Hyper-Realistic Rendering On A Multi-Focal Plane Stereo Display

Nicholas G. Timmons
Downing College

UNIVERSITY OF CAMBRIDGE

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

May 23, 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 an investigation into the effects of the presence of multiple focal planes, and different interpolations techniques on countering the vergence-accommodation problem and how it impacts the perceived realism of the images being shown. This is of particular interest in the areas of AR (Augmented Reality) where the user is faced with two conflicting depths for images appearing at the same location.

The aim of this research is to contribute to the knowledge base for VR(Virtual Reality) and AR techniques so that in the future similar techniques and ideas can be used in development of hardware to improve the user experience of 3D displays.

Add some of the results of the work here

It's often useful to bring forward some "highlights" into this chapter (e.g. some particularly compelling results, or a particularly interesting finding).

This report includes a background of current techniques being used for VR/AR and a short overview of similar methods using multi-focal planes. There is then a breakdown of the requirements for the software and hardware with details of how they have been implemented to achieve our goals. Finally there is a more in depth analysis of our results and how they can be built upon in the future.

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) which are in use in many commercial products have seen significant improvement over the past decade but still have some large technical and usability problems which could restrict the wide-scale appeal of the products [REF: The zone of comfort...] and do not yet successfully mimic the visual characteristics of the real world so as to be perceived as ”real”.

When considering the comfort of the user the results of many psycho-physical and usability studies have suggested that use of the current solutions can lead to various problems including distortion of perceived depth [REF], visual fatigue [REF], diplopic vision (double vision) and degradation in the oculomotor response (as measured by slight movements within the eye). There are many factors contributing to cause these conditions varying from low quality images such as the high persistence in Oculus Rift, incorrect interocular distances or inability to allow the eyes to rest but a major cause that is often mentioned is the discrepancy between the accommodation and vergence when using 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 visual axis to intersect at the desired 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 diagram.

Work by [Watt el al] has suggested 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 AR environments where the virtual environment is mixed with the real world and depth cues from the display would have to be correct to maintain consistency with the image of the real world.

2.1 What is realism?

In the context of this research we are considering realism to consist of many factors. 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.

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 over a scene which is lacking such cues.

2.2 Standard Stereo VR Implementations

The current standard for VR is shared by a number of commercial hardware implementations. They consist of a head mounted display with the screen split 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 can be seen and to reduce the "Screen Door Effect" caused by the pixel density. The software is then implemented using the interocular distances of screen in the display with correct field of view and perspective projection.

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 between the two images and perceiving them as a single object at an expected distance.

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

This triggers the mentioned Vergence-Accommodation problem mentioned earlier from a conflict in visual cues. There are some examples of trying to use simulated Depth of Field blur in early VR 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. cite Unreal talk

Papers:

Minification Affects Action-Based Distance Judgments in Oculus Rift HMDs
Speccifications: A Review Paper on Oculus Rift & Project Morpheus

Reference system requirements from current state of the art (75fps etc.)

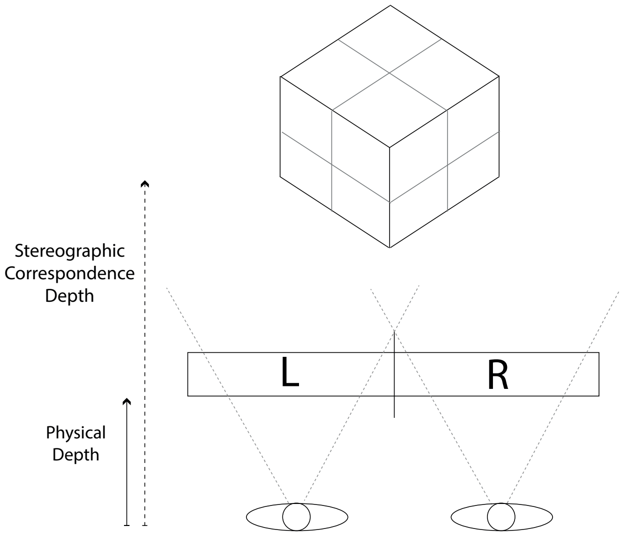

Figure 2.1: ADD DESCRIPTION

2.3 Vergence-Accommodation Problem Details

Papers:

-The zone of comfort: Predicting visual discomfort with stereo displays (2011) Immersive stereo displays, intuitive reasoning, and cognitive engineering (2009) Vergenceaccommodation conflicts hinder visual performance and cause visual fatigue(2008)

The Vergence-Accommodation problem in current single plane stereo implementations pose some issues when we are trying to simulate a real world using only a single focal distance.

While this conflict does have some effect on the comfort of the user REFD the zone of comfort paper, in the area we are looking at we are more interested in the effect this has on how the users perceives the world.

ADD DIAGRAM OF VERGENCE and ACCOMMODATION see Vergenceaccommodation conflicts hinder visual fig.1

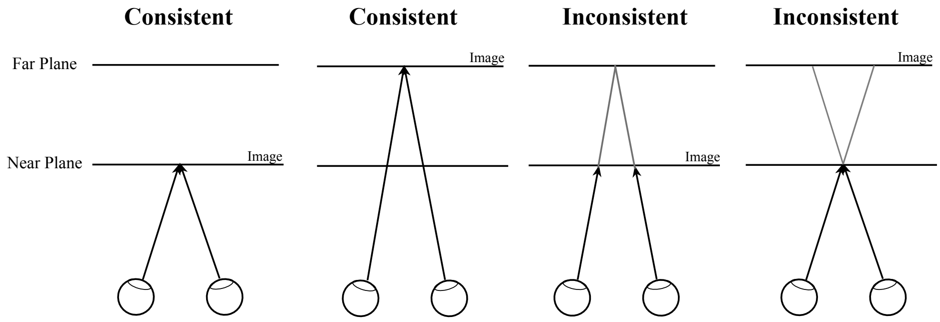

Figure 2.2: ADD DESCRIPTION

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 increased. These symptoms can all interfere with the stereoscopic method as it reduces the ability of the user to view 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 confusion in the viewed scene and allow it to be properly processed the same as any real-world scene.

This will need diagrams - borrow fancy camera and use images to help clarify

Chapter 3

Related Work

Discuss papers:

A Stereo Display Prototype with Multiple Focal Distances (2004) Super multi-view display with a lower resolution flat-panel display An Optical See-Through Head Mounted Display with Addressable Focal Planes A systematic method for designing depth-fused multi-focal plane three-dimensional displays

talk about pixel limitations with regard to requirements of modern setups as well as real distances.

The related work chapter should usually come either near the front or near the back of the dissertation. The advantage of the former is that you get to build the argument for why your work is important before presenting your solution(s) in later chapters; the advantage of the latter is that don't have to forward reference to your solution too much. The correct choice will depend on what you're writing up, and your own personal preference.

3.1 Multi-focal viewing

Multi-focal displays aim to simulate real world light ray angles from displays 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 and the problems the simple single focal displays they use cause.

The general idea is to use multiple displays, or sections of displays ref and a series of lenses, mirrors and beamsplitters 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 visible focal depth.

These screens were used to investigate the comfort and physical reactions of the human eye when viewing simple scenes ref 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 add diagram. This means that displays has to be invariant to motion and match the physical position of the users eyes.

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. They showed a mean distance as 6.47cm for men and 6.23cm for women, figure ref.

To prevent the movement of the user causing problems some very interesting solutions have been used in research ref such as gum-shields to bite down on and fixed eye slots. However, these are a little impractical for a study covering multiple people which might need to support multiple distances and interocular distances.

http://www.dtic.mil/dtic/tr/fulltext/u2/a209600.pdf

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

Figure 3.1: ADD DESCRIPTION

intraocular distances. Depth perception in the human eye / brain.

