DIY Head Mounted Displays Moving Towards Augmented Reality

Charles Han, Candice Murray and Arushi Arora
EE 367: Computational Imaging and Displays
Professor Gordon Wetzstein
Stanford University

Abstract

The field of Augmented Reality (AR) is growing rapidly. Many of the options available now are very expensive, which is limiting to development of applications of the technology. We have designed and developed several open-source do-it-yourself (DIY) head-mounted displays for augmented reality (AR) that cost between $4 to $40. Optical design and simulations were completed using Zemax for a design with a lens and a flat beam splitter (on-axis model) and for a single curved optical component (off-axis model). One demonstration of the on-axis model and two variations of the off-axis model were made. Each device aims to display the data from a projection source while at the same time allowing the user to see the real world around him/her with minimal obstruction. The effect is similar to Google Glass, but with some limitations to make it affordable and easily made by a wide range of potential users. The three main components in our design are: 1) a smart-phone screen as the projection source, 2) a curved clear plastic as the beam splitter and 3) cardboard as the body of the system.

1. Introduction

1.1. Motivation

Head-mounted displays (HMD's) have been heralded as the next generation in wearables. With the excitement surrounding such technologies as Oculus Rift, Microsoft Hololens and Google Glass in the past few years, many developers are starting to work in Augmented Reality (AR) and Virtual Reality (VR). One challenge in developing new applications and software for these platforms is the cost of the devices needed to generate the AR or VR experience. On the VR side, Google Cardboard has provided a low-cost, open source platform that can be easily obtained and modified. The AR field does not yet have an equivalent option. Our goal is to design an AR platform that can be made without any special equipment and with easily obtainable, low-cost materials. To further reduce cost, the design will use a smartphone screen as the illumination source.

1.2. Related Work

Ivan Sutherland is widely acknowledged as the first developer of AR. In his 1968 paper[8] he introduced the concept of overlaying information with the real world using optics and displays. The first practical AR-type systems were Heads-up Displays (HUD's), most commonly used in manufacturing or military environments. HUD's were display systems mounted inside helmets or other similar headgear which allowed the user to overlay digital information with their real-world surroundings. When first designed, the systems were very bulky and often tied to large computer systems, and so were mostly used in stationary settings, such as training [4][5] or manufacturing applications [2]. As the technology improved and became more mobile, people began to develop for applications in consumer electronics in addition to the previous training and manufacturing applications. In consumer electronics, the price point is generally very different than that for military applications or use by large, well-established companies. In order for the technology to be practical in an economical sense as well as desirable by consumers, it needed to be miniaturized and simplified. Simplification has generally taken two forms: simplifying the optical system and simplifying the image source.

In original HUD's, the projection source was generally a CRT controlled by a computer. As technology progressed, CRT's were replaced by digital projectors and micro-projectors, which give a much better image quality and can have much smaller form-factors and power consumption. This allowed the systems to shrink from large helmet-size packages to glasses-size packages [3][7]. Another solution that is being used to decrease overall cost is the use of a smartphone display instead of a special projecting system [11].

Examples of current technologies using projector-type systems are Microsoft Hololens, Google Glass, and Magic Leap. Generally, the projector setup allows better image quality and higher control over image properties. Examples of technologies currently using smartphone displays are the Seer (designed by Wei et.al.) and Metavision's Meta 2. While getting good image quality from these systems can be more difficult, the cost is significantly reduced ($120 for Seer, $950 for Meta 2 vs. $3000 for Hololens, $1500 for Google Glass).

The optical system is the other main component of HMD's. Many of the original generation of HUD's had complicated optical systems with many components in order to get the image the right size and distance for the user to view comfortably [10]. Since optical components make systems expensive, bulky and less robust, systems with fewer optical components are preferred. Two of the most common designs now are a combination of a single lens and a partially reflective surface [1], or a single curved optical surface [1][7][11]. The design parameters used in making the curved optical surfaces are in most cases proprietary information. In [1] equations are given for a general parabolic reflective surface, and in [7] it mentions that the curved optical surfaces are not rotationally symmetric in any Cartesian dimension, and gives some equations for calculating angular definition and surface normals of the optical component used.

