Wide Field-of-View Dual-Focal-Plane Augmented Reality Interactive Display with Gaze-Tracker

by

Uğur Yekta Başak

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Electrical and Electronics Engineering

KOÇ UNIVERSITY

January, 2019

Wide Field-of-View Dual-Focal-Plane Augmented Reality Interactive Display with Gaze-Tracker

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Uğur Yekta Başak

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

____
Prof. Hakan Urey

____
Prof. Yücel Yemez

____
Asst. Prof. Muhammed Fatih Toy

Date: ____________________

ABSTRACT

Stereoscopic displays have a fixed focus plane and they suffer from visual discomfort due to the mismatch between the focus and vergence of the eyes, known as the vergence-accommodation conflict (VAC). VAC is unavoidable in conventional stereoscopic head-mounted displays (HMDs). In this thesis research, I proposed a biocular (i.e, common optics for two eyes), two focal-plane based augmented reality (AR) system with real-time gaze tracker, which provides a novel interactive experience. Two separate liquid crystal displays (LCDs) are placed at slightly different distances to a Fresnel relay lens such that virtual images of LCDs appear at 25 cm and 50 cm to the user. Both LCDs are totally viewed by both eyes. While the system is limited to two depths and discontinuity occurs in the virtual scene, it provides correct focus cues and natural blur effect at the corresponding depths. This allows the user to distinguish virtual information through the accommodative response of the eye, even when the virtual objects overlap and partially occlude in the axial direction. Displays are driven by a single computer and the objects in the virtual scene are distributed over the LCDs according to their depths. Furthermore, a road scene simulation is realized as a convenient use-case of the proposed display so that a large monitor is used to create a background scene and the rendered content in the LCDs is augmented into the background. Field-of-view (FOV) is 60 x 36 degrees and the eye-box is larger than 100 mm, which is comfortable enough for two-eye viewing. The system includes a pupil and gaze tracker, which is implemented with a single camera and using computer vision algorithms. The gaze tracker, which is experimented on different users, is able to select the correct depth plane based on the shift in the interpupillary distance when the convergence angle of the users eyes changes. The rendered content can be distributed to both depth planes and the background scene simultaneously. Thus, the

user can select and interact with the content at the correct depth in a natural and comfortable way. The prototype system can be used in tasks that demand wide FOV and multiple focal planes and as an AR and vision research tool.

ÖZETÇE

Stereoskopik ekranlar sabit bir odak düzlemine sahiptir ve uyum-yakınsama çatışması (UYÇ) olarak da bilinen gözün odak ve hareketlerinden kaynaklanan uyumsuzluktan dolayı görsel rahatsızlıklardan mustariptir. UYÇ olgusu, geleneksel yöntemlerle hazırlanmış kafaya-takılabilen ekranlarda (KTE) kaçınılmazdır. Bu tezde; gerçek zamanlı bakış izleyicisi ile donatılmış ve yeni bir deneyim sunan bi-oküler (iki göz için ortak optik elemanlar), iki odak-düzlemli bir artırılmış gerçeklik (AG) ekranı sunulmaktadır. Iki ayrı sıvı kristal ekran (SKE) Fresnel lens ile farklı mesafelere yerleştirilerek sanal görüntülerin kullanıcıya göre 25 santimetre ve 50 santimetrede oluşması sağlanmıştır. Iki SKE de iki göz ile tamamen görülebilmektedir. Sistem iki derinlik ile sınırlandırılmış olmasına ve bu durum sanal sahnede bir süreksizliğe yol açmasına rağmen, bahsi geçen derinliklerde doğru odak kanıtlarını ve doğal bulanıklık etkilerini sunmaktadır. Bu durum, farklı derinliklerde gösterilen sanal bilgilerin aynı görsel eksende olsalar bile kullanıcı tarafından ayırt edilebilmesini sağlamaktadır. SKEler sadece bir bilgisayar ile sürülebilmektedir ve sanal sahnedeki objeler derinliklerine göre bu ekranlar üzerine dağıtılmaktadır. Ayrıca, bir yol ve sürüş simülasyonu da uygun bir kullanım alanı olarak sunulmaktadır. Öyleki; büyük bir monitör, arka plan sahnesini yaratmak için kullanılmış ve SKElerdeki içerikler bu arka plan üzerine düşürülmüştür. Görüş alanı (GA), 60 x 36 derecedir ve kullanıcı pozisyonunun değişimine rağmen sanal sahnenin görüntülenebilmesini sağlayan alan 100 milimetreden fazladır. Böylece, iki göz için konforlu bir sistem elde edilmiştir. Sistem, iki gözü de görüntüleyebilen tek bir kamera ve görüntü işleme teknikleri ile uygulanmış bir bakış izleyicisi içerir. Farklı kullanıcılar ile test edilmiş bu bakış izleyicisi, göz bebeği merkezleri arasındaki mesafeyi hesaplayarak kullanıcının baktığı derinlik düzlemini herhangi bir kızıl-ötesi aydınlatmaya ihtiyaç duymadan hesaplaya-

