

Focus Cue Enabled Head-Mounted Display via Microlens Array

Ryan Burke
Stanford University
Electrical Engineering Department
rburke2@stanford.edu

Leandra Brickson
Stanford University
Electrical Engineering Department
llbricks@stanford.edu

Abstract

A new head mounted display (HMD) is designed utilizing a microlens array to resolve the vergence-accommodation conflict common in current HMDs. The hardware and software are examined via simulation to optimize a depth and resolution tradeoff in this system to give the best overall user experience. Ray tracing was used to design the optics so that all viewing zones were mapped to the eye to provide focus cues. A program was also written to convert Lytro stereo pairs to a light field useable by the system.

1. Introduction

There has recently been a lot of research going into developing consumer virtual reality (VR) systems aiming to give the user an immersive 3D experience through the use of head mounted displays. One notable example of this would be the google cardboard system, which consists only of an objective lens and a smartphone display. This setup projects a stereo pair of images from the smartphone onto a virtual plane in front of the user. The pair of images (one image shown to each eye) gives stereo cues of depth, giving a 3D perception of the scene. This system is extremely useful and cheap, making virtual reality accessible to almost anyone. However, the main problem with stereoscopic displays, such as google cardboard is that there are no focus cues supporting accommodation. This leads to a vergence-accommodation conflict, which can be uncomfortable to a user after extended use.

There exist currently a few examples of HMDs which have already resolved this issue. Lanman and Luebke created a near-eye light field display using a microlens array, OLED panels, and a personal 3D viewer [1]. This system was created as a low form-factor HMD with focus cues. Additionally, Wetzstein, et al. created a light field stereoscope with focus cues using stacked transparent liquid crystal displays viewed through lenses. This system showed very promising results, giving a high resolution,

high field of view scene while still supporting focus cues.

While these systems are good solutions for the vergence-accommodation conflict, they still are not as accessible to the public as google cardboard. Our project investigates adding focus cues to the google cardboard system via microlens array, eliminating the vergence-accommodation conflict.

2. Optical System Design

The proposed optical design of the focus cue enabled head mounted system is shown in figure 1. In this, the google cardboard system is augmented by placing a microlens array (MLA) in front of the display to create a light field, which will then be projected to the virtual image plane by the objective lens. In this system, the MLA will create different viewing zones, which will provide depth cues for accommodation. However, with this added depth information, there will be a loss in spatial resolution of the projected image, as shown in the results section. As shown in figure 2, each pixel under each lenslet in the array will correspond to its own viewing zone.

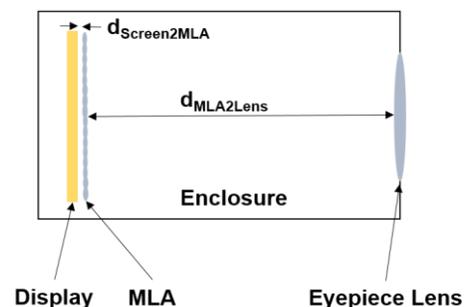


Figure 1: Diagram of proposed optical system design and associated distances. The design consists of a display, a microlens array and an eyepiece or objective lens.

This system must be designed so that all viewing zones can be resolved by an eye located at the eyepiece lens. This

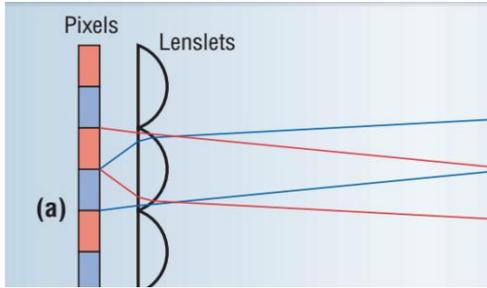


Figure 2: Graphical depiction of each pixel under the lenslet being mapped to its own viewing zone. In this example, a lenslet covering a 2x2 pixel space results in 2 viewing zones per dimension [4].

means that for this system, an eye relief of 0 was assumed so that everything on the objective lens would be redirected to the eye. Therefore, through optical system engineering, we can ensure that all viewing zones projected by the microlens array will be perceived by a user.

3. Image Processing Pipeline

In order to display content on this system, images displayed on the smartphone will need to be processed to send the correct information to each viewing zone so that the user will perceive the light field of a scene. The source of this information will be a stereo pair of light field images captured using a Lytro camera. These images are composed of many 14 x 14 sets of pixels corresponding to one lenslet inside the camera.

In order to fit the new resolution of the smartphone screen, the Lytro images are first downsampled, or interpolated at a subpixel level if necessary. The downsampling ratio is given by $n/16$, so that the current light field is composed of $n \times n$ sets of pixels for each lens. This gives us the correct light field, but the image size is too large. At this point, the image processing could be halted and the current light field could be cropped to create a stereo pair.