2. Design Considerations

The main goal for our design is to have the system be very low-cost and require as little special equipment to create as possible. These two characteristics will enable open-source development and ready access by any interested parties. This is a similar goal to that of Google Cardboard, which allows users to experience virtual reality for a very low-cost, but isn't as high quality of an experience as higher-end systems such as Oculus.

To achieve this goal, the system should project a virtual image of a smartphone screen at a distance of around two meters in front of the viewer. The image should be magnified to fit well with the real images in the users view at that distance. A wide field of view (FOV) is highly desirable, as is support of stereo cues. The system should have as few components as possible, and be easily assembled without extensive prior knowledge or training. Cost of the system should also be as low as possible.

When using optical components such as lenses it is also important to recognize that in a system the diameter of the lens is like the diameter of an aperture, since any rays passing outside of the lens will not be bent appropriately to magnify and direct the image. This puts a limitation on what can be done using standard lenses, since when using a phone screen as the illumination the source size is at least 2"x3".

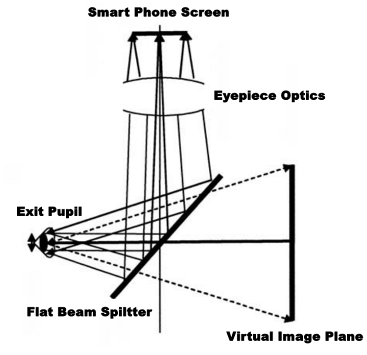


Figure 1: On-Axis HMD optical system.

3. Theoretical Modeling

Different from Google Cardboard or any other displays on the VR side, the AR HMD display is a see-through optical system. A critical component of the see-through optical system is the beam splitter, which is located in front of the eyes and enables the superposition of graphical and real-world information. In this project, we will discuss two basic theoretical models: 1) the On-Axis Model based on a flat beam splitter and 2) the Off-Axis Model based on a curved beam splitter.

3.1. On-Axis Model

The on-axis Model is similar to the folded version of the VR optical system in Google Cardboard. It consists of a display (smart phone screen), eye-piece optics, and a flat beam splitter (see Fig 1). The main advantages of the on-axis model are its simplicity and the high transparency. The spacings between the optical elements can be easily obtained by applying the Lens Maker's equation. The minimum required eye relief and the large size of the beam splitter set by the range of interocular distance impose the maximum FOV in theory of about 40° [6]. In practice as shown in Section 4, the FOV is below 30° limited by the lens size. This is one of the main drawbacks of the on-axis model. Another problem is the existence of ghost image created by the second surface of the flat beam splitter due to birefringence of the plastic. These two serious drawbacks lead us to consider the off-axis design.

3.2. Off-Axis Model

The off-axis model uses only a single, tilted component beam splitter. Its increased optical power compared to the flat beam splitter system allows wider FOV for the same effective eye relief. Adding power to a tilted beam splitter however introduces severe optical aberrations, such as coma astigmatism and asymmetric distortion [6]. There are two common approaches to realize the curved beam splitter. One is to fabricate a decentered prism optical system [9] and one is to use a free-form off-axial mirror [6]. Given that in an AR display system the image distortion can be compensated by algorithm computation, we use the easier approach of a free-form polynomial surface of the beam splitter. The profile of the x-y polynomial surface was described by the following equations [1]:

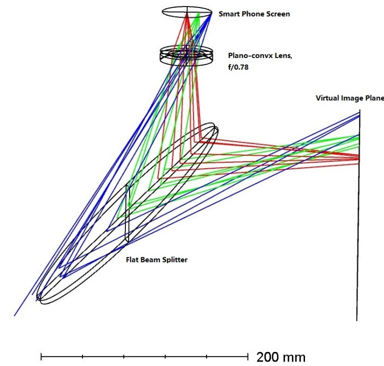
$$P(x,y) = \frac{cr^2}{1+\sqrt{1-(1+k)c^2r^2}} + \sum_{j=2}^{10} C_j x^m y^n \quad (1)$$

$$j = \frac{(m+n)^2 + m + 2n}{2} + 1 \quad (2)$$

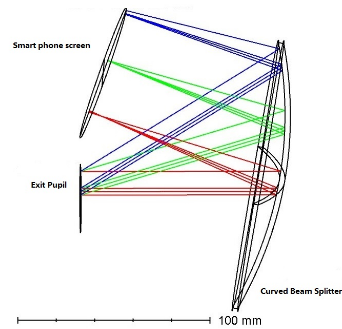
where c is the vertex curvature, r is the radial coordinate in lens units, $r^2 = x^2 + y^2$, k is the conic constant, and $C_j$are the coefficients of the various xy terms. The conic constant is less than -1 for hyperbolas, -1 for parabolas, between -1 and 0 for ellipses, 0 for spheres, and greater than 0 for oblate ellipsoids. The focal length of the mirror is half the radius of curvature.

3.3. Zemax Simulation

We build up the on-axis and off-axis theoretical models in Zemax OpticStudio 15.5 SP2. For the on-axis model shown in Fig. 2a, in order to be consistent with the actual prototype we made, we use a plano-convex lens with f/0.78. The spacing between smart phone screen and the plano-convex lens is 38.5 mm. The flat beam splitter is 100 mm from the lens. The virtual image plane is 225 mm in front of the beam splitter. The radius of curvature and spacings are optimized for the minimum spot size radius. For the off-axis model, the optical layout is shown in Fig. 2b. It consists of an 8-mm exit pupil, a free form mirror and the smart phone screen. We use a parabolic polynomial surface for the curved mirror where k = -1. The radius curvature of the curved mirror is 200 mm. The mirror is tilted at 10° in the x-z plane with respect to the optical axis of the human eye. The phone screen is also tilted at this angle to provide the least distorted image. The spacings are optimized for the minimum spot size radius. In both simulations we set the FOV to 32° to compare their image quality. Fig. 3 shows the point spread functions (PSF) and modulation transfer functions (MTF) of the two models. The result shows that the PSF of the on-axis model has asymmetry in the x-y plane while the PSF of the off-axis model is more symmetric and sharper; the on-axis model has a cut-off frequency of 60 cycles/mm while the cut-off frequency of the off-axis model is 100 cycles/mm. Therefore, the off-axis model performs with an overall better image quality than the on-axis model.



(a) The on-axis model consists of a smart phone screen, a plano-convex lens and a flat beam splitter



(b) The off-axis model consists of a smart phone screen and single parabolic curved beam splitter

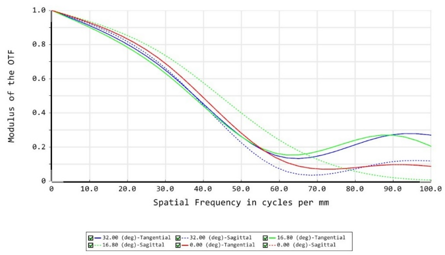
Figure 2: The on-axis and off-axis model layouts in Zemax

4. Implementation

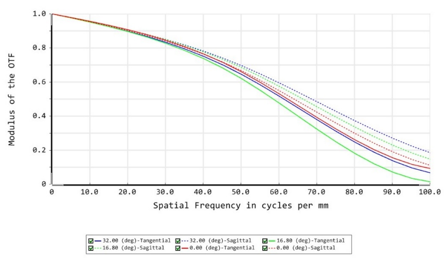
The implementations for the above designs were done using low-cost materials such as cardboard to make the holders and warped acrylic sheets for the curved visor. Three demonstrations are shown below: 1) a lens and flat beam splitter, 2) curved visor and the 3) stereoscopic curved visor. The first demonstration is based on the on-axis model. The second and third demonstration attempt to demonstrate the off-axis model. To accurately create the polynomial surfaces in the visor, fabrication techniques such as vacuum forming are required. Due to limitations of time and resources, we decided to form visors by warping plastic in a 1) spherical and 2) a abstract free-form shape. The spherical shape was chosen since the mathematical equations of conic sections are easily solved and manipulated. These two designs have been implemented in demonstrations 2 and 3.