Chapter 4

Multi-focal Theory

Maybe move this into the related work section?

Show and describe diagram of splitting the scene.

Show a diagram of rays at infinity mixed with nearby rays

4.1 Oblique Projection? Reducing disparity

Chapter 5

Implementation

Speak about how we want to measure the down sides of the current solutions and how we can compare them to the improved view.

5.1 Display Requirements

To be able to resolve our hypothesis the screen would need to support:

- At least four screens - Configurable screen distances - Support for alignment

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

Chapter 6

Display Design

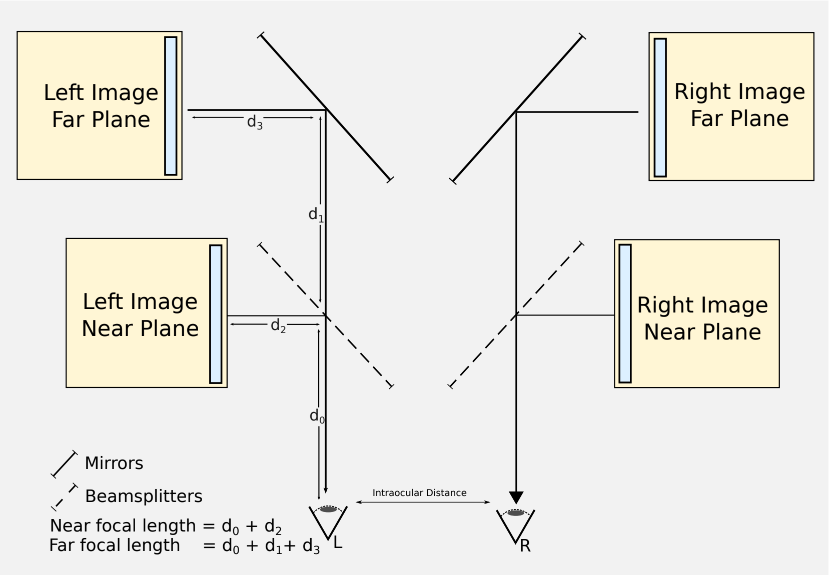

Figure 6.1: ADD DESCRIPTION

The display will be designed to allow for two real focal depths per eye in a configurable display for testing of the effects on perception of depth.

To achieve this we will be using four high resolution displays [ref ipad displays] to get maximum quality in rendering per view.

That gives us a near and far display per eye. The images on each display is seen through a 50/50 beam-splitter which merges the two images into one. See fig layout for a more detailed description of the physical layout.

As the image is merged through the beam-splitter it is essentially combining the 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 it 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 system 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 or error from light absorption from the mirrors or beam-splitter. Additionally this will also help correct for any external light or internal reflection we are not able to reduce.

6.1 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.

6.1.1 Display Distances

Add maths for calculating distance based on dioptres.

-Assumptions about parallel rays beyond far distance

Near Focal Distance	54cm — 1.0/0.54m Dioptres
Far Focal Distance	81cm — 1.0/0.81m Dioptres
Intraocular Distance	6.5cm ref
Convergence Distance	67.5cm