The remainder of the pipeline tries to integrate as much of the original image into the stereo pair as possible. We need to resize the images, but the light field cannot be resized using typical methods. To properly resize the light field we split it up into an image stack composed of $n \times 376 \times 541$ images. These images give n different perspectives on the scene and can be manipulated as normal images. The entire stack is resized such that the smallest dimension will match the smallest dimension of the screen of choice's resolution. The image stack is reformed into a light field and the light field is cropped and formed into a stereo pair with the same resolution as the smartphone screen. This process is outlined in figure 3.

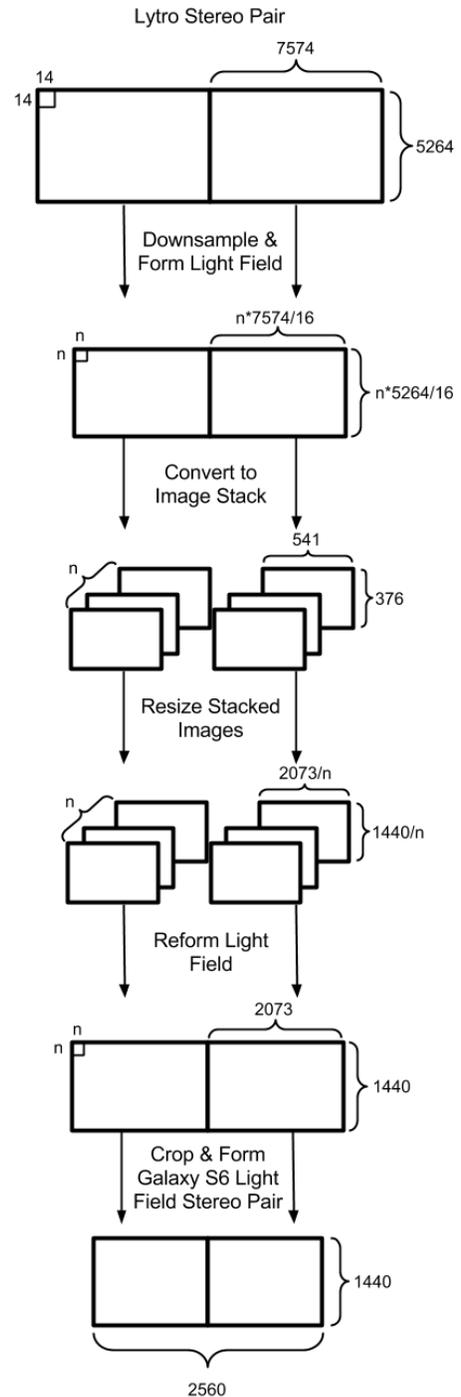


Figure 3: Image process pipeline outlining the process going from the original Lytro source images to what is displayed on the smartphone screen.

