

Depth Cues in VR Head Mounted Displays with Focus Tunable Lenses

Robert Konrad
Stanford University
Email: rkkonrad@stanford.edu

Terry Kong
Stanford University
Email: tkong@stanford.edu

Abstract—With the advent of new VR and AR technologies, it has become increasingly important to implement natural focus cues as the technology reaches more and more people. All 3D VR systems are limited by how well they handle the vergence-accommodation conflict. This project aimed to solve the vergence-accommodation conflict by using defocus blur and focus cues to simulate the perception of depth. The novelty of this project is the use of focus-tunable lenses, which until recently have been too small to have a usable field-of-view. The defocus blur will be implemented in OpenGL and a full integration of the hardware and software will be ported to the Oculus Rift. The long term goal of this project will be to incorporate an eye tracking module to automate the refocusing of the lenses.

I. INTRODUCTION

In head-mounted displays (HMD), there are many issues that prevent the technology from moving to mass distribution. One of these issues is the vergence-accommodation conflict. This problem with head-mounted displays, like the technology, is not new and is present in nearly all near-eye displays. The goal of VR is to simulate three-dimensional environments, so the technology needs to handle not only the quality of the simulations, but also how the human vision system responds to the presentation of these simulations.

II. VERGENCE-ACCOMMODATION CONFLICT

To correctly simulate a three-dimensional environment and display it to a user, the display system uses depth cues that provide information to the user's brain to help it get a sense of depth. There are two types of depth cues: monocular cues, which correspond to a single eye, and binocular cues, which correspond to two eyes. The general idea when creating an HMD that feels natural is *the more depth cues, the better*.

Many of these depth cues are neurally and necessarily coupled. This means that to properly present a 3D image to a user, the display needs to provide these cues synchronously. In stereo displays, the vergence cue (binocular) and accommodation cue (monocular) have a significant effect on the comfort of the user. *Vergence* is the simultaneous movement of the eyes in opposite directions that allows the fusion of the images captured in each eye. *Accommodation* is the process by which the eye muscles stretch the lens of the eye to focus at different depths. In HMDs, it is possible to decouple these two cues, which results in adverse physiological effects, e.g., nausea, similar to the effects that some experience in 3D movie theaters.

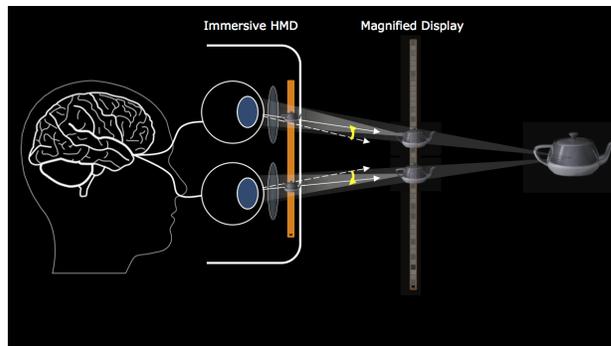


Fig. 1: Vergence-Accommodation Conflict Illustration¹

Figure 1 illustrates how the vergence-accommodation conflict arises. HMDs consist of a display placed within the focal length of a fixed lens. This creates a virtual image in the eye of the user. To provide the correct vergence cue, the display can encourage the eyes to verge at a different point, often off the virtual image plane, by shifting the image each eye sees as is depicted in Figure 1. However, the correct accommodation does not follow from this action since the user's eye will continue to focus on the screen as opposed to focusing farther off.² Thus, this motivates the idea of adding a focus-tunable lens to the HMD pipeline, which in principle, can encourage the eyes to focus at a distance off the plane of the display.

1) *Studies of the effect of Vergence-Accommodation Conflict*: The vergence-accommodation conflict has been a perpetual issue for near-eye displays. A paper in the vision sciences [1] discusses the comfort zones of stereo 3D displays and shows that the effect of the vergence-accommodation conflict is much more severe for viewing a near display. While the vergence-accommodation conflict may not be the only source of fatigue and discomfort in HMDs, this paper justifies the benefit of resolving the conflict.

III. RELATED WORK

Because different cues are extremely important in the success of HMDs providing a complete, immersive experience, there has been much research done in the area. Hua et. al. [2] implemented an optical see-through HMD with addressable

¹Image source: <http://stanford.edu/class/ee367/class17.pdf>

²As is the case in the illustration.

focus cues utilizing liquid, focus-tunable lenses. The paper discusses a vari-focal plane mode in which the accommodation cue is addressable, and a time-multiplexed multi-focal plane mode in which both the accommodation and retinal blur cues can be rendered.