6.2 Known limitations

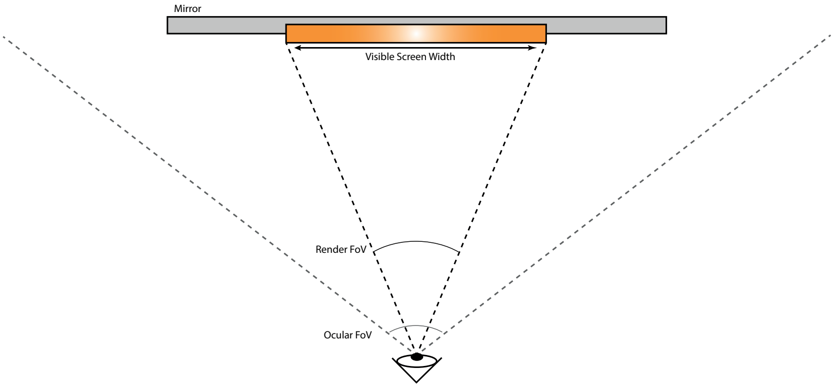

Figure 6.2: ADD DESCRIPTION

Limited eye coverage (low FOV). As our screens will only cover a limited FoV compared to a full coverage head mounted display we are limited in our ability to give as full an effect, rather it will look like peering through a window into another room. This may have an effect on the perceived realism.

Large display. The size of the rendered screen from the view point is limited to 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 view then the screen would have to be of the correct size to cover the full view at the given focal distance. In our display because we are using IPad displays we are limited to quite a small view of the scene due to the limited size. This could be overcome with magnification ? Limitations with multiple monitors at different distances compared to head mounted displays.

Lacks 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 can help a user feel a part of the scene. The fixed view point means we need to use motion of the scene rather than motion of the user. 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 correct fit the size of the distant screen which uses its full 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 lacking motion cues from ears [ref]. So in order to give the user cues of depth from different view we are relying on objects in the scene moving relative to the camera.

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. Not sure on this point. Testing Fri 20th May

plus others

Chapter 7

Software Design

This section will cover the considerations in the design of the software and the solutions to problems specific to multi-focal rendering.

7.1 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.

Testing: Depth blending methods - Linear, box, non-linear, clamped Model and texture support Camera motion

7.1.1 Perspective Modes

Talk about vertical parallax...

Talk about maintaining correspondence...

Toe-in - Explain problems + diagram

Off-Axis - Explain what it corrects + diagram

Oblique - Explain what it corrects and hypothesis + diagram

7.1.2 Depth Configurations

The scene is being modelled in metres 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. The corresponding points on each screen which do not pass this test will show black.

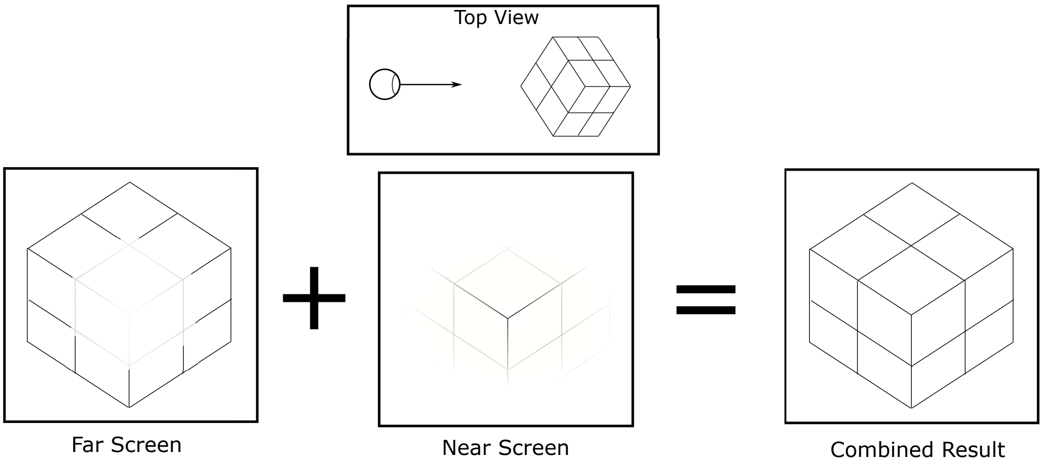

Figure 7.1: ADD DESCRIPTION

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 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.

The three 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.

$$n = \frac{n_i - n_1}{n_2 - n_1}$$

$$col_{out} = f(n) = \begin{cases} col_{near} & \text{if } n \text{ is } < 0.5 \\ col_{far} & \text{if } n \text{ is } > 0.5 \end{cases} \tag{7.1}$$

Note: We expect this to produce the effect of the scene feeling like it is made out of two pictures shown to the user.

