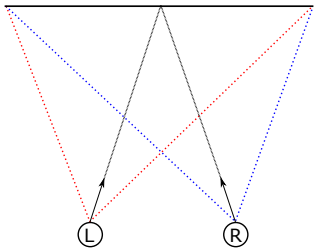



Figure 5.7: Oblique projection layout

Should I put the derivation of the Oblique matrix in here?

### 5.1.2 Depth Configurations

The scene is being modelled in real world units to easily match our physical configuration and allow for comparison in the future to real world scenes such as those taken from light field cameras.

When rendering the scene all the points closer than the near focal distance will only be displayed on the near screens and only the points beyond the far focal distance will be displayed on the far screens.

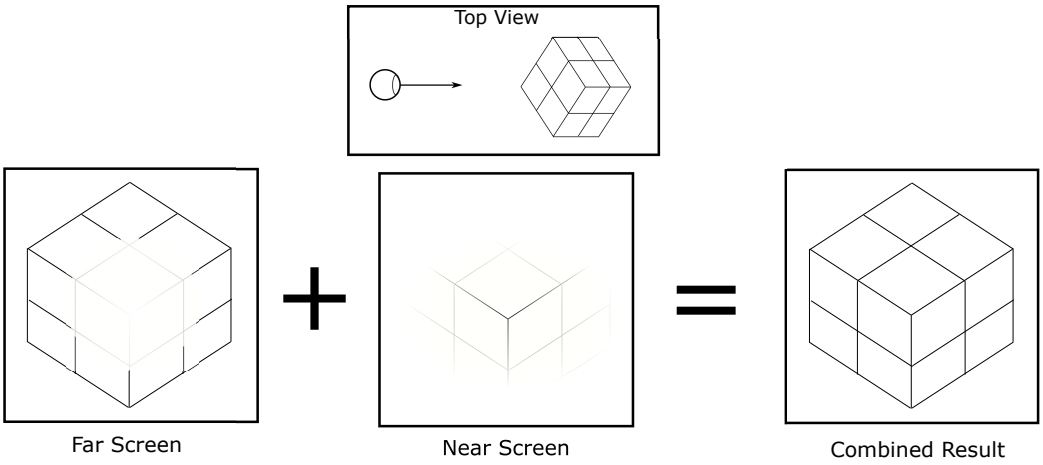



Figure 5.8: This is an illustration of how the near and far planes have different images rendered on them for each pixel depending on depth. In this example the cube is having the nearest corner which is in front of the near plane rendered onto the near screen and then the parts of the cube which are beyond the far plane rendered onto the far screen. The points which exist between these two depths are blended between the two screens. When these two images are combined through the beamsplitter and presented to the user they appear once again as a whole image, however the appropriate parts of the image will now be at the correct focal depth.

For the points which lie in the distance between the two planes we do not have a physical screen to display them at the correct focal distance so we will try different blending methods of the two distances to try and determine if it is possible to convince the user that these points exist at the appropriate distance between planes.

As we are using beam-splitters which work additively we display black for points which are being displayed on the other screen and not within the blend depth range.

The primary methods we will try are:

**Box:** All points less than half way across the middle space will

be considered on the near plane and all the ones more than half way across will be considered fully on the far plane.

$$
\begin{aligned}
n &= \frac{n_i - n_1}{n_2 - n_1} \\
col_{out} = f(n) &= \begin{cases} col_{near} & \text{if } n \text{ is } < 0.5 \\ col_{far} & \text{if } n \text{ is } > 0.5 \end{cases}
\end{aligned}
\tag{5.3}
$$

We expect this to produce the effect of the scene feeling like it is made out of two pictures shown to the user. As there is a clear "focal seam" where the edges of each depth are visible and this is exacerbated by any calibration error.

**Linear:** As the points move across the middle space they will be linearly interpolated between the two views.

$$
\begin{aligned}
n &= \frac{n_i - n_1}{n_2 - n_1} \\
col_{out} &= (n * col_{near}) + ((1 - n) * col_{far})
\end{aligned}
\tag{5.4}
$$

**Non-linear:** As the points move across the middle space they will be non-linearly interpolated using a modified sigmoid curve between the near and far plane.

$$
\begin{aligned}
n &= \frac{n_i - n_1}{n_2 - n_1} \\
blend &= \frac{1}{1 + exp((-n * 2 + 1) * 6))} \\
col_{out} &= (blend * col_{near}) + ((1 - blend) * col_{far})
\end{aligned}
\tag{5.5}
$$

**Projective:** In this mode the calculation of the depth interpolation value through the blending area is scaled inversely to match

the depth divide in perspective projection transform.

$$n = \frac{(1/n_i) - (1/n_1)}{(1/n_2) - (1/n_1)} \tag{5.6}$$
$$col_{out} = (n * col_{near}) + ((1 - n) * col_{far})$$

**Fixed:** Fix all to either the near or far plane. This method mostly exists as a way to test and configure the views and will only be used in the testing as a comparison to none focally split images.

$$col_{out} = col_{near} \quad \text{or} \quad col_{out} = col_{far} \tag{5.7}$$

The aim of the blending is to produce a sum result of combined rays which would approximate the rays from the target distance and to make the shift from one viewing plane to another less noticeable.

### Blend Comparison

It showed in our early testing that linear and non-linear blending gave inaccurate results when compared to the projective blend. The linear and non-linear blends caused a false sense of depth as objects remained in the near focal image too long and the depth appeared inconsistent as well as showing a visible seam.

The non-linear blend was particularly susceptible to calibration errors similar to that of the box blend as the change in the focal plane was too quick in the middle ranges causing very strong border artefacts.

It makes sense that these symptoms were relieved with the projective blend as it is mapping the change in focal depth to the same '1/Depth' projection that is used in the projection matrix. Combining this with the linear depth

calculation we are maintaining correct focal depth consistency throughout the blend.

RAFAL: Does this need pictures of the software showing the separate blends?

### 5.1.3 Reflection Depth

In our setup we are interested in providing correct depth cues through light rays reaching the eye from the correct focal distance.

For a given diffuse object when light hits it the light is scattered with varying amounts of uniformity which results in the light hitting the eye with an angle appropriate for distance to the object.

This is different for reflective objects where a portion of the light is directly reflected without diffusion towards the eye. In this case the rays of light are arriving at an angle similar to that of the object at the distance to the reflecting object plus the distance from the reflected object to the source of the reflection.

As we are not modelling the scene to take into account multiple reflections or the depth of those reflections, we are not able to successfully map these reflections onto the depth that is being used to split the scene into different focal ranges.

A naive approach could be attempted but any mismatches could potentially break the illusion for the surface we are mapping. REWORD We will be able to fully test whether this method of distance splitting is effective using a purely matte test scene.

Show diagram of diffuse distance vs specular distance

### 5.1.4 Rendering costs

#### State Change

For rendering simple scenes with OpenGL a high proportion of the costs can be the switching of GPU render state [20] which can cause stalls if enough data is not being submitted and the GPU is sat idle during the process.

When using two cameras the common approach is to render all the left view and then all of the right view. This means that the camera state is being changed many times per object per camera view.

A more optimal approach is for each set of objects being rendered switch the currently bound camera and render target (or multiple render targets and mask out the opposing view).

In this configuration it is only a maximum of two sets of camera and render target state changes per object instead of twice the textures, meshes and data per object.

$$\begin{aligned} &m = \text{Object count} && c = \text{View count} n = \text{State changes per object} \\ &\text{Per view} = c * m * n \\ &\text{Per object} = m * n + m * c \end{aligned} \tag{5.8}$$

#### GPU Memory Usage

Memory consumption is an important consideration when using multiple screens, particularly at a high resolution and high bit rate.

The display being used with this software requires four screens outputting at eight bits per pixel so we start with a minimum of 96MB of GPU memory being dedicated to the screen buffer.