In Maeillo et.al. [3] a low-cost, gaze-contingent display was created which displayed a stereoscopic image to the observer with dioptric (retinal) blur. The system was implemented using light-field photographs taken with a Lytro camera. These images support digital refocusing anywhere in the images. The group coupled this with an eye-tracking system and stereoscopic rendering and observed the effects of these depth cues on the time course of binocular fusion.

IV. OUR APPROACH

Our approach can be divided into two parts: the optics and hardware, and the software. The focus-tunable lenses (optics) can encourage the eye to accommodate, and the software is needed to simulate retinal blur, which works together with accommodation. *Retinal blur* is related to the circle of confusion that objects in out of focus image planes contribute. It is fundamental to establishing a natural impression of depth.

A. Choice of HMD

One of the requirements of our system is to integrate it with a head mounted display (HMD). We had the choice of either creating one from scratch, or augmenting our lenses into an existing HMD. Because of the limited time we had, we decided to integrate our lenses with a current HMD.

Of the many HMDs that are currently on the market, there are a few that have a developer kit available. The Sony Morpheus, Epson Moverio, and Oculus Rift are a few such products. The main requirement that the HMD had to satisfy was it must be possible to place our focus-tunable lenses directly in the optical path. Since this project was time-sensitive, we were restricted by the availability of HMDs that we could get our hands on within the span of a few weeks.

Neither the Epson Moverio nor the Sony Morpheus have the ability to remove the lenses. Also, we did not have immediate access to either of these products. The Oculus Rift, immediately available through Professor Wetzstein's lab, fortunately comes with multiple removable lenses that can be easily swapped out. The lenses are attached to cups that can be twisted into the headset. Because of this we chose to use the Oculus Rift as the HMD that we would integrate our focus tunable lenses into. Terry designed cups similar to those of the Oculus that housed our lenses and could be twisted into the Oculus headset.

B. Optics

Since the focus-tunable lenses used in this project are convex lens, we can restrict our discussion to focusing farther than the virtual plane.³ A convex lens has the property of converging light rays to a point. In our application, we desire that the light rays emanating from the display plane to

converge on our retina. If we use a fixed focal length lens that focuses light from the display to our eye, everything on the display plane will be in focus and appear at the virtual plane. This is illustrated in Figure 2. If we try to accommodate, or focus farther than the screen, the light coming into our eye will not focus at a point and the image will appear blurred. This is the limitation of a fixed focal length lens. The inability to accommodate at any other distance is precisely why there is a mismatch between vergence and accommodation.

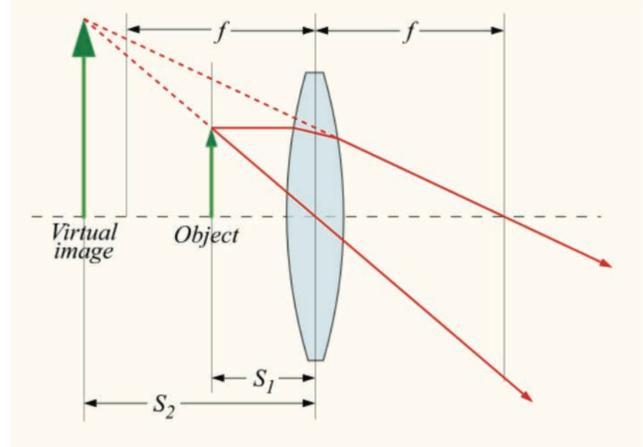


Fig. 2: Virtual Image Plane⁴

So what we desire is a lens that can change its focal length in order to encourage the lens of the eye to focus closer or farther. Now considering a focus-tunable lens, if an object moves away from the eye off the virtual plane, which was previously in focus, then the eye needs to increase its focal length. According to the effect described above for a fixed focal length lens, this would result in a blurred image. However, if the focus-tunable lens increases its focal power, i.e., decrease its focal length, the lens can focus the gaze of the eye back to the display plane which brings the object back into focus. Hence, the user should perceive an object at a farther distance.