4.1. Demo 1: Lens and Flat Beam Splitter

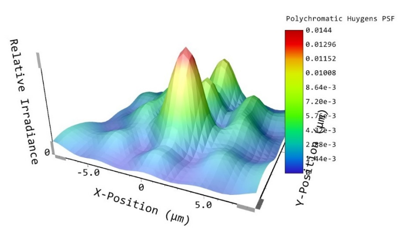
This design required three main components: a phone, a condensing lens, a beam splitter and a cardboard holder to



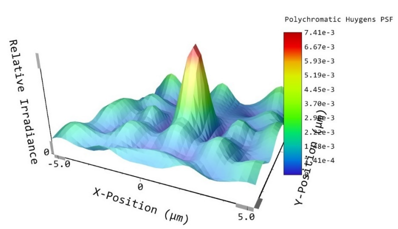
(a) On-Axis Model MTF



(b) Off-Axis Model MTF



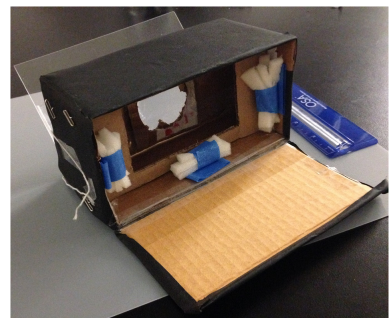
(c) On-Axis Model PSF



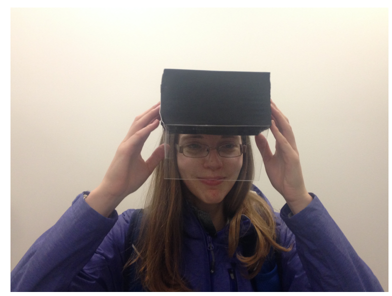
(d) Off-Axis Model PSF

Figure 3: The point spread function (PSF) and modulation transfer function (MTF) of the on-axis and off-axis models in Zemax

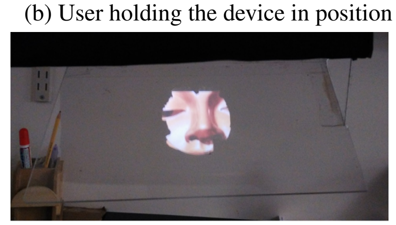
hold the three components (see Fig. 4). The phone to be used as the display source need not be a smart phone in this demonstration (demo) since only a 2-D image of the screen is to be displayed. This demo was created for an iPhone 5c which has a 4 inch (diagonal) screen. The lens was an uncoated off-the-shelf circular plano-convex lens with a focal length of $\sim$4 cm on its spherical side, a f/0.78 and a thickness of 2 - 3 mm. An acrylic sheet with 1.52 mm inch thickness was used as the beam splitter to give $\sim$10% reflection of incident rays perpendicular to the sheet. A cardboard box (16 cm x 8 cm x 8 cm) was constructed to hold two shelves; one each for the phone and the lens. The beam splitter was hinged to the end of the box by clear tape and a short string was used to hold the beam splitter at $\sim 51^\circ$ with respect to the horizontal plane.



(a) Cardboard box containing phone shelf, lens holder encapsulating a lens and a flat plastic beam splitter hinged to the side



(b) User holding the device in position



(c) Image (of a face) displayed by the phone as seen through the beam splitter

Figure 4: Demonstration 1 involving lens and flat beam splitter

The phone and the condensing lens were held perpendicular a little above the eyebox. The phone is kept a distance of 3.5 cm (smaller than the focal length) away from the lens thus creating a magnified virtual image on the same side of the lens. The diverging rays that escape from the other side of the lens are partially reflected by the beam splitter that is held at an appropriate angle to the eye. The eye looking into the beam splitter then sees an inverted image of the phone (at some distance in front of it and behind the beam splitter) overlaid over the view of the real world from transmission through the beam splitter. The image was measured to be approximately 10 cm away from lens (5 cm from lens to beam splitter and 5 cm from beam splitter to image). The horizontal FOV is then measured to be 28°. The device does not support occlusion at this point and so the image is much like that of Pepper's ghost.