bilmektedir. Oluşturulan içerik, iki farklı derinlik düzlemine ve arka plana eş zamanlı olarak dağıtılmaktadır. Böylece, kullanıcı gözlerini kullanarak farklı derinlik düzlemlerindeki içerik ile etkileşime konforlu bir şekilde girebilmektedir. Sunulan sistem prototipi, geniş görüş alanı ve çoklu odak düzlemi gerektiren uygulamalarda veya bir AG araştırma aracı olarak kullanılabilir.

ACKNOWLEDGMENTS

My sincere appreciations go to my supervisor, Prof. Hakan Urey, for providing me to a chance to be a member of Optical Microsystems Laboratory (OML) at Koç University and teaching me a lot in different fields patiently during my master studies. Everything would be different without his trust when I was feeling desperate.

A very special thanks go to Seyedmahdi Kazempourradi and Erdem Ulusoy for their great effort and creative ideas related to my thesis.

I would also like to thank my colleagues Cemalettin Yılmaz, Koray Kavaklı, Kıvanç Hedili and Burak Soner for having a key role in my projects and sharing their knowledge with me.

I would like to thank European Research Council and Graduate School of Science and Engineering institution of Koç University for their financial support as well.

Finally, I am grateful to my whole family for all the love and support I received over the years.

TABLE OF CONTENTS

LIST OF FIGURES

NOMENCLATURE

3D: Three dimensional

AR: Augmented reality

CGH: Computer generated hologram

DOF: Depth of field

EI: Elemental image

EPD: Entrance pupil diameter

FOV: Field-of-view

HMD: Head-mounted display

HNED: Holographic near-eye display

HOE: Holographic optical element

IPD: Inter-pupillary distance

LCD: Liquid crystal display

MLA: Micro-lens array

PSF: Point spread function

ROI: Region of interest

SLM: Spatial light modulator

VAC: Vergence-accommodation conflict

Chapter 1

INTRODUCTION

This dissertation explores existing augmented reality display technologies and presents a novel dual-focal-plane gaze-contingent interactive augmented reality display system
.

The following chapter firstly introduces the fundamental principles of the human visual system and some frequently used terms related to the human visual system for better understanding of the proposed research project. After the brief introduction to the human eye, brain and depth perception phenomenon; an overview of different methods for mitigating visual discomfort in conventional head-mounted display systems are presented. The basic concepts and major types of augmented reality display systems are discussed with their advantages and drawbacks. Finally, the contribution of the thesis to the literature is stated.

Chapters 2 and 3 present design and optical simulations of proposed display together with gaze tracking system principles, interaction details, and experimental results.

1.1 Human Visual System

The human visual system is a complex mechanism consisting of two main parts: the eye and the brain. While the eye functions as a complicated sensory system that provides sight of surroundings, the corresponding part of the brain is responsible for processing this visual input.

What the eye and the brain perceive as a scene is the combination of light rays emitted from light sources and light reflected by real objects. When a light ray hits the eye, it passes through the cornea and the other parts of the eye before reaching the retina.

If the light rays are within the range of the visible electromagnetic spectrum (300-700 nm), a healthy eye conveys the signals to the responsible part of the brain via optical nerves as shown in Figure 1.1 and the brain manages the perception of the scene. A typical pair of eyes are separated from each other with inter-pupillary distance (IPD) that varies between 50 - 75 mm with a mean of 63 mm [1].

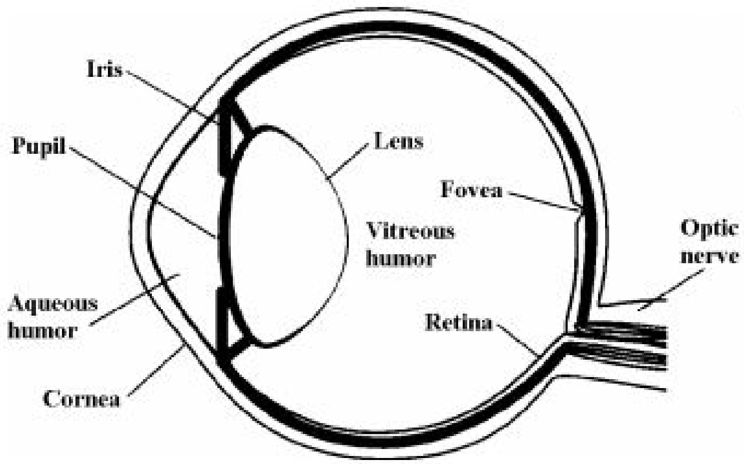


Figure 1.1: Basic anatomy of the human eye

1.1.1 Terminology

To understand the working principles of the distinct display systems and three-dimensional (3D) perception, some frequently used terms in the human visual system must be understood for the proposed thesis.

