Static Scene Light Field Stereoscope

Kevin Chen
Stanford University
350 Serra Mall, Stanford, CA 94305
kchen92@stanford.edu

Abstract

Advances in hardware technologies and recent developments in compressive light field displays have made it possible to build inexpensive displays which support multiple viewing angles and provide focus cues. Current stereoscopic display technologies provide a wealth of depth cues such as binocular disparity, vergence, occlusions, and more. However, focus cues which provide significant sense of depth are not yet in commercialized products because of tradeoffs such as form factor or cost. In this paper, we apply these compressive light field techniques to a head-mounted display where the views are limited over a small eyebox, allowing for low rank-1 approximations of the light field which are sufficient for focusing at different depths. This could be a cheap alternative to other display technologies which allow for high resolution displays that support accommodation with reasonable form factor. Specifically, this paper focuses on the display that shows only static scenes rather than dynamically changing scenes.

1. Introduction

Head-mounted displays (HMD) and virtual reality (VR) have gained significant interest in recent years especially after the advent of Oculus Rift and Google Glass. Although this marketing is primarily aimed towards the gaming audience, many other applications arise in areas such as simulation and training, scientific visualization, phobia treatment, remote-controlled vehicles, and therapy for various disorders. VR simulations have improved operating room performance for residents conducting laparoscopic cholecystectomy [Seymour et al. 2002] and VR has also been shown to be effective at treating post-traumatic stress disorder [Rothbaum et al. 2001].

To provide a truly immersive experience, VR HMDs need to provide depth cues such as shadows, motion parallax, binocular displarity, binocular disparity, binocular occlusions, and vergence. However, one important cue that has not been implemented in commercialized products is focus cues. Nearly correct focus cues significantly improve depth perception, stereoscopic correspondence matching [Hoffman and Banks 2010], and 3D shape perception becomes more veridical [Watt et al. 2005].

Furthermore, it is essential that HMDs support these focus cues because of the vergence-accommodation conflict, which creates discomfort in users and can cause nausea, headaches, and possibly even pathologies in the developing visual systems of children. The vergence-accommodation conflict arises from a decoupling of two cues, vergence and accommodation. Vergence arises from rotation of the eyeballs and accommodation comes from the changing focal length of the eye lens depending on object location. Without solving the vergence-accommodation problem, HMDs cannot be practically be used over long periods of time.

We propose a near-eye stereoscopic light field display that presents a 4D light field to each eye, each of which encodes focus cue information. Similar to the compressive light field displays [Lanman et al. 2010; Wetzstein et al. 2011; Wetzstein et al. 2012], we use stacked spatial light modulators to approximate the light field. Since the light field only needs to be defined over a small eyebox, a rank-1 approximation is sufficient for providing focus cues. In particular, in this paper we discuss the prototype that displays static scenes which allows the device to be lightweight and portable. For the dynamic prototype, please see the paper by Huang et al. [2015].

2. Related Work

There is much research being done in the virtual reality field in attempt to solve the vergence-accommodation problem in order to create a comfortable and immersive viewing experience, but many of these attempts include tradeoffs such as form factor or resolution that prevent them from being commercialized. For example, holography [Benton and Bove 2006] provides all depth cues but require complex systems and high computational power, making it impractical for use in VR headsets. Volumetric displays using mechanically spinning parts [Favalora 2005; Jones et al. 2007] are also infeasible for use in wearable displays.

There are also multi-focal plane displays which are able to provide nearly-correct focus cues but often require complex hardware. They often use expensive liquid lenses

	multi-focal plane displays	near-eye light field displays	factored near-eye light field display
resolution	high	low	high
hardware complexity	high	medium	low-medium
form factor	large	small	small-medium
brightness	normal	normal	low
computational cost	medium	medium	medium
accommodation range	high	low	high
retinal blur quality	medium	low	high

Table 1. *A comparison of current displays that support focus cues with our dynamic factored near-eye light field display.*

which have limited field of view, high speed displays such as 240 or 300 Hz displays to allow for time-multiplexing, and/or bulky form factor [Liu et al. 2008; Love et al. 2009; Mackenzie et al. 2010]. This makes them non-ideal for commercialized use.