The cost of the various components is as follows: the lens costs \$35 (varies depending on focal length and coating), the acrylic sheet, cardboard, tape and pins combined cost ∼\$4. The total cost of the setup is ∼\$40.

4.2. Demo 2: Curved Visor

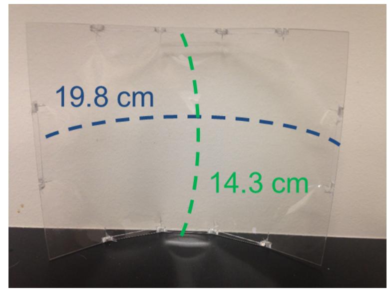
This demonstration uses a 0.51 mm thick acrylic sheet that is warped to give a visor with a radius of curvature of 20 cm (see Fig. 5). On applying simple conic sections, it was found that to warp a sheet of flat plastic into a radius of curvature of 20 cm both vertically and horizontally, with final dimensions of 19 cm by 14 cm the arc lengths must be 19.8 cm x 14.3 cm. Thus, by making multiple cuts along the edges of the sheet and sewing them together, a visor with those approximate dimensions was created. These final dimensions are chosen to accommodate the smartphone screen size plus some extra space so that the image is not overlaid on areas warped by the cuts. The warping generally occurs over about 1.5 times the length of the cut. A phone holder made from cardboard was then attached to the top center of the the sheet. An inverted image of the phone screen could then be seen in the visor. The visor was hinged from the top in such a way that it could be adjusted by each user to get the best image.

The best image was found to occur ∼18 cm away from the eye box and the image measured ∼12 cm horizontally. The horizontal FOV therefore for this system was larger and equal to 36.8°. Due to limitations of time (2 weeks) and resources, the curvature formed using this procedure was not quite perfect and the image suffered from aberrations especially towards the edges. It is expected that a visor formed using a mould or 3D printed with more accuracy would perform better and come closer to the Zemax modeling showed in the previous section.

The cost for this design consisted of a plastic sheet worth less than \$2.50 and a needle, thread, cardboard, tape, pins costing less than \$2. The total cost therefore was ∼\$4.

4.3. Demo 3: Stereoscopic Curved Visor

The stereoscopic curved visor was formed by modifying a Google Cardboard in two respects: 1) the two small lenses in the Google Cardboard were replaced by a large curved visor made from 0.51 mm thick acrylic sheet and, 2) the phone was placed at the top of the system instead of in front of the eyes as was its original position (see Fig. 6). The Google Cardboard design was used primarily to take advantage of the setup with individual eye boxes and a separator between them. In this demo, it was important to use images that supported stereo cues. By using the Google cardboard app on the phone we were able to see a stereoscopic image on the curved visor. The image was found to occur ∼38 cm away from the eye box and the image measured ∼18 cm horizontally. The horizontal FOV therefore for this system was equal to 26.6°. The total cost of this system taking into account the cardboard and the sheet was also around ∼\$4.

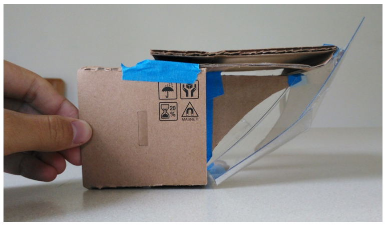


(a) Curved visor made from 0.51 mm thick acrylic sheet with dimensions of curvature shown as an overlay.