- **Accommodation:** Accommodation is the focus response of the eye that determines the focal distance of the eye by changing the shape of the eye's lens. When the eyes focused in a far distance or near distance, the muscles of the eye's lens relaxes or contracts respectively. The amount of the blur in the retina drives this accommodation action of the eye to create a blur-free image.

- **Vergence:** Vergence is the rotation of the eyeballs that creates an angular difference between the visual axis of the eyes. Vergence action of the eye is driven by the distance to the object of interest so that the visual axis of the eyes intersect at the fixation distance.

- **Iris and pupil:** The iris is a thin, colorful circular area in the eye that controls the amount of the light reaching the retina by changing the diameter of the pupil which is the opening in the center of the iris. The average size of the human iris is 11.8 mm and the pupil aperture of the human eye varies between 2-8 mm under different circumstances.

- **Visual acuity and resolution:** Visual acuity is a measure of the spatial resolution of the eye and it can be considered as the clarity of vision. Rayleigh resolution criterion states that the resolution limit of a system depends on the system's aperture stop diameter. Thus, the angular resolution of a system can be calculated as:

$$\sin\theta = 1.22(\lambda/D)\,, \tag{1.1}$$

  where λ is the wavelength of the light and D is the aperture stop size.

  Thus; the angular resolution of a human eye is $\theta = 0.295\,\text{mrad} = 0.0169^\circ$ when the diameter of the pupil is assumed as 2.2 mm and the wavelength is equal to 532 nm. Therefore, a human eye can resolve approximately 60 pixels/degree = 30 cycles/degree where a cycle consists of 2 pixels.

- **Point spread function:** Point spread function (PSF) can be described as the response of an imaging system to a point object. From a mathematical point of view, it is the impulse response of the imaging system. PSF is a good indicator of the degree of blurring of a point object and quality of an imaging system. In addition, it provides information about other types of aberrations such as higher-order aberrations, astigmatism, and chromatic aberration. When PSF at one wavelength is considered, it is proportional to the square of the modulus of the Fourier transform of the complex aperture function [2] as shown in Equation 1.2. Once PSF of a system is known, the image of the scene in the retina can be calculated via convolution and appears as an Airy disc for a point source in the retina of an aberration-free eye as shown in Figure 1.2.

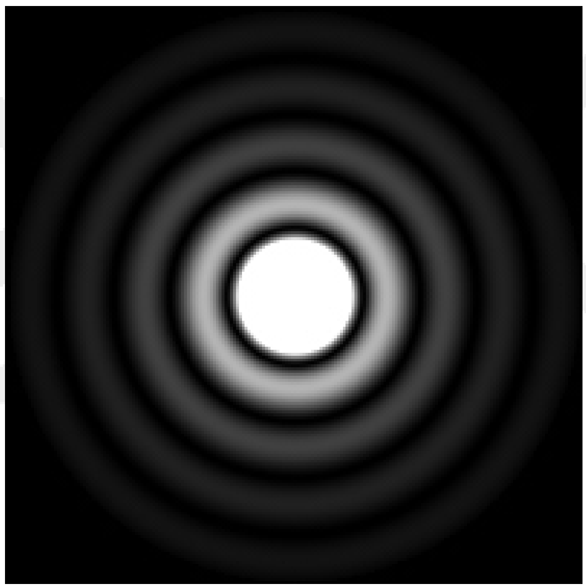


Figure 1.2: The image of a point source is an airy disc on the retina of a human with an aberration-free eye

$$PSF_{\lambda}(\theta, \phi) \propto \left| \mathcal{F}(Ae^{\frac{2\pi i}{\lambda}(Z_d + Z_{HOA} + Z_{LCA})}) \right|^2 , \quad (1.2)$$

where λ is wavelength, A refers to the transmittance of the pupil and Z defines the wavefront aberration which is separated with subscripts as defocus (d), higher-order aberrations (HOA) and longitudinal chromatic aberration (LCA) [3].