Recently, however, researchers have begun developing light field displays. Lanman et al. [2013] recently constructed a light field display using a micro-lens array with an extremely lightweight and portable form factor. This display allowed the user to accommodate, but the downside was the resolution since several pixels in the display screen became one effective pixel. Instead, the approach we used to create a light field display which supports focus cues is similar to that of Maimone et al. [2013]. This allows for higher resolution displays with better form factor than that of the multi-focal plane displays. Also similar to the work of Wetzstein et al. [2011], we used stacked spatial light modulators to create a compressive light field display, but specialize this towards near-eye light field displays such that the eyebox is much smaller and the generated light field can be derived using a low rank-1 approximation. This allows for a cheap solution with off-the-shelf inexpensive parts, and also uses multiplicative image formation to provide better depth cues. The main downside to this approach is reduced brightness which is almost a non-issue in a HMD since the world light does not have to be taken into account, so the user's eyes can adapt to the reduced brightness.

Figure 1. *The final prototype as a Google Cardboard with modified lenses and an additional backlight.*

3. Method

3.1. Hardware

The components involved in building the HMD were: acrylic sheets, inkjet (or laser) transparencies, Google Cardboard, batteries, LCDs, and 50 mm 5x aspheric lenses from eBay (purchased at http://www.ebay.com/itm/5x-pocket-loupe-magnifier-with-Aspheric-Lens-/251148139033?pt=LH_DefaultDomain_0&hash=item3a79987a19).

Although we initially planned to use the my3D viewer, the focal length of the lens and the construction of the viewing device makes it difficult to work with. In particular, we would need to cut into the plastic housing to put the image planes in the right locations, and the shape of the housing makes it difficult to hold the transparencies and acrylic in a stable position. Furthermore, swapping out the scenes would be very difficult.

I advocate the use of a Google Cardboard because of its simplicity, versatility, and easy access. It is a very cheap solution that is easy to work with and allows the user to quickly swap scenes by just sliding the previous scene out and inserting a new one. The downsides to using the Google Cardboard are lack of robustness compared to plastic housings and alignment issues. The cardboard is clearly not as robust as plastic such as the my3D Viewer but it should still be sufficient for our uses. Any accidental drop of the Cardboard should not result in significantly more damage to the backlight than a plastic housing (unless if the backlight can fit entirely inside the plastic housing). Aside from wiring, battery holders, and the backlight, no other

components in the device should be damaged from a drop.

I also decided to use the 50 mm aspheric lenses from eBay because they were the same that were used in the dynamic prototype. This way, lens distortion was easier to account for, which can be very troublesome to deal with in a static prototype.

The acrylic sheets I used to space the transparencies were 1/16" thick. They were laser cut to match the size of the Google Cardboard (12 cm x 7 cm). To make the static prototype display the layers at the same magnified virtual distances as the dynamic prototype, the transparencies were spaced four layers apart.

3.2. Construction and assembly

The first step was to modify the Google Cardboard to fit the larger lenses. This was done by using a box cutter to cut holes in the Cardboard.

Extracting the backlight from the LCD varies on the specific LCD, making it difficult to find the right one to purchase. For example, they may use different connectors and run at different voltages. However, they all still have similar structure. To take apart the backlight, I first removed the LCD portion from the display. Generally, this can be done by removing the clips on the frame using a flathead or knife. After removing the top frame, the LCD can be separated from the backlight, but they may or may not be connected by a ribbon cable. In one of my LCDs, the backlight had a separate connector consisting of a red wire and a white wire (ground) for turning on and turning off the display. Using the backlight was a simple matter of hooking up 9V to the two wires (or fewer volts if hooked up in parallel to the LEDs). These were then connected to a switch to allow the user to conveniently turn on and off the backlight. However, in two other LCDs, the backlight was connected to the LCD board by a ribbon cable. This seems more common. Along an edge of the display, there is a narrow circuit board consisting of several LEDs in a line. These light up the backlight by passing through several materials used to make the light uniform. Therefore, to power the backlight, the LEDs need to be powered. Soldering wires to the ribbon cable can be very difficult since there is not much space to work with. Some ribbon cables consist of two wires or three wires (similar to a switch). Connecting the wires directly to the ribbon cables would require more voltage since the LEDs are wired in series. It is recommended to connect the wires in parallel to the LEDs to reduce the number of batteries required to power the device. This is described later in the paper.