We would like to render in a linear XYZ colour space High Dynamic Range setup to allow for future tests with HDR displays. This means we are required to use a linear (non-srgb) format. When we come to present the rendered result on screen will be transforming that result with an XYZ to RGB colour space matrix and gamma correction, as such as want to ensure that each texel colour has a high enough bit-rate to ensure that the rendered result will be transformed accurately into its corresponding RGB colour value without banding or artifacts in the resulting image. It is because of this we will be using one of the floating point formats and aim for a higher bit rate per pixel. As this is purely an experimental setup we have selected to use the 32 bit per pixel floating point format, GL_RGBA32F, to ensure maximum quality and reduce the chance of errors from unwanted quantisation before the transform to RGB colour space. This has big impact on the total memory being used for the render buffers as it is four times as large as the standard RGB8 formats but it is worth the cost to ensure correct results in our tests and to ensure that any colour calibration can be correctly supported now, or in the future.

| Texture | Resolution | Bits/Pixel | OpenGL Format | Size (MB) |
|---|---|---|---|---|
| Near Plane Render | 2048 x 2048 | 32 | *GL_RGBA32F* | 128 |
| Far Plane Render | 2048 x 2048 | 32 | *GL_RGBA32F* | 128 |
| Near Screen Buffer | 2048 x 1536 | 8 | *GL_RGBA8* | 24 |
| Far Screen Buffer | 2048 x 1536 | 8 | *GL_RGBA8* | 24 |
| **TOTAL:** Single GPU | | | | 304 |
| **TOTAL:** Both GPUs | | | | 608 |

Table 5.1: Memory consumption of the display buffers used in this display. The large linear colour buffers used for rendering the scene are the most costly. We are using a significant portion of the GPU memory with just the rendering pipeline. If we were to test a more complex scene this could become prohibitive and accuracy of the colour calibration transform might need to be sacrificed.

## 5.2 Software Configuration

### 5.2.1 Rotational Consistency

A major part of this software is deciding where we need to partition the images being displayed. That partitioning needs to be correct for the real world distance to objects and should only take into account position of the user view and distance to the point in space.

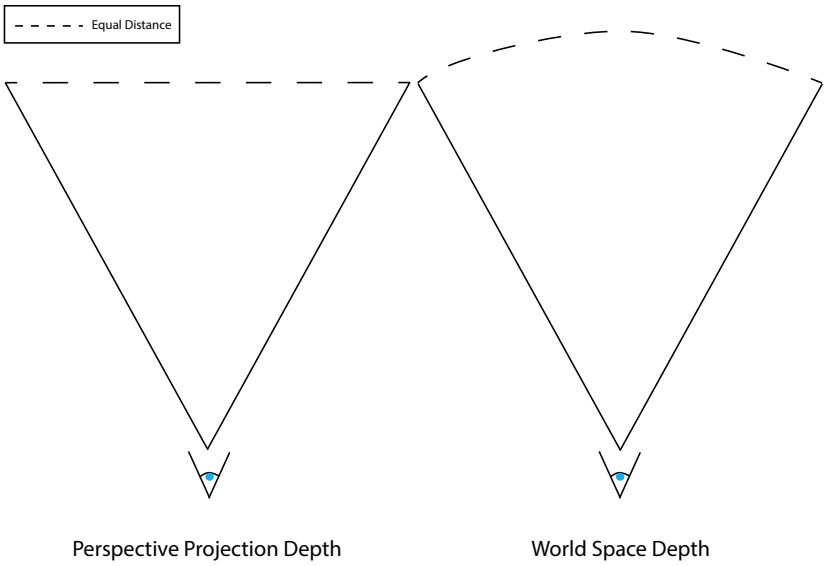



Figure 5.9: This is an illustration to show the different of position relative to the eye position of the same depth when using projective depth instead of world space depth.

The depth that is produced from a standard projection matrix is not the real world scene depth. The non-linear divide gives a non-linear depth from zero to one when what we need for calculating depth is the real distance from the ocular centre to the point in the scene. The perspective depth also gives all points at the far plane an equal distance from the camera, as shown in Figure. 5.9 which causes changes in depth as the screen is rotated as points move from being on the edge of the far plane to the centre.

To calculate accurate depth, we will be using the objects positions multiplied

by the world matrix to get its world coordinates and then subtract from them the world space position of the camera and calculate the length of the resulting ray.

This will give us a linear depth to each rendered point in the scene which will be consistent under any scene projection, translation or rotation.

Show image of different depth

### 5.2.2 X11 Window Controller

With modern Operating Systems a windowed applications maximum frame rate is tied to the refresh rate of the desktop. This would not be suitable for our application as we need to maximise frame rate to maintain realism and ensure persistence of vision.

Only full-screen applications are completely decoupled from the desktop screen refresh rate. Since we are using multiple discrete GPU's it is not possible to create a graphics context being controlled by both, so we need to support multiple full screen context windows which is not possible in the standard "extended desktop".

In our setup we are using X11 as the windowing system which supports a mode to allow the user to run multiple independent screens rather than just extending one and the user can select which screens are controller by which GPU. Using this we are able to launch windows directly on the separate screen and run them full screen which allows us full control of the presenting of rendered images up to the maximum supported refresh rate of the displays and GPUs.

We want to remove as many barriers to presenting to the screen as possible to try and ensure that the screens are kept in sync and we do not fall out of alignment which could cause strange effects when objects are in motion,

### 5.2.3 OpenGL MRT's

When rendering to multiple screens we want to try and do it as optimially as possible to reduce wasted CPU or GPU time so that this display can support as complex a scene as possible.

The hardware setup has two GPUs powering four screens in a non-SLI configuration. As we are not able to share data between the two GPUs we are forced to render the scene at least twice. As this is also a requirement of rendering stereo views without reprojection this is not a limiting factor in performance.

In order to take advantage of the data sharing we do have available we can make use of multiple render targets in our shaders where it is possible so that we only processing the vertex data once per eye and then in pixel shader we perform our depth calculation and write the appropriate blend of the lighting value to separate render targets representing the near or far screen. This limits our lighting calculations to once per eye and by running through the same GPU and OpenGL context we can be more sure of matching VSync on both screen so we are less likely to suffer from screen mismatches from the screens not being synchronised.

This is more of a factor for matching the views between each eyes as we currently have no method to ensure that the left and right views are refreshing on the same schedule. To overcome this, we are using screens with a high refresh rate and low persistence and ensuring the test scenes are running above the required frame rate so that even if they are out of sync the difference between the two scenes should be very low.

If we are not able to share context across multiple screens then the more costly method of using four individual contexts will have to be used. This would involve storing the data twice on each GPU and will prevent any fragment shader optimisation to write to multiple render targets, essentially doubling the pixel shading cost of the application.

### 5.2.4 Colour Calibration

When sharing the rendering of an image across multiple displays it is important that the parts of the image from each display are indistinguishable by colour and intensity.

In our display we are using four separate displays and a number of mirrors and beam-splitters. Each of these can interfere with the colour and intensity of light that is reaching the user and to maintain the merging of images for multi-focal rendering and correspondences for stereographic rendering this has to be avoided.

The source of error in the displays come from how they are provided from the manufacturer. Some are produced with variance in light intensity and colour ranges.

The beam-splitters suffer error as light is absorbed or diffused in a direction other than the expected reflected direction. To try and reduce this we are using high quality first surface mirrors to prevent diffusion or absorption within the protective medium. They are more susceptible to damage but as in our design they are covered and can be calibrated without touch them they should be safe from damage or problems.

To correct for this error we need to calibrate the screen to show the true colour being represented to the eye from each display.

#### Implementation

To begin to calibrate the screen we must use a high quality spectro radiometer to measure the output from each screen for a range of colours and intensities. We do this with the display setup as if it is going to be used normally so that it is being configured for the real lighting conditions.

For each screen the spectro-radiometer is setup and all the other screens are set to display black, to account for the light they still emit when in a blank state which will be mixed in the beam-splitters. The screen then displays a

wide range of colours at different intensities which are measured to produce a map of what the spectro radiometer expected to see and what it actually received. From this map we can construct the calibration parameters needed to ensure that each display for each colour input will output a matching colour out.

The calibration from each screen produces 'black level', 'gamma' value and matrix to map from the CIE1931 XYZ colour space to RGB for each display [23].

We have chosen to use XYZ colour space as our scene rendering colour space as it is a good approximation of the human eye colour reception and has a larger supported colour gamut than RGB. It is also a good fit for our 32-bit per pixel linear render target textures which we are using to gain more accuracy during the transformation into RGB colour space than the standard 8 bit per pixel render targets would allow.