It is easier to discuss the analysis of the focus-tunable lens starting from the eye and how the lens must change its focal length to allow the eye to focus back on the screen. However, when setting the focal length of the lens, it is not necessary to know the focal length of the eye.⁵ This is because once the focal length of the lens is set, the eye will automatically accommodate to make the image on the display appear in focus. So if the focal length of the lens decreases, the eye's focal length will increase to ensure the image is in focus. The equation to set the the focal length of the focus-tunable lens is simply the len's maker equation:

$$\frac{1}{f} = \frac{1}{S_{display}} - \frac{1}{S_{virtual}}$$

⁴Image source: <http://stanford.edu/class/ee367/class3.pdf>

⁵The focal length of the eye is not irrelevant. The point is just that the focal length of the focus-tunable lens dictates the focal length of the eye, not the other way around. This idea that the lens can force our eye to accommodate at different depths is a remarkable fact.

³Although the same analysis can be made for a concave lens

where f is the focal length, $S_{display}$ is the distance from the lens to the display, and $S_{virtual}$ is the distance from the lens to the virtual plane.⁶

C. Benefits of Using Focus-Tunable Lenses

The discomfort that one experiences due to the vergence-accommodation conflict is not binary. Rather, each individual experiences variable levels of discomfort related to the severity of the vergence-accommodation conflict. In Figure 3 from [1], the black diagonal curve represents the natural viewing and any point off the diagonal represents a vergence accommodation mismatch where points farther away represent more mismatch. With this visual in mind it is easy to see what the focus-tunable lenses do: they move a point on this graph closer to the diagonal curve. Hence, it should become more comfortable to focus on an object that is off the virtual image plane.

It is important to note that the focus-tunable lenses used in this project have a refresh rate of approximately $15ms$, which is roughly on the order of the refresh rate of a standard display.⁷

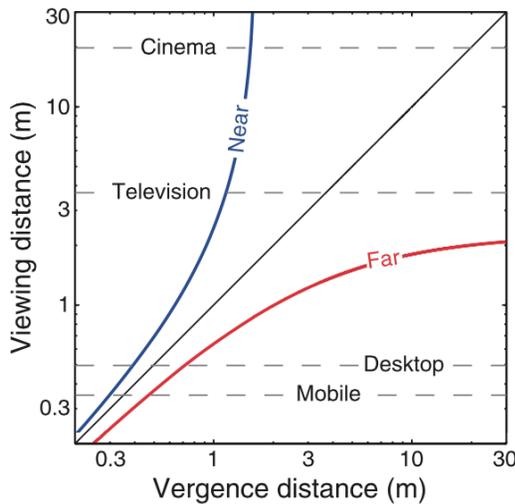


Fig. 3: Comfort Zone: The region between the blue and red curve is the comfort zone and the black diagonal curve represents the natural viewing of a display.⁸

D. Limitations of Focus-Tunable Lenses

There are a couple limitations imposed by focus-tunable lenses available now:

- The field-of-view on the focus-tunable lenses used in this project is quite small (30° for the Optotune EL-10-30-VIS-LD). To utilize the full field-of-view, the lenses

⁶In OpenGL, $S_{virtual}$ is related to where the object is in world space.

⁷Standard displays have a refresh rate of 60Hz, which corresponds to $16ms$ per frame.

⁸Image source: Shibata, T., Kim, J., Hoffman, D., & Banks, M., (2011). The zone of comfort: Predicting visual comfort in stereoscopic displays. *Journal of Vision*, 11(8), 11.

should be moved farther back. However, moving the lenses too far back is not practical for an HMD, not to mention that the range of objects the lenses can force the eye to accommodate on diminishes. Focus-tunable lenses with larger field-of-views are available, but these rely on a mechanical input to change the focus.

- As a practical issue, these electrically tuned lenses are controlled via an electromagnetic actuator. This requires a significant amount of current, which can cause the lens to heat up. This has two implications:
 - Heating up the fluid in the lenses has an effect on the focal power of the lenses since the pressure of the fluid within the lens is related to the focal power of the lens. Fortunately, Optotune has worked out a heuristic for the relationship between heating up and the focal power, which is accounted for in certain modes of operation.
 - When the lens reach the highest current, which corresponds to the highest focal power, the lenses can become significantly hot. This may be an issue since given the small field-of-view on these lenses, the user must bring the lens as close to their eye as possible to maximize the usable field-of-view. There may be a workaround for this, but this project did not explore this issue.
- For this project, the lenses were placed very close to the eye. This complicates the matter of placing any other module between the eye and the lens, e.g., an eye tracker.