Figure 2. *The LEDs are located on the back side of the strip of circuit board. There is very little room to work with in order to solder wires to individual LEDs, but this should be done in order to reduce number of batteries on the device.*

Figure 3. *The LCD is separated from the backlight.*

3.3. Light field parameters

I attempted to match the static prototype with the dynamic prototype as closely as possible. Therefore, I tried to put the light field origin (unmagnified) at 4.3 cm from the lens, the first transparency layer at 4.0 cm, and the second transparency layer at 4.6 cm. With a 5 cm focal length lens, this puts the light field, first transparency, and second transparency at virtual distances of 30.71 cm, 20 cm, and 57.5 cm according to the thin lens approximation. Using a different pair of lens, one should make sure that the image plane distances are the same with the formula below, where f is the focal length of the lens, o is the distance from the lens

to the object, and i is the distance from the lens to image (which should be negative).

$$\frac{1}{f} = \frac{1}{o} + \frac{1}{i}$$

The light field resolution and size should also match each other in aspect ratio. For example, a light field resolution of 1440 x 900 pixels should have a display size with the same aspect ratio as shown in the formula below which keeps the ratios the same. If the display size height is chosen to be 7 cm, the width should be 11.2 cm. The observer distance was set to 7.35 cm. This should not be changed.

$$\frac{lf\ width}{lf\ height} = \frac{display\ width}{display\ height}$$

The parameters for the dynamic prototype were the following. The light field resolution was 1440 x 900 pixels with 5x5 views. The pupil size was 0.5 x 0.5 cm, the distance between the viewer and screen in config was set to 7.35 (observer distance). The layers were 0.3 cm from the light field origin, so layer 1, light field origin, and layer 2 were located at 4 cm, 4.3 cm, and 4.6 cm. The number of NTF iterations was 5 and the focal length of the lens was 5.

The parameters for the Google Cardboard prototype were the following. Only the light field size, layer depth positions, distance to light field, and lens focal length should be changed. The user should verify that the layers and light field origin are at approximately the same distance. Moreover, the observer distance should be the same, otherwise the results will be different. The Google Cardboard prototype used a resolution of 1440 x 900 pixels, 5x5 views, an eyebox of 0.5 x 0.5 cm, and distance between viewer and screen was 7.35 cm. The layer distances from light field origin were 0.287 and -0.287 cm measured using a caliper. Always use a caliper to double-check the thickness of the acrylic sheets. The display size was 11.2 cm by 7 cm and the layer offsets were set to 0. The number of NTF iterations was left at 5 and the focal length of the lens was also 5 cm since we used the same lenses. The light field origin was left at 4.3 cm. This resulted in a distance from lens to magnified layers of 20.329 and 55.533 cm, fairly close to our dynamic prototype which had distances of 20 cm and 57.5 cm. The lens distortion values for the front layer were left the same as the dynamic prototype. The front layer k1 was 0.44 and k2 was 0.376. For the rear layer, the k1 was 0.525 and the k2 coefficient was 0.844. To calibrate this, one can print transparencies of crosses and try to make sure they are aligned and straight, without any distortion. If the layer distances change, then the lens distortion parameters should change as well.