Once the scene is rendered in XYZ colour space it is then transformed with the equation showing in figure ref into the calibrated RGB colour space which is sent to the screens.

$$col_{rgb} = (screen_{XYZtoRGB} * (col_{xyz} - screen_{black}))^{(1/screen_{gamma})} \tag{5.9}$$

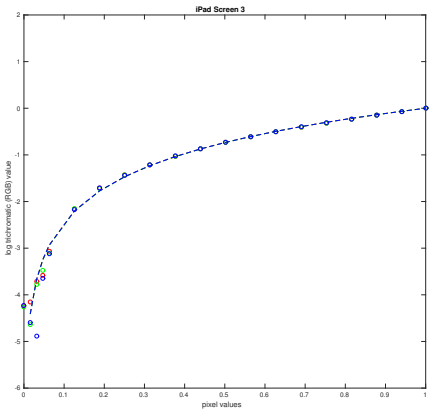
Add pictures showing calibration error

#### Calibration Results

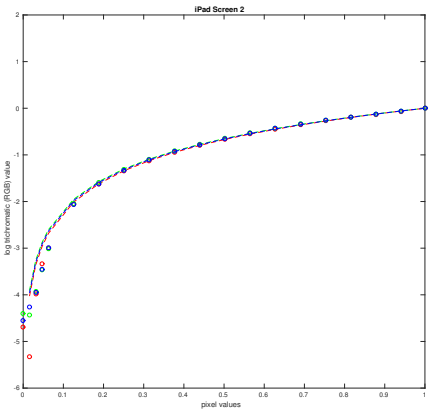
Explain different screen colour configurations. Sort latex to not throw the figures everywhere...

### 5.2.5 Full pipeline

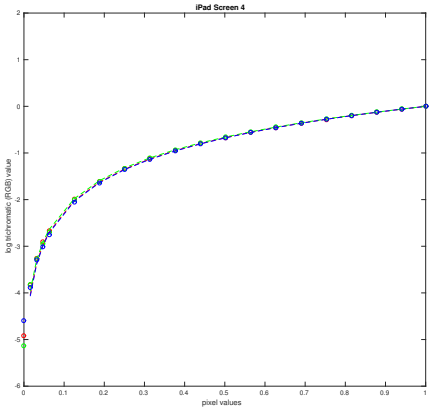
tidyup.... Once the screen is fully calibrated and the environment has been



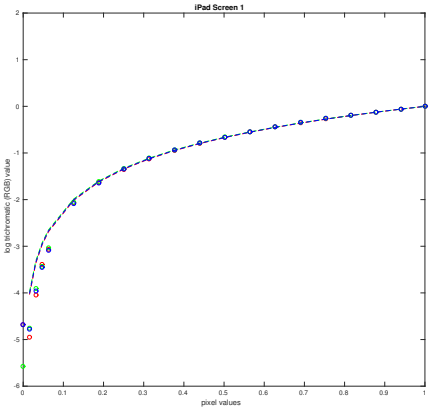

(a) Far Plane Left Display Panel.




(b) Far Plane Right Display Panel.




(c) Near Plane Left Display Panel.




(d) Near Plane Right Display Panel.

Figure 5.10

| Screen ID | Black Level Adjustment(CIE31) |
|---|---|
| Screen 1 | [0.5074, 0.4216, 1.1063] |
| Screen 2 | [0.5045, 0.4233, 1.0869] |
| Screen 3 | [0.4420, 0.4321, 0.8115] |
| Screen 4 | [0.4373, 0.4325, 0.8009] |

Table 5.2: Black levels - Representing the light leaking through the LCD panels when attempting to display a completely black screen.

| Screen ID | RGB Gamma value |
|---|---|
| Screen 1 | [2.2368, 2.2064, 2.2247] |
| Screen 2 | [2.2319, 2.1663, 2.1936] |
| Screen 3 | [2.4438, 2.4475, 2.4438] |
| Screen 4 | [2.2406, 2.1956, 2.2554] |

Table 5.3: Gamma correction values for each display panel.

setup, the process of rendering scenes for the multifocal view is relatively straight forward if you are used to rendering regular 3D scenes.

First you load all of the geometry and configure the position of the cameras and lights in the scenes. The cameras should be setup to match the projection that you would like to use and a blend mode should be chosen and implemented in the shaders.

The geometry is then loaded into each OpenGL context, and the cameras and projection matrices for the correct views are applied to each from the appropriate camera and projection are computed.

When it comes time to render the scene for each view we set the fragment shaders to render to two linear 32-bit render targets (Frame buffer objects) each. One represents the near plane and the other the far plane. We then apply the pixel shader which we have implemented some lighting and the blend functions as described in Section.5.1.2.

When the geometry is submitted is then correctly positioned in the scene and lit using the shader. At the end of the shader the blending is performed on its real world depth and then based on the depth and the blend mode selected the output is then applied to both render targets. For example, if

$$\begin{bmatrix} 0.0324 & -0.0095 & 0.0006 \\ -0.0147 & 0.0180 & -0.0017 \\ -0.0051 & 0.0006 & 0.0081 \end{bmatrix}$$

(a) Screen 3

$$\begin{bmatrix} 0.0345 & -0.0078 & 0.0003 \\ -0.0159 & 0.0149 & -0.0011 \\ -0.0058 & 0.0007 & 0.0062 \end{bmatrix}$$

(b) Screen 2

$$\begin{bmatrix} 0.0115 & -0.0038 & 0.0004 \\ -0.0050 & 0.0070 & -0.0009 \\ -0.0018 & 0.0002 & 0.0034 \end{bmatrix}$$

(c) Screen 4

$$\begin{bmatrix} 0.0118 & -0.0031 & 0.0001 \\ -0.0055 & 0.0059 & -0.0005 \\ -0.0020 & 0.0002 & 0.0028 \end{bmatrix} \tag{5.10}$$

(d) Screen 1

Figure 5.11: CIE XYZ to RGB colour matrices for each display panel.

a triangle is submitted which sits in front of the near plane and we have selected "projective' blending then 100% of the value of each pixel in front of near plane will be applied to the near plane render target and black will be written to the far plane. This effectively splits the scene between the two planes as shown in Figure.5.12.

After all of the rendering is complete the resulting render targets are then written to a plane on each screen in their configurable position.

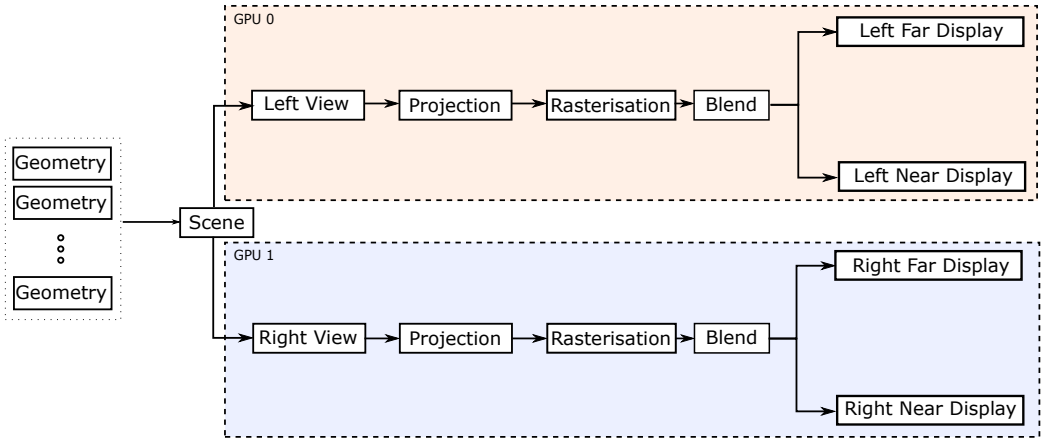



Figure 5.12: A high level overview of the steps to separate work between GPU's and display panels.

# Chapter 6

# Methodology and Testing

In experimenting with the device we are aiming to find out to what degree, if any, the use of multi-focal viewing planes increases the users perception of depth in a scene when compared to stereo and standard viewing conditions. Our hypothesis is that by alleviating the vergence-accommodation conflict we should be introducing a more comfortable and realistic depth viewing experience which should allow the user to more accurately judge depth than when the focal cue is missing. However, we are open to the idea that the change in focal depths as the user looks around a scene could introduce discomfort from other means which are less researched as it is not what users are typically accustomed to when using stereoscopic 3D devices.

The display and software design has given us lots of options for testing through varying projection, blending and real focal distances. However, to keep the experiment simple and within scope we have performed some early testing to narrow the the total number of repeat experiments. As such we will be investigating using the "Oblique" projection method as it provided correct near and far plane correspondence and with the projective blending gave us the least error from misalignment in the scenes.

Misalignment is a large concern for us during experimentation. We have seen that misalignment can completely ruin the effect we are trying to achieve and

can lead to a lot of discomfort the user.

In order to address our hypothesis' we will be testing the users perception of depth by showing two objects at different depths to the user and asking them to select which object appears closer. This will be performed with the user viewing objects with zero disparity, stereoscopically and stereoscopically with mutltiple depth planes. This will allow us to compare how accurately the user was able to determine the relative depth when given different depth cues and therefore determine if the additional focal depth cue allowed for a higher perception of depth.