V. SOFTWARE

While the first part of the project concentrated on providing focus cues by changing the focal length of the lenses, the second part focused on simulating retinal blur. In order to dynamically change the scene, retinal blur was implemented in software. We created an OpenGL program written in C++, which renders a stereoscopic scene and blurs it in real-time to emulate retina blur.

A. Stereoscopic Scene

In order to render a stereoscopic scene we used asymmetric frusti that matched with the dimensions of the screen. A diagram of how we constructed the frustum can be seen in Figure 4.

We are able to use similar triangles to come up with the left and right vertical clipping planes of this asymmetric frustum. The frustum is asymmetric because the lens is not placed at the center of the screen, but rather at a position corresponding to the intra-pupillary distance. Calculating the top and bottom vertical clipping planes is very similar. However, because we assume that the lenses are placed at the vertical center of the screen, the frustum in that case is symmetric.

B. Retinal blur in Software

Retinal blurring is a depth cue experienced prominently for objects approximately within $10m$ of the observer. Objects outside of the plane at which the eye is focused on are

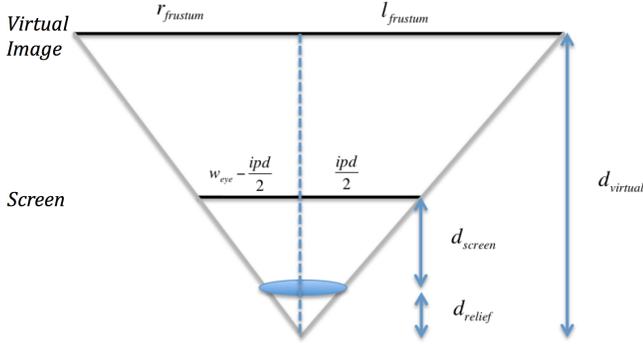


Fig. 4: Frustum

blurred by some amount. In our project, retinal blur is modeled analogously to the circle of confusion experienced by cameras. The equation for the diameter of the circle of confusion, c , is as follows:

$$c = M \times D \times \frac{|S_f - S_o|}{S_o}$$

where M is the magnification factor of the lens, D is the diameter of our pupil, S_f is the distance from the lens to the focal plane, and S_o is the distance from the lens to the object. We are rendering this retinal blur on the virtual image which is already magnified, therefore we set $M = 1$. We set $D = 2cm$ which is the average pupil size for humans under normal lighting conditions. The circle of confusion equation is applied at every fragment that is to be displayed, and the natural choice is to place this calculation inside of the fragment shader.

The circle of confusion represents the optical spot not coming to perfect focus when imaging a point source. For our scenario, this represents the size of a blurring kernel. In our implementation we used a separable 9-tap Gaussian filter to implement the blurring. We chose Gaussian filtering because it is simple to implement, and also computationally efficient. Again, this blurring is done on a per fragment basis.

Clearly, from the above equation, we must have depth information in order to accurately simulate retinal blurring. Because we are rendering a computer-generated image, the depth information about the scene is known a priori, and can be accessed through the OpenGL framework. This problem would be more difficult if we were to use images from the real world where perfect depth information is not available or accurately recorded.

In an ideal system we would be able to set the focal plane based on where the eye is looking via eye tracking, and couple the movement of our eyes with retinal blurring. However, due to the difficulty of the problem and our decision to create a proof of concept, we decided to set the focal plane via a click of the mouse. This is something we are planning on improving in the future and will be discussed in greater extent in the Future Works section below.



(a) Single Teapot: Demonstrates effect of moving the teapot to different virtual planes and refocusing.



(b) Teapots with Occlusion: Demonstrates the effect of focusing at one depth has on objects at other depths.



(c) Depth Test: Gauges the effect that the focus-tunable lenses have on the user.

Fig. 5: Oculus Rift Results

VI. EVALUATION/RESULTS

We accomplished our goal of implementing the system on an Oculus Rift. With the click of a mouse we are able to set the focal plane, which sets the correct focal length of the lens and renders an appropriate retinal blur. Because our results are very subjective we would have liked to implement a user study to better understand the effects of these focus and depth cues on the visual system. However, due to lack of time we were not able to implement such a study.