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

$$\begin{aligned} n &= \frac{n_i - n_1}{n_2 - n_1} \\ col_{out} &= (n * col_{near}) + ((1 - n) * col_{far}) \end{aligned} \tag{7.2}$$

Non-linear: As the points move across the middle space they will be non-linearly interpolated using a modified sigmoid curve between the two distances as shown in fig.blend.

$$\begin{aligned} n &= \frac{n_i - n_1}{n_2 - n_1} \\ blend &= \frac{1}{1 + exp((-n * 2 + 1) * 6))} \\ col_{out} &= (blend * col_{near}) + ((1 - blend) * col_{far}) \end{aligned} \tag{7.3}$$

Inverse Depth: In this mode the calculation to determine the depth through the blending area is scaled inversely to match the depth divide in perspective transform.

$$\begin{aligned} n &= \frac{(1/n_i) - (1/n_1)}{(1/n_2) - (1/n_1)} \\ col_{out} &= (n * col_{near}) + ((1 - n) * col_{far}) \end{aligned} \tag{7.4}$$

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.

$$\begin{aligned} col_{out} &= col_{near} \\ &\text{or} \\ col_{out} &= col_{far} \end{aligned} \tag{7.5}$$

The aim of the blending is to produce a sum result of combined rays which would approximate the rays from the target distance.

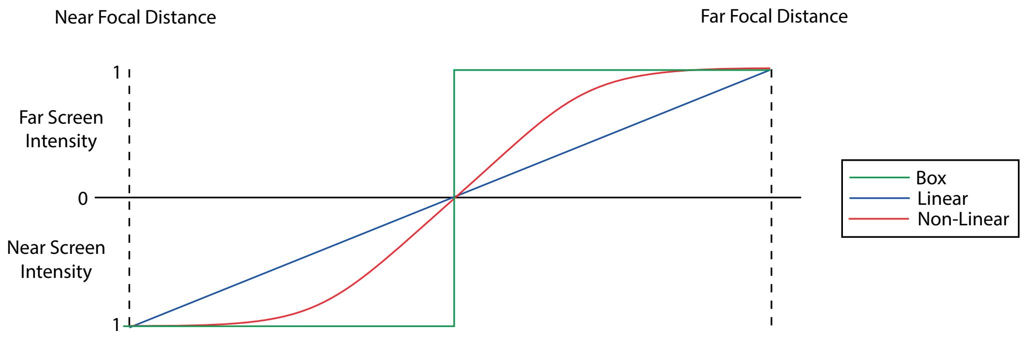

Figure 7.2: ADD DESCRIPTION

Add examples of this working - with camera on nice scene.

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 minification action-based distance judgements in oculus rift 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 FoV will have to be correct to maintain that depth cue accurately.

7.1.3 Reflection Depth

In our setup we are interested in providing correct depth cues through light rays reaching the eye with the appropriate angle for the distance to the object.

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 off the object without diffusion towards the eye. In this case the rays of light have a parallelism like that of the object at the distance to the reflected 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 solution for the surface we are mapping. 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

7.1.4 Rendering costs

for (all object) (all eyes) is better than (all eyes) (all objects) For rendering simple scenes with OpenGL a high proportion of the costs can be the switching of state if enough data is not being submitted.

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).

This way it is only a maximum of two state changes per object instead of twice the textures, meshes and data per object.

$$\begin{aligned} m &= \text{Object count} & c &= \text{View count}n = \text{State changes per object} \\ \text{Per view} &= c * m * n \\ \text{Per object} &= m * n + m * c \end{aligned} \tag{7.6}$$

Depth clip and shared depth

7.1.5 Simulated DoF

Explain which method we will implement to get the fake effect used in standard stereo setups.