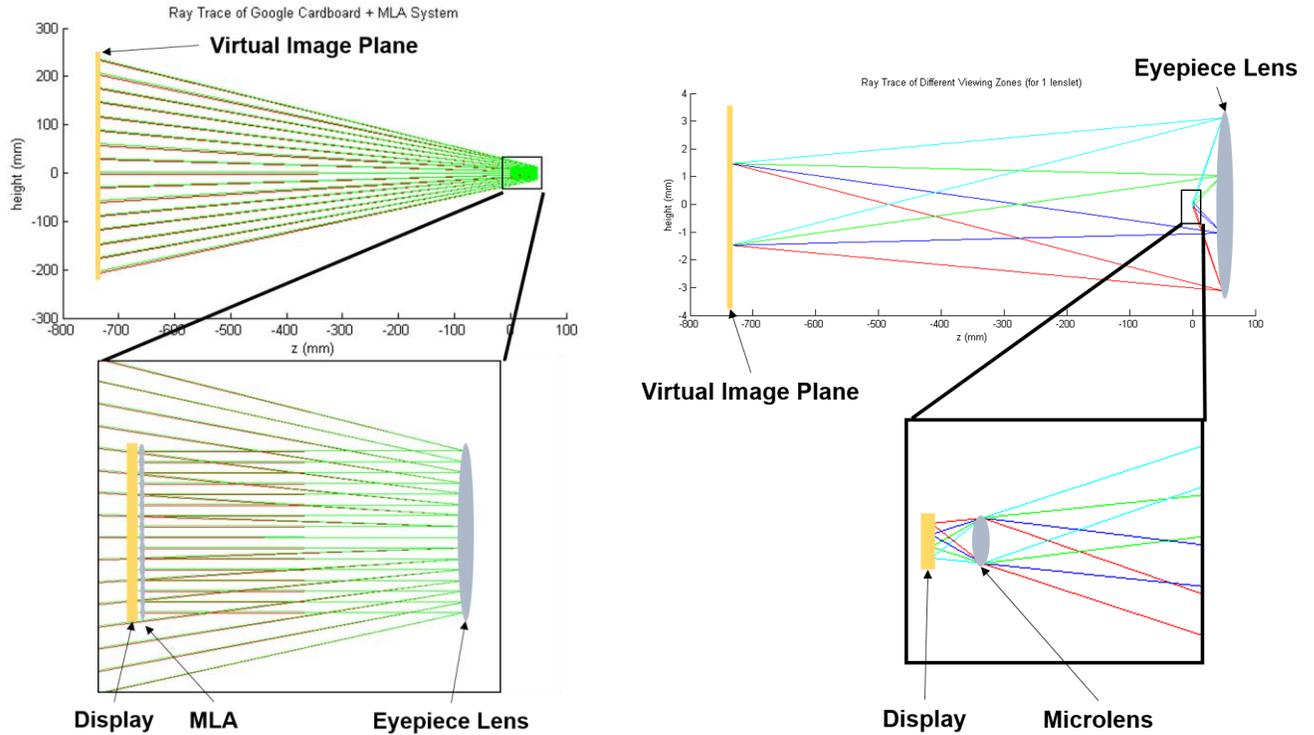


Figure 4: Ray tracing software results. The left side shows the ray tracing of the proposed optical system, the full ray trace is shown above, a zoomed in portion of the region between the MLA and the eyepiece lens. This shows the focus of the microlens array onto the objective lens and the loss of resolution at the virtual plane. The right side shows a ray trace of 4 pixels under a single microlens. The top shows the full system and the zoomed in portion shows the 4 pixels and the microlens. This ray trace illustrates that all viewing zones are being mapped to the user's eye.

4. Results and Discussion

4.1. Ray Tracing the Optical System

Ray tracing calculations using ray transfer matrices were done to characterize the system. Figure 4, left shows a marginal ray trace of single pixels through the google cardboard + MLA system. It can be seen that at the virtual plane, the image doesn't focus, this indicates the resolution loss in the image due to the addition of the microlens array and can be used to calculate the spatial resolution loss, for the parameters given in Table 1, the spatial resolution is 2.93mm/pixel.

Figure 4, right shows a marginal ray trace of each pixel under a single microlens in the array. Each color represents the viewing zone projected by each pixel. In this case, the lenslet covers a 4x4 pixel region, giving 4 viewing zones in each direction. Each viewing zone is mapped to a different location on the objective lens, giving a different view of the scene. Since an eye relief of 0 is assumed, as long as marginal ray is within the diameter of the objective lens,

then it will be seen.

Table 1 gives the final details of the optical system parameters.

4.2. Depth Resolution Tradeoff

Increasing the number of viewing zones improves the depth quality of the light field, but reduces the effective resolution of the system for a fixed screen size. To decide the optimal number of pixels per lens, the depth quality and effective resolution are compared independently for 3, 4, 5, and 6 pixels per lens (meaning a 3 x 3, 6 x 6 of pixels under each lens).

Depth quality is assessed by creating all-in-focus images from the downsampled Lytro light fields. As the number of pixels per lens increases, so does the depth image quality. In particular very low pixels per lens creates duplicates of objects that are far from the original focus plane. The all-in-focus images are shown for 3, 4, 5, and 6 pixels per lens in figure 5.

Eyepiece Diameter	37 mm
Eyepiece Focal Length	50 mm
Screen to MLA dist	1 mm
MLA to Eyepiece dist	47 mm
Pixels Per Lenslet	4 pixels
MLA Focal Length	979.2 μm

Table 1: Final optical design parameters for the proposed system. All values were calculated using the ray tracing software.

Resolution quality is assessed by passing a Lytro image through the main image processing pipeline for varying numbers of pixels per lens. The resulting light fields appear to have similar resolutions because the pixels are being viewed without the microlens array. For larger numbers of pixels per lens however, the display squares corresponding to each lens become visible. Additionally, the effective resolution of a single image of the stereo pair is calculated and considered, because that is what the viewer will see with the final product. The images showing the resolution tradeoff for 3, 4, 5, and 6 pixels per lens are shown in figure 5.

The Depth and resolution quality images were used to determine an acceptable number of pixels per lens. Upon inspection, the design team decided 4 pixels per lens to be appropriate. A stereo pair of Lytro images was captured and processed using the chosen 4 viewing zones to create the image shown in figure 6, which can be displayed on a Galaxy S6.

5. Lenslet Array Hardware Implementation

An attempt was made to fabricate the system using an Eyefly 3D lenslet array, an iPhone 5C and a google cardboard head-mounted unit. However after examining the lenslet array under a microscope, it was discovered that the lenses are arranged diagonally on the screen. Though diagonal encoding of zones could be done, we deemed this too in-depth for this project.