However, we were able to show a demo at the project fair which displayed three different scenes. The scenes are shown below in Figure 5. In the first scene in Fig. 5a, the user was able to control the placement of the teapot and accommodate at different focal planes. In the second scene in Fig. 5b, the teapots were static but placed at different depths. The most interesting of the scenes is the 3rd one in Fig. 5c, where all of the teapots appear to be the same size. In reality they are each sized differently and are at different depths. During our demo session we showed this scene to our users and asked which one appeared closest. We were observing the effect of the focus cue on our ability to estimate depth. In the informal

setting, we saw that a majority of the people is able to correctly classify the closest teapot, but because this was not a rigorous user study these results mean little. We plan on conducting an extensive user study in the future.

VII. DISCUSSION/CONCLUSION

In this paper, we proposed a near-eye display technology that supports focus cues via focus tunable lenses, and the retinal blur depth cue. We believe that moving forward, focus cues along with depth cues other than stereopsis will be critical in the success of HMD's. Our visual systems are complicated and take many different cues in order to accurately understand a scene. We believe that successful HMDs will be the ones that are able to capture as many of these cues as possible, creating a scene most similar to that viewed in the real world.

Our bodies do not solely understand a scene via visual cues. Audio and haptics are extremely important for us to immerse ourselves in a scene, and sometimes are critical for us to view certain actions comfortably. Similar to how the mismatch in the cues of vergence and accommodation can cause adverse effects in many users, other mismatches in cues can cause similar, and often more drastic, effects. An example of this is the mismatch between our visual system and our vestibular system when a jumping action is performed inside of virtual reality. Unless the user is actually jumping in real life, many users experience immediate nausea from such an action. The visual system is telling them that they are jumping, while their vestibular system tells them that they are stationary. It will be interesting to see which mismatches our bodies will be able to tolerate and which ones will limit the capabilities of virtual reality headsets.

VIII. FUTURE WORK

Eye tracking would be an interesting extension to this project, because it would allow the entire system to be automated, giving the user seamless transitions between focus and depth cues. Integrating eye-tracking into an HMD is a difficult problem and in general difficult to solve, especially in compact devices such as HMDs. There are only a handful of companies that currently offer to modify the Oculus to incorporate eye-tracking. Unfortunately, the price of modification is astronomical, not to mention the modification only supports the standard Oculus lenses.

In general, we see two possible solutions to the eye-tracking problem for our system. One solution is we can place the camera behind the lens and track the eye through the lens. This approach has problems. Since the focal length of the lens is varying, the eye tracking algorithm would have to be robust and invariant to the changes in the lens' focal length, making this a very challenging problem. Another solution would be to track the movement of the eye with a camera set in front of the eye between the eye and the lens. The issue here is that the camera must have line of sight with the eye, requiring an offset of the eye from the lens. Because the diameter of our focus-tunable lenses is small (10mm), they have very limited field-of-view and we cannot afford to offset the eye even a

little. With the eye pressed up against the lens, we experienced a field-of-view of around 30 degrees. With an offset of even one cm, this field-of-view would drop to around 10 degrees.

We have also thought to improve the quality of the retinal blurring, in the sense that we want to make it more similar to the natural blurring experienced by our eyes. We would like to implement a bilateral filter instead of the currently used Gaussian filter. A bilateral filter preserves edges better than a Gaussian filter and is also a separable filter, making it computationally efficient. We are also interested in higher order models of the eye to implement more accurate blurring, perhaps those presented in the graphics community. However, we are constrained by the strict latency requirements of virtual reality in order to provide an immersive feeling.

Finally, we would like to improve on the field-of-view the focus tunable lens. Adding elements to the optical path could potentially increase the field-of-view of the lenses, resulting in a more immersive experience for the user.

REFERENCES

- [1] Shibata, T., Kim, J., Hoffman, D., & Banks, M., (2011). The zone of comfort: Predicting visual comfort in stereoscopic displays. *Journal of Vision*, 11(8), 11.
- [2] Sheng Liu, Dewen Cheng, Hong Hua, "An Optical See-Through Head Mounted Display with Addressable Focal Planes", Proc. Of IEEE and ACM ISMAR '08, Cambridge, UK, 2008
- [3] G. Maiello, M. Chessa, F. Solari, P. Bex. "Simulated disparity and peripheral blur interact during binocular fusion." *Journal of Vision*, vol. 14, July 2014.

APPENDIX

Photos of the setup



Fig. 6: Oculus Rift with Lenses Installed



Fig. 7: Oculus Rift with Lenses in Holders