## 6.1 Method

This is an experiment which will be measuring the perception of depth in humans so it is necessary for this to be tested on a range of volunteers rather than just ourselves to ensure that any conclusions are conclusions for a more general case of viewers and not just our own. The participants selected for this experiment will not be selected for any visual strengths or weaknesses such as stereoscopic perception strengths as we want the participants to map as closely as possible the general population. This experiment is going to be ran on a relatively small number of people than would be ideal so we will be taking that into account during our conclusions.

We have setup the display to have the distance between the far and the near plane approximately 0.6 dioptres apart, as this is seen to be the optimal distance for retinal image quality [15] and is the a large enough distance to for optimal stimulation of the accommodation response [16]. For this experiment the near plane will be at 0.54m and the far plane set to 0.81cm.

### 6.1.1 Depth Comparison Setup

The depth comparison scene is a small game in which the user is asked to look at two objects in the display and then report which of the two objects is

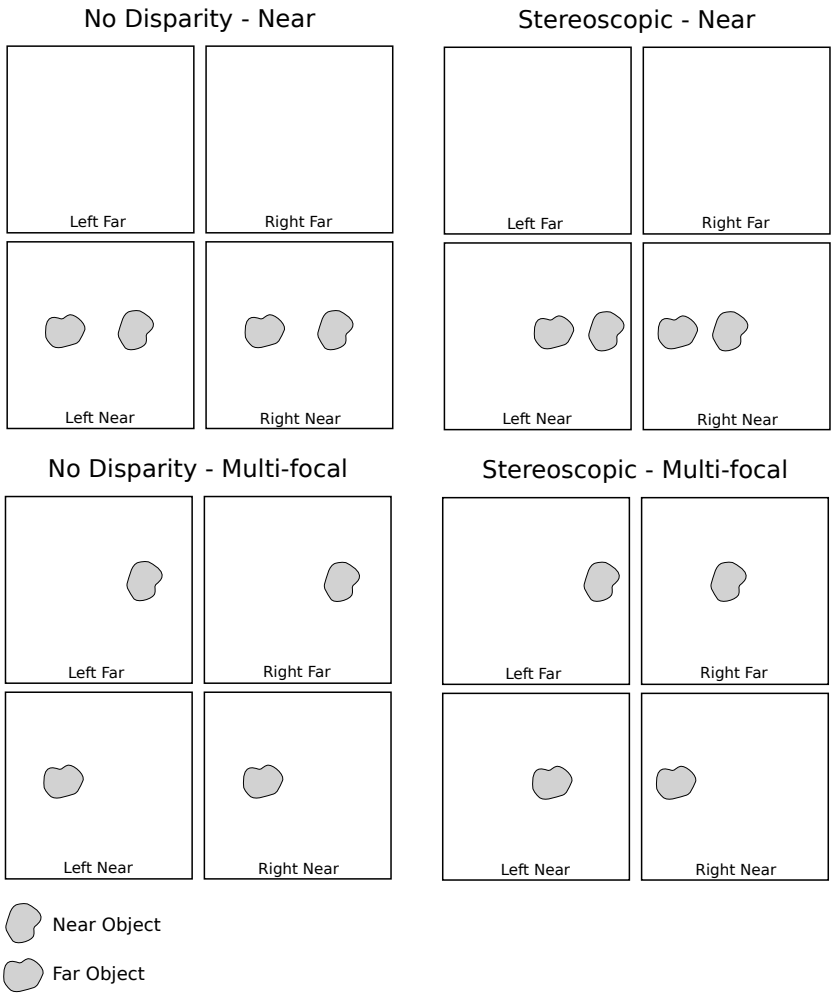
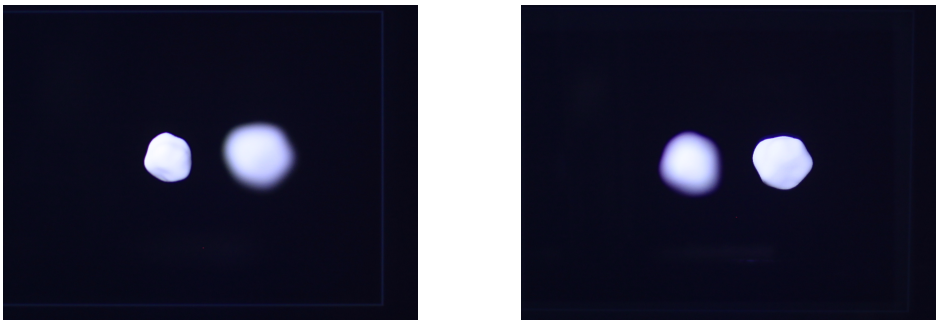



Figure 6.1: Example of a setup explain...

| | |
|---|---|
| Near Focal Distance | 54cm — 1.85 Dioptres |
| Far Focal Distance | 81cm — 1.23 Dioptres |
| Near-Far Range | 27cm — 0.62 Dioptres |
| Interocular Distance | 6.5cm |

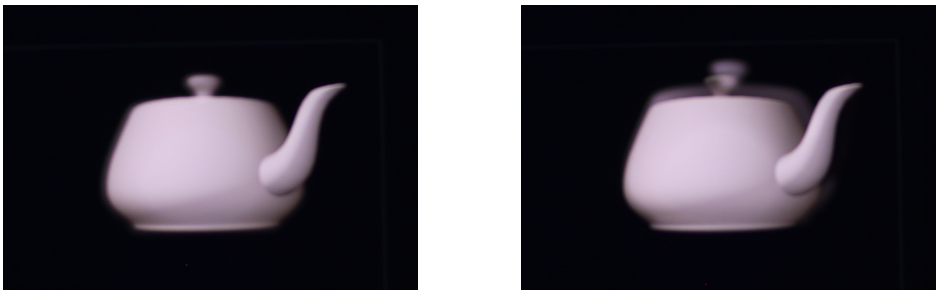
Table 6.1: This is the distance configurations the display has been setup to use in this experiment. It maintains the optimal 0.6 Dioptres between focal planes so that the difference in focal depth should be very perceptible to the user.

closer. This is then repeated for different distances between the two objects. The aim of this scene is to accurately measure the effects of the different



(a) Near Plane Object Focus. (b) Far plane object Focus.

Figure 6.2: These photographs are from a run of the experiment showing a large offset between the two object and displaying the images in the multi-focal render mode. As you can see by changing the focal length of the camera from 0.54m to 0.81m we are able to selectively focus on both objects and there is a large visible difference when that changes. The images were taken with an apperture of get app from file.



(a) Correctly calibrated. (b) Incorrect calibration.

Figure 6.3: This image shows the error from incorrect calibration. It shows the far plane out of alignment with the near plane and the 'shadowing' artefact becomes apparent. When this happens it is very clear to a user that the image is comprised of two separate planes.

depth cues to be able to determine if the addition of focal depth improves user perception of depth. To do this we need to be able to isolate the user from all other cues except the ones we are investigating, namely the presence binocular disparity, vergence, accommodation or none.

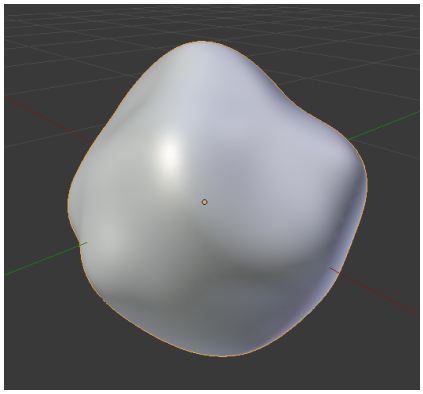


Figure 6.4: This is an example of the deformed sphere models that are being used in the experiment.

In a regular scene the user is given many hints at the depths of objects, see Section.2.2. In our scene we have taken many measures to achieve this:

- **Perspective**: To eliminate perspective cues we have scaled the objects so the size is invariant to depth in the scene. The objects being displayed also have no parallel lines for comparison of convergence. Additionally we have introduced a random scale of up to 5% of the objects size

- **Known Size**: We are using randomly extruded spheres so they are not a recognisable object.

- **Detail**: The objects have little no texture detail. The only cues from detail come from a slight pattern of shading from the shape of the object but as the two objects have different shape this is not comparable as a cue.

- **Occlusion**: The objects sit either side of the optical axis and do not occlude each other.

- **Lighting**: We have used fixed directional lighting so the lighting is invariant to position in the scene. This prevent the distance being judged from the rate of change of lighting on the surface.

- **Relative motion**: There is no motion in the scene to allow the user to judge depth from rate of change of size and position.