6. Future Work

The most crucial step of future work will be to get the microlens array and to find a way to accurately apply it to a phone. Once this step is complete we will be able to verify our calculations and qualitatively analyze the depth quality

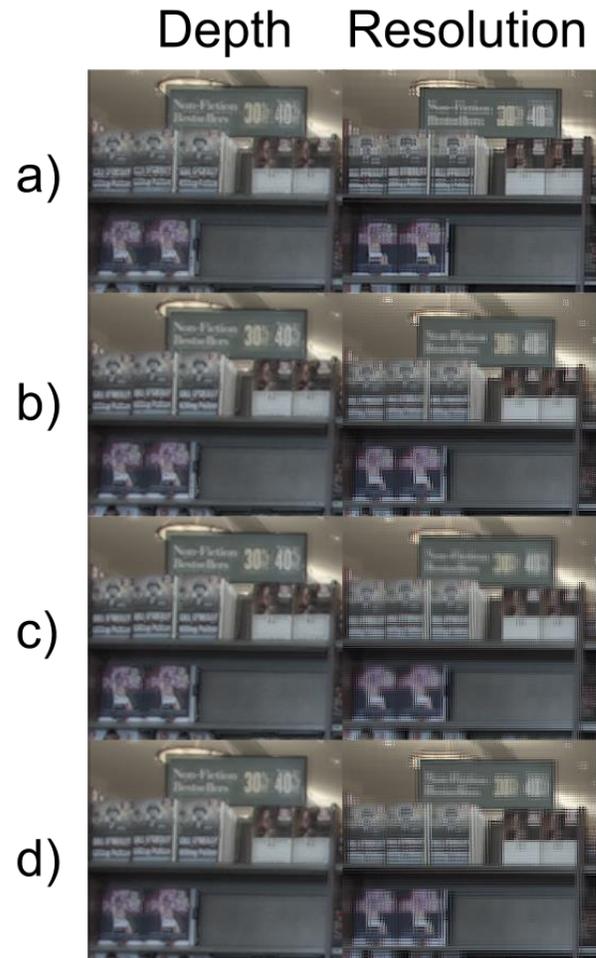


Figure 5: Images assessing the depth and resolution quality for a) 3, b) 4, c) 5, and d) 6 pixels per lens.

of the google cardboard stereo image. We can also calculate product pricing from manufacturing costs if we want to sell the microlens array/software.

Further work should also be done to investigate the effects of the microlens array on pixels located near the corners of the lenslets. For Lytro images, these pixels are frequently black, because they capture less light. We may need to account for this disparity by increasing the intensity of these pixels, when possible.

Finally, another possible step for improved user experience is aberration correction. The most prominent aberration is inward curvature caused by the objective. To account for this, the image to be displayed is usually bowed outward such that the user observes a square image on the virtual plane. An additional consequence of this aberration



Figure 6: A stereo pair for the proposed system with 4 pixels per lens, to be displayed on an Android Galaxy S6

correction is that some screen resolution has to be sacrificed to allow room for the image to bow outwards. This would be accomplished in software, but additional considerations would have to be made keeping the effects of the microlens array in mind.

7. Conclusion

In this project, a Google Cardboard HMD was augmented to resolve the vergence-accommodation conflict. This goal was achieved by adding a microlens array to the HMD phone screen to provide different viewing zones, giving focus cues for accommodation. A set of Lytro images were processed to find the number of pixels per lens with an appropriate depth quality/resolution tradeoff. Ray tracing software was created in MATLAB to define optical system specifications and to verify that all viewing zones are mapped to the eye. Finally a program was created to take two Lytro images and create a stereo pair ready for display. Using this framework, a low-cost virtual reality system with depth and focus cues can be easily created.

References

- [1] Lanman, Douglas, and David Luebke. "Near-eye light field displays." *ACM Transactions on Graphics (TOG)* 32.6 (2013): 220.
- [2] Perlin, Ken, Salvatore Paxia, and Joel S. Kollin. "An Autostereoscopic Display." *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 2000.
- [3] Huang, Fu-Chung, David Luebke, and Gordon Wetzstein. "The Light Field Stereoscope." *ACM SIGGRAPH Emerging Technologies* (2015): 24.
- [4] Dodgson, N. "Autostereoscopic 3D Displays" *IEEE Computer Society, August, 2005 0018-9162*
- [5] Schwerdtner, A, Heidrich, H "The Dresden 3D Display" *SPIE's Stereoscopic Displays and Applications IX Vol. 3295, 0277-786X*
- [6] Maimone, A, Wetzstein, G, Hirsch, M., Lanman, D. Raskar, R. "Focus 3D: Compressive Accomodation Display" *ACM Transactions on Graphics*.
- [7] Ren Ng, "Digital Light Field Lithography", Stanford Univeristy Department of Computer Science phd Thesis, July 2006.