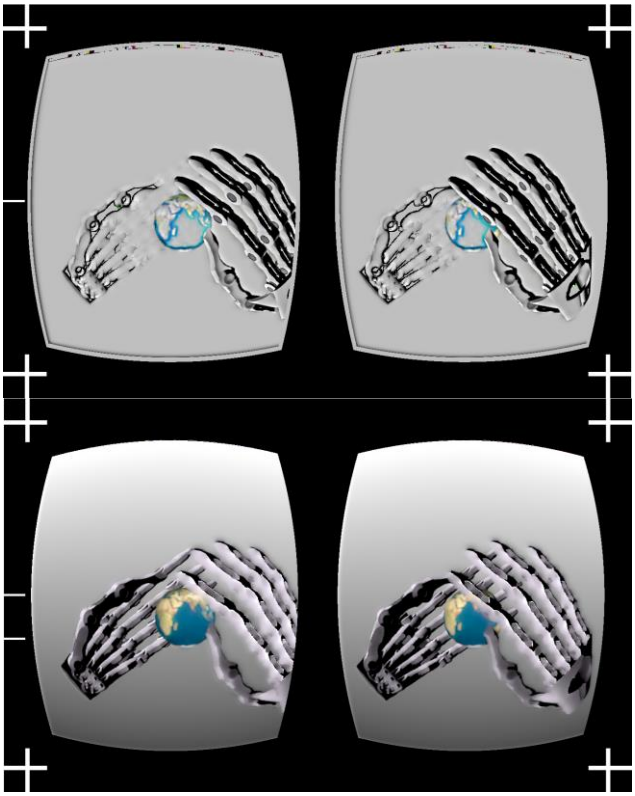

Figure 4. *The transparencies used for the scene. The top shows the transparency on the front layer and the bottom shows the transparency for the rear layer. Crosses are in each corner to aid in alignment, and the ticks on the left signify which side faces the lens and which layer is the front layer and which is the rear.*

4. Scene selection and evaluation

Many of the scenes in the dynamic prototype work very well. I personally found that the tree/bench scene and the scene with buildings composed of columns work extremely well. Other people like the chess scene or the scene with a robot and plane. Thus, user experience can vary from person to person even on a well-calibrated device.

On a static prototype, the user's head is free to move, making it harder to pinpoint the correct locations that the eyes should be in. Since the prototype is in the user's hands, it is also not on a stable platform. In other words, it is hard to maintain proper and exact alignment. It is also much more difficult to have exactly the right inter-pupillary distance (IPD) for the user. In the dynamic prototype, the eyes are always in a fixed location and the IPD can be adjusted. As a result, having the proper alignment of the two transparencies and the eyballs is an important characteristic to get perfect.

It turns out that the scene can make quite an impact on user experience. Some scenes provide a better sense of depth but are stricter towards the IPD of the user, whereas

some scenes have more tolerance towards different IPDs. Certain scenes such as the one with columns or with a tree and bench are also more difficult to fuse. For example, for a scene with robot hands and an earth in the middle, the scene worked fine for two people, but for three others, it did not work at all. They could not fuse the image and it was very uncomfortable. Simply by swapping the scene to the chess board without changing any settings such as IPD, all five users immediately found that the chess scene fused well. Unfortunately, the original two users that the scene (with robot hands and earth) had worked well for did not like the chess board scene as much, since the hands and earth felt like they had more depth. It really felt like the rear hand was behind the earth, which was behind the front hand. So, it was originally intended to have the robot hands and earth scene as the demo scene. Once I found that many people had difficulty fusing this, I decided to use the chess scene instead. The sense of depth may have felt a bit worse in the chess scene also because of the print quality. The resolution on the rear display seemed worse and some users thought that the background seemed blurred, but this was because the farther objects are smaller but still have the same DPI on the transparency. Therefore, each far chess piece has very little dots compared to the closer chess pieces as discussed later in the paper. Fixing this issue would most likely result in better focus cues.

The problem of fusing scenes and providing reliable focus cues would be a non-issue with proper IPD calibration for each user. A solution to this topic is proposed and discussed later in the paper. I found that the following three scenes worked the best for calibrated IPD on the static prototype: robot hands and earth, panther, and chess board. For non-calibrated IPD, the chess scene works well. Again, the user experience also depends person-to-person. No quantitative measurements were taken.