We have also had to consider hints at position that can come from the display setup:

- **Colour**: As we are using multiple mirrors and beam-splitters in the display there is a small amount of chromatic aberration which was quite obvious when using separate colours for each object in the scene. This gave hints to the number of mirrors being used to display that object which in turn told the user whether it was on the near or far plane when testing the multi-focal part of the experiment. As a result of this we had to switch to using a single colour per object in the scene.

- **Head Position**: If the user is misaligned when they change scene they will be able to notice outlines of any objects partially drawn on the near plane overlaid with the percentage of the object on the back plane. This will show the user which object is physically closer. To avoid this have added a calibration step between each display of the scene at different depths and advised the participants to try and avoid head motion while viewing the scene.

With these changes we are left with two randomly extruded spheres which will allow us to isolate and measure the cues we are interested in by varying which render mode we use.

From our early pilot tests we found that in stereo and multi-focal viewing modes we were able to very accurately determine the relative depths of objects when the distance between the objects were large. To affirm this and determine at what distance this breaks down we have selected to keep at least one object on the far plane and then move the other object away from the far plane at fixed offsets up to the near plane. With more samples be-

| **RENDER MODES** | Near Plane Only | Both Planes |
|---|---|---|
| 6.5cm Disparity | Stereo Cues | Stereo + Multi-focal cues |
| No Disparity | No depth cues | Multi-focal depth cues |

Table 6.2: This table shows us the four possible combinations of the render modes we will be testing and which cues they allow. We are particularly interested in the combination of 'Near Plane' and 'No Disparity' as this combinations will show us how well we have isolated other depth cues. If that combination is still able to be accurately judged for its depth then there is other depth cues available.

ing taken with smaller offsets, See Figure. 6.3, on and near the far plane to help determine the largest range that works within the multi-focal setup and determine where it breaks down.

In our chosen offset distance we have included a zero offset. So the two objects are at the same distance from the camera. In this situation the computer chooses a random object as the correct answer. This has been included in the test as a check for our data collection. The correct answers on this offset should trend towards 50%± error for the number of samples, see Figure.6.6. If this is significantly different in our data it will indicate a problem.

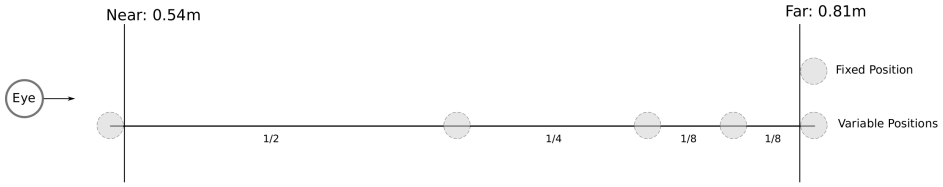



Figure 6.5: This diagram different offset locations and the relative distances between them. In this diagram it shows the right object at different locations but in the experiment either blob can be in the offset position but one is always on the far plane.

In order to get accurate and unbiased results we have made sure that the selection of which sphere is closer is randomly selected.

In experiments such as these the participant will become accustomed to what they are seeing and are expected to improve as it goes on. Because of this we will be randomising the order of offsets which the user is shown so that

| Index | Offset (cm) | Depth (cm) |
|---|---|---|
| 0 | 0 | 86.25 |
| 1 | 4.03125 | 76.97 |
| 2 | 8.0625 | 72.94 |
| 3 | 16.125 | 64.88 |
| 4 | 32.25 | 48.75 |

Table 6.3: This table shows the offset positions in distance from the camera. The depth is central position of the object. On the far plane at zero offset the value is not 81cm as we have adjusted for the size of the object and random scale to ensure that is entirely behind the far plane in its starting position.

either the small or large offsets, or certain view modes are not unevenly improved. This is random for each user performing the experiment so that we are accounting for any pattern that may arise in a single randomisation.

For the experiment we are asking the participants to do 20 samples per render mode comparison. Ideally we would have wanted to perform up to 40 comparisons per offset per rendering mode but time is a limiting factor for our participants. With 20 samples we are able to reduce the margin of error from random mistakes to be low enough to extract the information we need but for further study on any findings we would want to want to be more thorough.

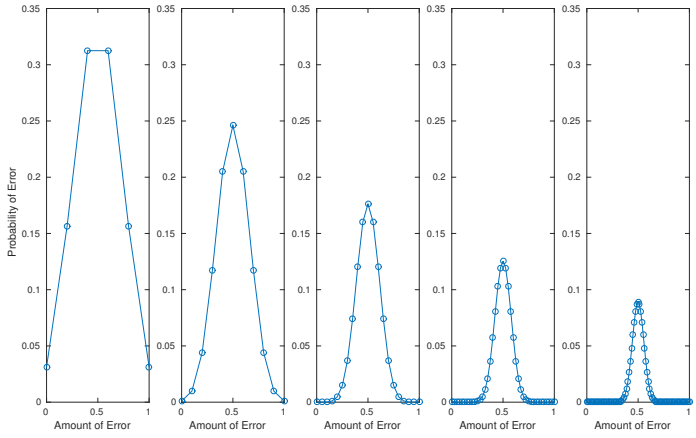
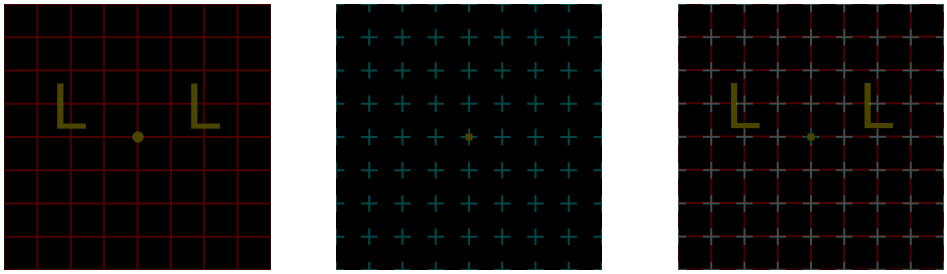



Figure 6.6: These graphs show the binomial graph of expected error.



(a) Near Plane (b) Far Plane (c) Combination.

Figure 6.7: Software Geometric Calibration: The near plane calibration screen shows a central point and two *'L'* characters. The point helps in centering the image at the correct position and the *'L'* is a simple way to confirm the orientation of the screens are correct. The far plane has a series of cross marks to align with the near plane as well as central point to align.

From our pilot studies with this test we found 20 samples for all render modes and offsets took around 20 minutes. As we are used to using the display we are assuming participants will take between 20-30 minutes including calibration steps.

Calibration can take a short amount of time as it is not automated. The display is setup in its default calibration assuming the mean interocular distance with a convergence point on the near plane. If a users interocular distance varies from this then we will need to spend a few minutes using the software's calibration screens (Figure. 6.7) to adjust the positions of the displayed visual planes. The software supports scaling and translation of the different planes to accommodate for participants eyes which may differ from the average in vertical alignment as well as horizontal interocular distance. This also allows us to ensure that the display is calibrated for a head position which is comfortable for the user.

### Experiment Procedure

The process of performing this experiment with a participant is relatively straight forward. The user will first be given a guide explaining what the experiment is and what is required. See Appendix.9.1 for the user guide.

Then the participant will be shown the display with the calibration screen and asked to get into a comfortable position. The screens are then aligned so that the user can comfortably see the near and far plane and converge the images into a correct stereo pair.

The participant will then be shown an example of the scene and shown the controls for how to select between the two blobs.

Once the participant is comfortable at the display with the controls the experiment proper can start.

The participant will be shown the scene, make a decision on which they believe is closer and then they will be shown the calibration screen to ensure the calibration didnt change during the scene. This is repeated twenty times for each combination of render mode (Table. 6.2) and each offset distance in a random order.

During the experiment the participant is advised to try and avoid head motion during each scene.

After the participant has completed all the required decisions an image will inform show them that the test is over.

### Data Gathering

From our experiment we will be storing an ID to represent each participant and a list of whether the participant was correct or incorrect for each sample of each offset and render configuration.

We will then analyse each participants results individually to determine any obvious patterns in multi-focal depth perception for that user and compare it to other users and the average results of all participants.

We are looking at each individually first as we expect there may be some quite interesting results that may be unique to each participant. In our early testing we found some quite large discrepancies in the perception of different depth cues which we hope to elucidate. We haven't filtered any of

our participants for ability to see stereo or judge depth from focal cues so this is something we predict will occur.

Once we have analysed the data individually we are also going to look for any trends in the whole data, with consideration to anything we may find in the individual analysis.