(b) The acrylic sheet is attached to a holder for the phone. Figure shows user holding the setup in position

(c) Image from the phone as seen on the visor

Figure 5: Demonstration 2 involving curved visor



(a) The Google Cardboard box has been modified in two respects: 1) the lenses have been replaced by a curved visor made from 0.51 mm thick acrylic, 2) the phone is placed at the top perpendicular to its original position

(b) User holding the setup in position

(c) Stereoscopic image from the Google Cardboard app on the phone (top) as seen on the visor

Figure 6: Demonstration 3 showing stereoscopic image with curved visor

5. Conclusions and Future Work

We have demonstrated extremely low-cost simple on-axis and off-axis designs for head mounted displays involving a phone as a projection source. The on-axis model is based on a flat beam splitter and the off-axis model is based on a curved beam splitter. The minimum required eye relief and the large size of the beam splitter in the on-axis model impose the maximum FOV in theory of about 40° [6]. For the off-axis model, the optical layout consists of an exit pupil, a free-form curved mirror surface chosen to be described by a polynomial function and a phone screen. We then did the Zemax simulations for the above designs and found that the PSF of the off-axis model is more symmetric and sharp than that for the on-axis model and thus, the off-axis model shows a better image quality. We then implemented three different designs as a first-principles demonstration of the simulations. The first design involves a glass lens that magnifies the image from the phone and a flat beam splitter that reflects part of the image back into the eye. The second design uses a curved visor to create magnification and removes the need for the lens. The third design incorporates stereo cues in the curved visor. While these ideas have been around since the time of Ivan Sutherland [8], there are three differences between the previous designs and the ones we display here. The first key difference is the do-it-yourself nature of these designs. They use cardboard, pins, tape and plastic sheets to create simple designs that allow viewing of the virtual image. The second key difference is the cost. Except for demo 1 which requires a lens worth $35, all of the other components cost less than $5. Despite the low-cost nature of the raw materials, the device holds up exceedingly well and the image formed is quite good. This is unique since the other AR products on the market are significantly more expensive. Third, many of the head-mounted systems utilize LCDs as their projection source which are significantly smaller, easier to modulate and design optics for. In order to achieve the aim of a low-cost HMD, we decided to stick to the projection source of a phone which acts as an extended light source.

A true AR would be formed when the viewer can interact with the virtual display - through sight, sound, touch, smell etc. These heads up displays need to be augmented with at the very least head-tracking and gaze-tracking to enable use in manufacturing (adding rivets in airplane construction) to commercial (video games) applications. In addition, the quality of the user experience would be much better when the device supports occlusion - a feature that is a hot topic of research and yet to be seen implemented in commercial AR systems.

6. Acknowledgements

We would like to acknowledge the support and advice given by Prof. Gordon Wetzstein, Shikhar Shrestha and Robert Konrad.

References

[1] O. Cakmakci and J. Rolland. Design and fabrication of a dual-element off-axis near-eye optical magnifier. *Optics Letters*, 32(11):1363, 4 2007.

[2] T. Caudell and D. Mizell. Augmented reality: an application of heads-up display technology to manual manufacturing processes. In *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*, volume 2, pages 659–669. IEEE, 2 1992.

[3] M. Gribetz, S. Mann, and R. Lo. Extramissive spatial imaging digital eye glass apparatuses, methods and systems for virtual or augmediated vision, manipulation, creation, or interaction with objects, materials, or other entities, 7 2014.

[4] F. P. Heller and S. M. Ellis. Headgear with spherical semi-reflecting surface, 3 1978.

[5] W. Kraemer. Head mounted visual display, 12 1996.

[6] J. P. Rolland. Wide-angle, off-axis, see-through head-mounted display. *Optical Engineering*, 39(7):1760, 7 2000.

[7] D. A. Smith, G. A. Harrison, and G. E. Wiese. Head-mounted display apparatus employing one or more reflective optical surfaces, 1 2014.

[8] I. E. Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I on - AFIPS '68 (Fall, part I)*, page 757. ACM Press, 12 1968.

[9] T. Togino. Decentered prism optical system, 1997.

[10] D. van Krevelen and R. Poelman. A survey of augmented reality technologies, applications and limitations. *The International Journal of Virtual Reality*, 9(2):1 – 20, 2010.

[11] R. Wei. Head mounted perspective display device, 4 2015.