5. Future work and recommended modifications

I recommend many modifications to this prototype to make it more immersive, practical, robust, and durable. The first modification is to use a different housing, such as a larger Google Cardboard or to build a larger Google Cardboard from scratch. To build a Google Cardboard from scratch, purchase a cardboard sheet or box and enlarge the Cardboard template available on the Google website. The larger housing allows for more room for electronics but, more importantly, a more secure insert for the large 50 mm lenses. Of the two options, it is better to build a Google Cardboard from scratch. This is due to the fact that the lens section of the Google Cardboard consists of three layers of cardboard strongly glued together, making it very difficult to cut through and modify even with a sharp box cutter. Therefore, the next prototype should be a custom-built Google Cardboard. In particular, the lens inserts should compose of at least three cardboard sheets stacked on top of each other. The middle layer should be as wide as the diameter of the lens, whereas the outer cardboard layers should be smaller than the diameter of the lens to hold the lenses in place. The structure should be glued together securely such as with epoxy. This should fix the lenses so that they do not have much room for movement. Since the lenses are quite thick, it may be wise to use more than three layers of cardboard. Note that the black lens holder should be removed from the lens when inserting it into the custom-built Google Cardboard.

The next recommendation I have is to use higher quality prints. The image quality is limited by the DPI of the printed transparency. When magnified through the aspheric lenses, these dots become quite noticeable. Moreover, images that are at a distant virtual distance have the same resolution as objects that are at a very close virtual distance. This makes it appear like the nearby objects are very crisp and clear and the far away objects are blurred. Furthermore, it makes it more difficult to focus far away, depending on scene. Therefore, having high quality prints is crucial to having an immersive user experience. I used an HP 8600 inkjet printer printed at 1200 DPI, but I found that the 1200 DPI prints did not look very different from the 600 DPI prints, which may have been a result of the transparencies themselves. The results varied depending on the scene, but none of the scenes looked as sharp as on the dynamic prototype. This could have been due to a variety of issues: dust, dirt, and smudges on the acrylic sheets, improper alignment of transparencies and/or user's eyeballs (vs. the dynamic prototype which fixes the user's head to a secure position after alignment), and image DPI. In the chess scene for example, the rear layer did not look clear, most likely due to the limited DPI of the printer on distant objects, resulting in fewer dots per far-away chess piece compared to the nearby chess pieces. I would stay away from 600 DPI prints if possible and use an inkjet printer.

A third note is regarding the backlight. The size of the backlight depends on the size of the (possibly custom-designed) Google Cardboard. For my prototype, I used a standard Google Cardboard, so a 5" display that spans roughly 12 cm x 7 cm was ideal for me. In purchasing an LCD, it is important to ensure that the bezel of the display is not too thick so that the display uniformly lights up the entire transparency without sticking out awkwardly from the Cardboard. To reduce the number of batteries required to power the device, it may be wise to wire up the LEDs in parallel rather than in series. Many of the LCDs of this size require a backlight voltage of about 19.2V if wired in series, but only about 3V if wired in parallel. Unfortunately, the LEDs in LCD backlights are (in my experience) always wired in series, so this requires soldering in some very, very constrained areas. The LEDs

are usually located on a narrow circuit board only a couple millimeters wide, and there is not much room in the LCD/LED housing for extra wires to be added in to wire the LEDs in parallel. So, the form factor may change slightly once the LEDs are wired in parallel.