# Chapter 7

# Results and Evaluation

The experiment was ran with x participants. With most experiments taking between 20 and 25 minutes, as we expected. The results have fallen roughly in-line with what expected from our hypothesis with a few notable exceptions. They also have noise from error which is consistent in most entries with what we predicted. There are a few discrepancies which fell outside of this expected error which are explained in Section 7.1.

### 7.0.1 Problems during the Experiment

There was initially a few issues when we ran the test with the setup that we had not considered in our initial setup.

The first minor problem was that we had not taken into consideration the height of the user. The desk was too tall for some of our participants and it is not height configurable. This meant that we had to repeat an experiment after it was noticed and the participant mentioned it was difficult to keep the correct head position due to stretching to reach correctly use the mounted goggled.

The next minor issue was caused because we had not taken into consideration that people have varying sensitivity to bright images. The combination of the

brightness of combined images, particularly the calibration and end screens, which were not averaging the brightness to be the same as one display caused some participants discomfort and had to have the brightness reduced.

For some participants the test was a long period of time to maintain head position. That led to more breaks and slight head motion during the test than was expected and could have impacted the results of the test particularly the later comparisons which could impact the amount of noise in the data.

### 7.0.2 Individual Results

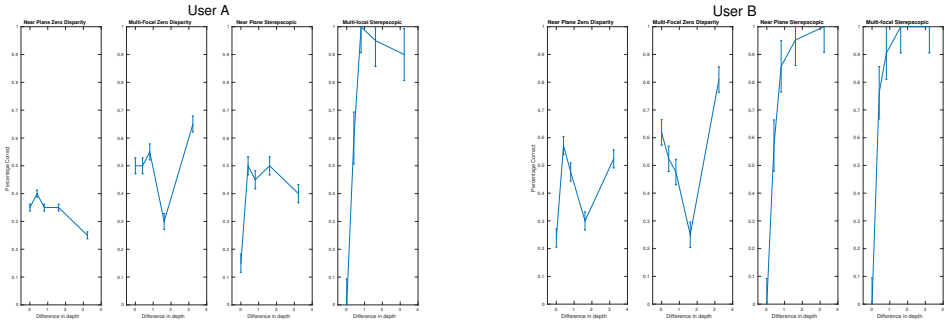
This section contains the raw results of each individual who took part in the experiment.

For each participant we have graphed the four different render mode combinations along with the results of what ratio of answers were correct for each offset position. For each result we have included the 'Standard error of the mean' as an indication of expected error and the participants names have been anatomised to letters.

### 7.0.3 Combined Results

As well as looking at the individual performance of each participant we have also found it useful to group the data together so that we can identify any overall trends.

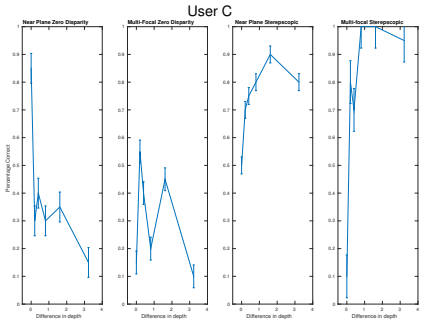
There was a very noticeable grouping in our individual results of those who were not able to discern the depth of objects using only stereoscopic depth cues. As such for further analysis we have separated those into two groups and plotted the results separately (Figures. 7.4a and 7.4b).



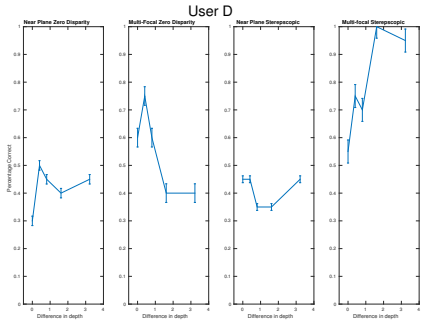

(a) Participant A



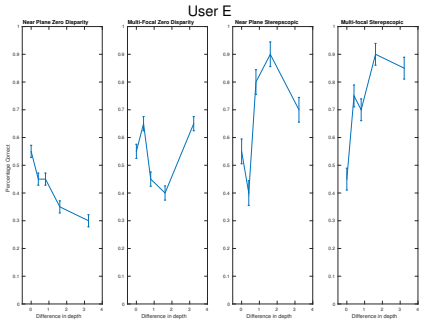

(b) Participant B



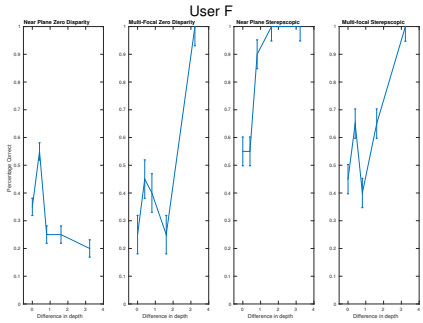

(c) Participant C




(d) Participant D

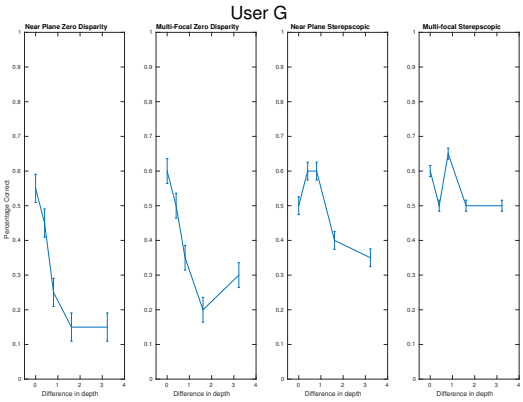



(e) Participant E




(f) Participant F

Figure 7.1: (Part 1 of 2) Individual results from participants in the depth comparison experiment. The graphs show the ratio of correct answers at each offset. The error range on each point is the 'Standard error of the mean'.




(a) Participant G

Figure 7.2: (Part 2 of 2) Individual results from participants in the depth comparison experiment. The graphs show the percentage of corrects answers at each offset. The error range on each point is the 'Standard error of the mean'.

## 7.1 Evaluation

### 7.1.1 General Overview

Overall we are very pleased to see that the results show that differences in depth can be accurately percieved from multi-focal depth cues where they may possibly may not be with just binocular disparity (Figure.**??**).

However, the results for multi-focal alone are a bit disappointing and we expected a stronger result in favour of correct depth selection especially at the higher offsets where each object would appear entirely on a separate plane, removing all possible error from the projective blending. The results of this are particularly curious as when setting up the experiment with participants in our calibration steps we asked the participants to focus between the two planes on two displayed objects and they all were able to do this. We will discuss this further in Subsection. 7.1.3.

For the same reasons as this we were surprised by the lack of stereoscopic depth perception from some participants as they also confirmed that they were able to see the objects in the scene correct in the stereographic rendering