7.2 Software Configuration

Which blend methods were used? How will it accommodate different people. What problems will it overcome (colour correction, etc).

7.2.1 Rotational Consistency

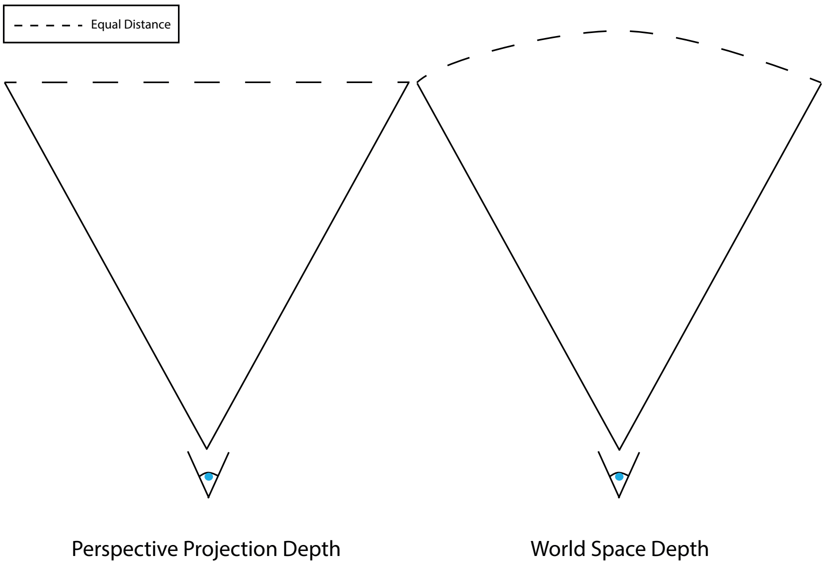

Figure 7.3: ADD DESCRIPTION

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 fig fig dist 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 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.

get screen shot of different depths!

Depth has to be distance from the eye position, not distance in view space. Show comparison of different depths.

7.2.2 X11 Windows controllers and shared contexts

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 maximise the effects of persistence of vision ref].

Only full-screen applications are completely decoupled from the desktop screen present 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 systems which does support a mode called "Zaphod" to allow the user to run multiple 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 as full screen which allows us full control of the presenting of rendered images up to the maximum supported refresh rate of the displays.

As we can select which GPU controls which displays it is possible to align the left side displays to share a context on one GPU and the right side to share a context on the other.

7.2.3 OpenGL MRT's

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 ref] this is not a limiting factor in performance.

In order to take advantage of the data sharing we do have available we are making use of multiple render targets in our shaders 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 recommended 75fps [ref] so that even if they are out of sync the distance between the two scenes should be low enough to be imperceivable by the human eye.

7.2.4 Colour Calibration

Floating point render targets after split apply separate colour correction from XYZ -¿ RGB with correct gamma and offset.

Explain XYZ, RGB and sRGB space

Explain why we render in linear lighting

Explain different screen configurations - Get graphs from Rafal

$$col_{rgb} = (screen_{XYZtoRGB} * (col_{xyz} - screen_{black}))^{(1/screen_{gamma})} \tag{7.7}$$

7.3 Known limitations

Wasted render time. Not very extensible. Fixed setup. No Alpha support - transparency would need to be a special condition and would have similar distance considerations as specular due to refraction - or any real lens/refraction based behaviour.

Chapter 8

Testing

Perceived Depth

Compare Stereographic Consistency to just the single focal

Compare Oblique projection to regular

Ask the user to determine the size of an object they are looking at in Multifocal or regular.

Define what is being tested. Survey, what is to be tested and on which scene

8.0.1 Method

How is the test taking place

8.0.2 Results

Raw results

Chapter 9

Evaluation

For any practical projects, you should almost certainly have some kind of evaluation, and it's often useful to separate this out into its own chapter.

What are the results? How did it perform? GRAPHS.

Chapter 10

Summary and Conclusions

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

potential future work.