5.1. Alignment

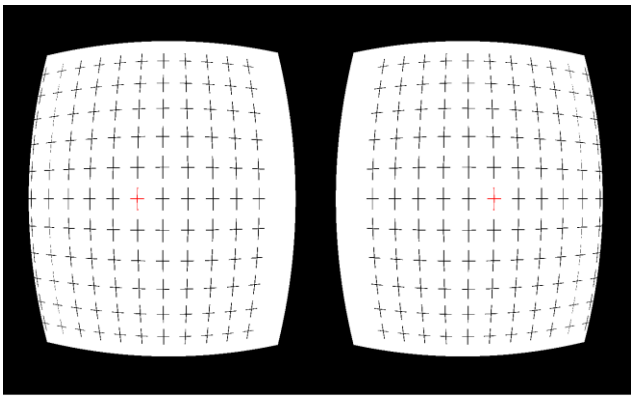

Figure 5. *Crosses on transparency used for calibrating the lens distortion.*

Alignment is a key issue that needs to be perfect, otherwise the display will not work at all. The transparencies included crosses and notches to help with alignment. They were taped to the acrylic sheets, which in turn were then taped together to make sure there was no movement once they were aligned. However, it is also important to account for different IPDs. With a Google Cardboard, this can easily be done by swapping out the scene. For example, the IPD can be measured using a ruler, and the scene for the corresponding IPD can be inserted into the Google Cardboard. Or, rather than measuring the IPD of the user, the user can select from a number of cross scenes, each of which displays a set of crosses for different IPDs. Once the user has found a scene that shows aligned crosses, he or she can select a different scene corresponding to the same IPD to get an aligned light field. In other words, the Google Cardboard allows for easy swapping of scenes and if there is a set of crosses for different IPDs and a set of scenes for different IPDs, then then the IPD calibration process should be fairly straightforward.

6. Conclusion

We have constructed a HMD that displays static scenes, is light weight, and portable. It is able to solve the vergence-accommodation conflict and provide focus cues to the user. This allows for prolonged use of HMD in the consumer market and is an alternative to other expensive, bulky solutions. However, there is still much room for improvement including higher DPI prints to allow for better user experiences and better hardware such as housing and backlighting to make the device more robust. However, this is a significant first step towards making a refined product which provides focus cues and ultimately gives a comfortable viewing experience.

References

[1] Benton, S., and Bove, V. 2006. *Holographic Imaging.* John Wiley and Sons.

[2] Favalora, G. E. 2005. Volumetric 3D displays and application infrastructure. *IEEE Computer 38*, 37-44.

[3] Hoffman, D. M., and Banks, M. S. 2010. Focus information is used to interpret binocular images. *Journal of Vision 10*, 5, 13.

[4] Huang, F.C., Chen, K., Wetzstein, G. The light field stereoscope.

[5] Lanman, D., Hirsch, M., Kim, Y., and Raskar, R. 2010 Content-adaptive parallax barriers: Optimizing dual-layer 3D displays using low-rank light field factorization. *ACM Trans. Graph. (SIGGRAPH Asia) 29*, 163:1-163:10.

[6] Liu, S., Cheng, D., and Hua, H. 2008. An optical see-through head mounted display with addressable focal planes. In *Proc. Ismar*, 33-42.

[7] Love, G. D., Hoffman, D. M., Hands, P. J., Gao, J., Kirby, A. K., and Banks, M. S. 2009. High-speed switchable lens enables the development of a volumetric stereoscopic display. *OSA Optics Express 17*, 18, 15716-15725.

[8] MacKenzie, K. J., Hoffman, D. M., and Watt, S. J. 2010. Accommodation to multiple focal plane displays: Implications for improving stereoscopic displays and for accommodation control. *Journal of Vision 10*, 8.

[9] Maimone, A., Wetzstein, G., Hirsch, M., Lanman, D., Raskar, R., and Fuchs, H. 2013. Focus 3d: Compressive accommodation display. *ACM Trans. Graph* 32, 5, 153:1-153:13.

[10] Rothbaum, B., Hodges, L., Ready, D., Graap, K., and Alarcon, R. 2001. Virtual reality exposure therapy for Vietnam veterans with post-traumatic stress disorder. *Ann Surg 62*, 8, 617-22.

[11] Seymour, N., Gallagher, A., Roman, S., O'Brien, M., Bansal, V., Andersen, D., and Satava, R. 2002. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg 236*, 4, 458-63.

[12] Watt, S., Akeley, K., Ernst, M., and Banks, M. 2005. Focus cues affect perceived depth. *Journal of Vision 5*, 10, 834-862

[13] Wetzstein, G., Lanman, D., Heidrich, W., and Raskar, R. 2011. Layered 3D: Tomographic image synthesis for attenuation-based light field and high dynamic range displays. *ACM Trans. Graph. (SIGGRAPH) 30*, 1-11.

[14] Wetzstein, G., Lanman, D., Hirsch, M., and Raskar, R. 2012. Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph. (SIGGRAPH 31*, 1-11.