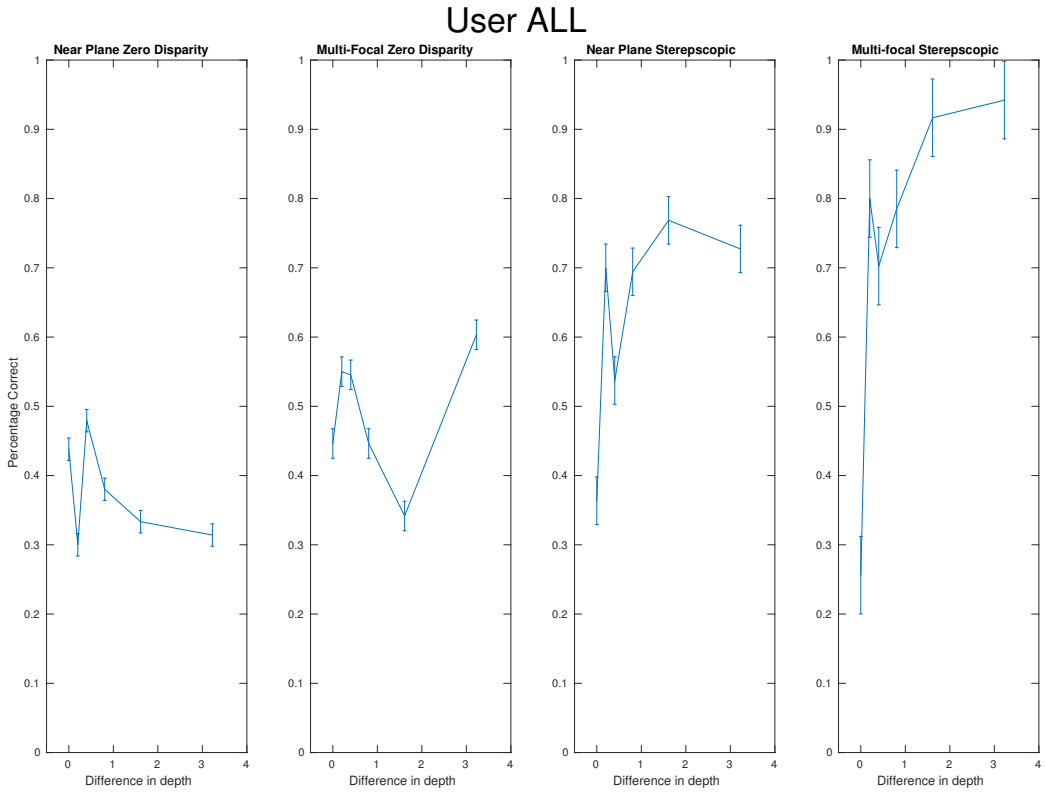
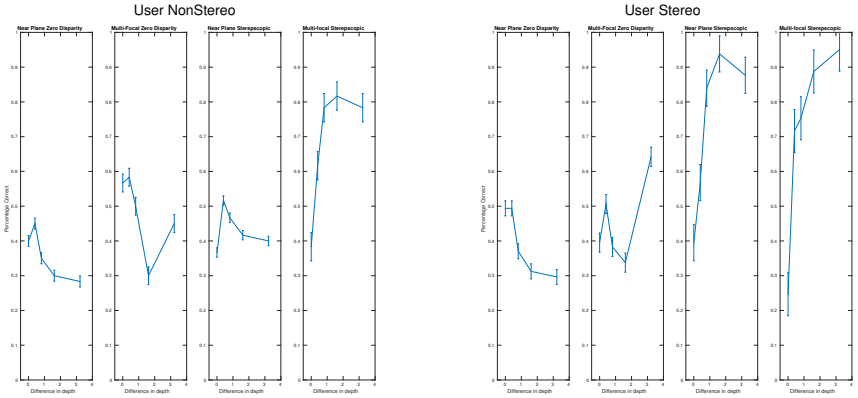



Figure 7.3: This graphs shows the combined results of all users for all rendering modes and offsets.




(a) Participants unable to see stereo depth.

(b) Participants able to see stereo depth.

Figure 7.4: These graphs show the separate results of participants who were and who were not able to percieve depth from only stereographic cues.

modes as part of the calibration. This is discussed further in Subsection.7.1.2.

### 7.1.2 Stereoscopic Cues

As mentioned in Subsection.7.1.1 we were surprised by the lack of some participants to be able to select the depths of different objects when given only stereo depth cues as users had reported being able to view the objects stereoscopically.

What makes this more interesting is that those participants without strong stereo depth scored consistently high on the stereoscopic multi-focal scene but not much different than random input on the multi-focal only scene.

This points to the possibility that there is some relationship between the two cues and that perhaps when the stereo vision depth perception is poor but present it is greatly helped bu the presence of a depth cue from accommodation, or the vergence cues which will be missing in the multi-focal only scene play a part in how the visual cortex processes the signals from accommodation and without those cues the accommodation alone may be hard for the participant to comprehend.

Another factor to consider at this stage is that some people have experience with stereographic systems from going to see 3D films or using VR devices and regular binocular vision but others may not. If the participant does not have good stereopsis in a the real world, such as if they have stereo-blindness caused by amblyopia, then unless they are trained to work with stereo they will not be able to correctly process the depth cues as well as someone who has experience processing those cues [14].

There is even a significant variance in the percentage of correct answers at the maximum offset for those who were able to correctly recognise the binocular disparity cues. At the highest offset, where the difference in depth should be the most obvious, there are some users who were able to recognise the depth difference correctly but not as frequently as others. It would have been better if we had measured the interocular distance of these users as a difference in

interocular distance than the 6.5cm that was used in this experiment has been shown to have the effect increasing or decreasing the sense of depth from an object depending on whether the mismatch is greater or less than the real interocular distance. In the situation where the users interocular distance is larger than 6.5cm it would have the effect of making the distance between the two objects appear smaller which could be a cause of the different rates of fall-off in accuracy.

### 7.1.3 Multifocal Depth

In the scene showing just multi-focal depth cues we had expected that those which saw a benefit in the from the cues in stereo rendering would similarly see a strong response when the cue was isolated. However, this seems to have not been the case for all participants.

A few participants were able to get a strong sense of depth with just multi-focal cues for the very large offsets, with one participant able to correctly identify all twenty examples correctly (Figure.7.1f) but that was not the case for most. Somewhat more interesting is that a few users have consistently answered incorrectly (Figures.7.1c,7.2a). This suggests that in some cases some depth was detected by the user but the distant objects appeared to be closer consistently. Further study would have to be taken to get a more conclusive conclusion about what caused that, but a hypothesis is that without extra visual cues the participant could only detect a change in focal depth but not which focal length represented points closest to them. This could tie in with our thoughts on the improvement for stereoscopic rendering, as the participants which were able to correctly identify depths from binocular disparity were able to do so more effectively with focal cues. It could be that the participants are not trained to identify depth through accommodation alone and although it is present they cannot ascribe it to a depth accurately without assistance from other cues. Accommodation cues by being less prominent in some people, could simply be used to amplify other depth cues.

However, this may just be the result of our use of a relatively small distance

between planes, 0.6 Dioptres. This distance was selected due to research showing it allowed for the smoothest transition of accommodation in the human eyes. In some of our results we can see a relatively strong correlation in correct answers for the largest offset. If we were to use a much larger distance we could measure whether it produced a stronger response than we currently are, in that situation it would be interesting to see if we still saw the inversion of results or the low perception of depth we are seeing in some of our current results.

Another important consideration when discussing the middle offsets is that this is where the blending of the near and far plane will have the largest effect. In this area the projective blend is selecting to split the brightness of the screen across both screens dependent on the depth of the point in the scene and distance between the near and far planes. When the object is at the highest offset it is entirely on the near plane and being compared to an object entirely on the far plane. This scenario is the closest to the real world and should give the strongest accommodation response, so it is somewhat understandable that this should get the strongest depth response. Error within the middle offsets could be in part due to the blending. The blending methods being applied in this experiment assume the users eye as a point camera, where as in fact the eye samples across a different range based on the pupil diametercite pupil response paper. A more complex implementation such as one which uses ray tracing with ray bundles would be able to approximate the spread of pixels hitting the eye based on pupil diameter and this would reduce errors in the blend. Without having to do any major changes to our setup we could test the effect of this by varying the brightness of the displayed objects. If the objects are brighter the pupil will contract which will reduce the spread of light rays by reducing the angles light can enter the eye.

Our blending is also highly affected by the calibration of the geometric alignment. We have done our best to ensure this is correct at the start of experiments and between each comparison however the user is not in a fixed position and particular when the participant wears glasses it is very easy for

this to become disrupted. When the blending is not correct it is much more difficult for the participant to see the image correctly and will prevent the user from focusing correctly on the objects in the scene. If this has happened then we could expect a lot more noise in our results, which appears to be the case for some offsets when only using multi-focal depth cues.

This could be a result of the relatively low depth in the depth. Maybe placing the planes at the distances for the smoothest lens accommodation response wasnt the best.

Could have been affected by geometric calibration. Can't get a perfect blend because eyes are not point cameras. Would need approximation. Try under different screen brightnesses to effect pupil diameter to measure how this effects the focus at the blending boundaries which is where the results drop off in the curve so could effect that. Eyes dont wont on single point ray. Ray clusters may be a solution if we wanted to avoid rasterisation.

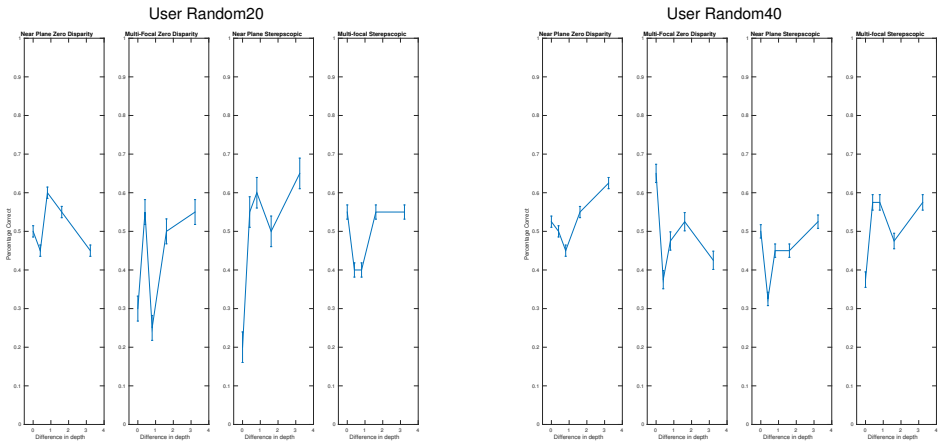
### 7.1.4 Error Margins

Slightly higher error than expected for the zero offset positions and no depth cues render mode. There may be slight depth cues, or things people see as depth cues in the spheres. maybe slight error in randomisation?

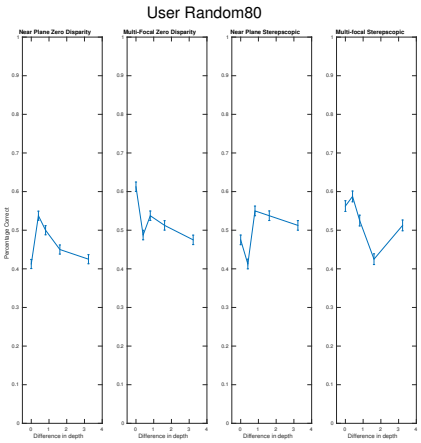
Margins of error make this more difficult to parse accurately than hoped. A further study with more samples would give better results. Also repetitions on different times or day may help as accommodation could be compromised by tiredness, ref tiredness effects on accommodation.

Show purely random input responses.

Error could be from misalignment. Perhaps a more thorough method of maintaining alignment is needed?




(a) 20 samples (400 data points overall) (b) 40 samples (800 data points overall)




(c) 80 samples (800 data points overall)

Figure 7.5: Random choice for different samples sizes per offset and render-mode. The random noise is consistent with what expect for this distribution (Figure. 6.6).

### 7.1.5 General Trend

Generally it looks like depth is percieved only to within 10cm at a distance of 80cm. Ref other papers looking into this. In some people this can be enhanced slightly by multifocal.

What new information does this show? SHOWS THE PRESENCE OF THE FOCAL DEPTH BEING HELPFUL

By what margin is it better or worse? NOT A LOT IN STEREO

What does this point to for only multi-focal viewing? Considering people are trained for stereo. STEREO ENCHANCES.

Talk about alignment problems. Point camera not pupil distance means that edges are not approximated correctly. Ref mark banks

# Chapter 8

# Summary and Conclusions

Mention if it reduced the vergence problem. Mention if it appeared more real. Which blend methods were convincing? Was alignment consistent and believable?

potential future work.

Test different distances.

Test comfort.

Test depth feeling.

Longer user testing for middle blend distances.

Auto-calibration.

# Chapter 9

# Appendix

## 9.1 User Guide

**BRIEFING FORM**

Experiment: Depth Perception with a Multi-focal Stereo Display Date: 31 May 2016

Thank you for taking part in the experiment. The experiment will take approximately 20 minutes and will be comprised of one 20-25 minutes section. Please read the following instruction carefully before starting the experiment.

The purpose of this experiment is to compare the effect on percieved depth when given extra depth cues. The results of this will help in the design of future displays.

You will be shown two objects at different depths. The size, shape and lighting on these objects is no idication of the depth of the object. You will have to select which of the objects is nearest using the '¡' and '¿' keys to select the left or right object. After each selection you will be shown a calibration screen to ensure the screens remain calibrated correctly for your eyes, press '/' when this is correct to move onto the next comparison. This will be performed for 400 times.

In this experiment you are asked to avoid making any decision which based on misalignment of the screens and are encouraged to take a break between comparisons if you feel any discomfort.

[END]

# Bibliography


[1] Alex Paul Pentland. "A new sense for depth of field". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 4 (1987), pp. 523–531.

[2] Claire C Gordon et al. *1988 Anthropometric Survey of US Army Personnel-Methods and Summary Statistics. Final Report*. 1989.

[3] Mark Mon-Williams, John P Warm, and Simon Rushton. "Binocular vision in a virtual world: visual deficits following the wearing of a head-mounted display". In: *Ophthalmic and Physiological Optics* 13.4 (1993), pp. 387–391.

[4] Andrew J Woods, Tom Docherty, and Rolf Koch. "Image distortions in stereoscopic video systems". In: *IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology*. International Society for Optics and Photonics. 1993, pp. 36–48.

[5] John P Wann, Simon Rushton, and Mark Mon-Williams. "Natural problems for stereoscopic depth perception in virtual environments". In: *Vision research* 35.19 (1995), pp. 2731–2736.

[6] Tetsuri Inoue and Hitoshi Ohzu. "Accommodative responses to stereoscopic three-dimensional display". In: *Applied optics* 36.19 (1997), pp. 4509–4515.

[7] Kurt Akeley et al. "A stereo display prototype with multiple focal distances". In: *ACM transactions on graphics (TOG)*. Vol. 23. 3. ACM. 2004, pp. 804–813.

[8] Neil A Dodgson. "Variation and extrema of human interpupillary distance". In: *Electronic imaging 2004*. International Society for Optics and Photonics. 2004, pp. 36–46.

[9] Simon J Watt et al. "Focus cues affect perceived depth". In: *Journal of Vision* 5.10 (2005), pp. 7–7.

[10] Rafal Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. "A perceptual framework for contrast processing of high dynamic range im-

ages". In: *ACM Transactions on Applied Perception (TAP)* 3.3 (2006), pp. 286–308.

[11] David M Hoffman et al. "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue". In: *Journal of vision* 8.3 (2008), pp. 33–33.

[12] Sheng Liu, Dewen Cheng, and Hong Hua. "An optical see-through head mounted display with addressable focal planes". In: *Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium on*. IEEE. 2008, pp. 33–42.

[13] Gordon D Love et al. "High-speed switchable lens enables the development of a volumetric stereoscopic display". In: *Optics express* 17.18 (2009), pp. 15716–15725.

[14] Daphne Bavelier et al. "Removing brakes on adult brain plasticity: from molecular to behavioral interventions". In: *The Journal of neuroscience* 30.45 (2010), pp. 14964–14971.

[15] Sheng Liu and Hong Hua. "A systematic method for designing depth-fused multi-focal plane three-dimensional displays". In: *Optics express* 18.11 (2010), pp. 11562–11573.

[16] Kevin J MacKenzie, David M Hoffman, and Simon J Watt. "Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control". In: *Journal of Vision* 10.8 (2010), pp. 22–22.

[17] Takashi Shibata et al. "The zone of comfort: Predicting visual discomfort with stereo displays". In: *Journal of vision* 11.8 (2011), pp. 11–11.

[18] Yasuhiro Takaki, Kosuke Tanaka, and Junya Nakamura. "Super multi-view display with a lower resolution flat-panel display". In: *Optics express* 19.5 (2011), pp. 4129–4139.

[19] Parth Rajesh Desai et al. "A review paper on oculus rift-A virtual reality headset". In: *arXiv preprint arXiv:1408.1173* (2014).

[20] Cass Everitt. "Beyond Porting". In: *SteamDevDays*. NVidia. 2014.

[21] Bochao Li, Ruimin Zhang, and Scott Kuhl. "Minication affects action-based distance judgments in oculus rift HMDs". In: *Proceedings of the ACM Symposium on Applied Perception*. ACM. 2014, pp. 91–94.

[22] Nick Whiting. "Lessons from Integrating the Oculus Rift into Unreal Engine 4". In: *Oculus Connecy*. Epic Games. Oculus, 2014.

[23] Rafal K. Mantiuk. "Perceptual display calibration". In: *Displays: Fundamentals and Applications*. Ed. by Rolf R. Hainich and Oliver Bimber. 2nd. CRC Press, 2016.

[24] Michael Abrash. *Down the VR rabbit hole: Fixing judder*. http://blogs.valvesoftware.com/abrash/down-the-vr-rabbit-hole-fixing-judder/. Accessed: 2016-06-05.

[25] Paul Bourke. *Calculating Stereo Pairs*. http://paulbourke.net/stereographics/stereorender/. Accessed: 2016-06-05.

[26] Ben Lewis-Evans. *Designing to Minimize Simulation Sickness in VR Games*. http://www.gdcvault.com/play/1022772/Designing-to-Minimize-Simulation-Sickness. Accessed: 2016-06-05.

[27] OpenGL. *OpenGL - GLFrustum*. https://www.opengl.org/sdk/docs/man2/xhtml/glFrustum.xml. Accessed: 2016-06-05.