- **Field-of-view:** Field-of-view (FOV) is the area that a person can see through his/her eyes or through an optical device. A healthy person has approximately

190° horizontal monocular, 120° horizontal binocular and 135° vertical FOV as illustrated in Figure 1.3 [4].

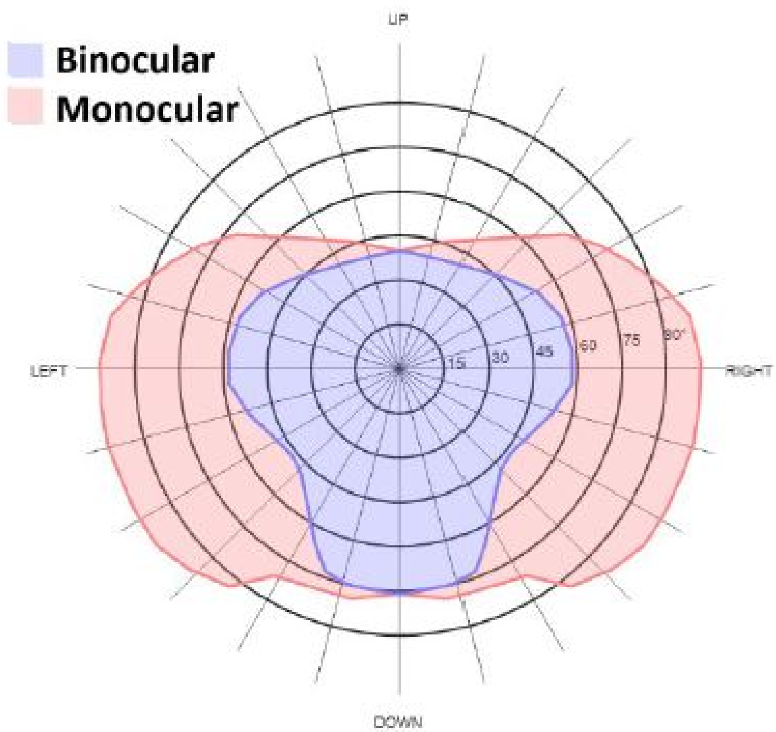


Figure 1.3: The field-of-view of human visual system [4]

1.1.2 Depth perception

The human visual system involves an extended range of depth cues. These depth cues can be classified into two main categories: psychological and physiological cues. Psychological cues include perspective occlusion, correct 3D perspective, shadow and shading, apparent size and texture gradient; whereas the physiological cues include binocular disparity, motion parallax, retinal blur effect, accommodation as well as vergence [5]. Alternatively, it is also common to classify depth cues as monocular and binocular cues [6]. In this subsection, some important depth cues for the proposed thesis are explained in detail for better understanding in the following sections.

- **Binocular disparity:** Binocular disparity is the positional difference between the two retinal projection of a given point in space [7]. The reason behind this positional difference is that the world seen by each eye is slightly different due to the IPD. This positional (horizontal) difference is the primary source of depth extraction and 3D perception of the human brain.

- **Motion parallax:** Motion parallax is one of the most significant monocular cues for the depth perception in the human visual system. Closer objects in the scene appear to move faster than further objects due to the visual perspective. Thus, the user can perceive the depth of the objects according to their movement speed even with a single eye and that is why it is identified as a monocular cue.

- **Retinal blur effect:** When an object of interest at a particular distance is observed, it is imaged onto the retina sharply via accommodation action of the eye's lens. Nonetheless, parts of the scene that lie at different distances than accommodation distance cannot be focused on and as a result, a blur pattern in the retina occurs. Points which are further away from the fixation distance creates a greater degree of blurring effect. The point-spread function (PSF) diameter is a good indicator of the amount of defocus and it can be calculated as:

$$\beta = A\left|\frac{1}{z_0} - \frac{1}{z_1}\right| = A\left|\Delta D\right| \ , \tag{1.3}$$

  where β in radians, A is pupil diameter, z_0 is the distance to which the eye is focused, z_1 is the distance to the object that appears blurry and ΔD is the difference in those distances in diopters [8].

1.2 Stereoscopy and Vergence-Accommodation conflict

It is possible to fool the human eyes and the brain artificially in terms of 3D perception through a method called stereoscopy. In this method, a pair of parallax images for two eyes are rendered in flat displays so that the virtual image is created at different depths according to the amount of binocular disparity. While stereoscopy was firstly proposed by C. Wheatstone in 1852 [9], it continues to be the main idea behind the conventional stereoscopic head-mounted displays (HMD) which are commonly used in today's consumer applications.

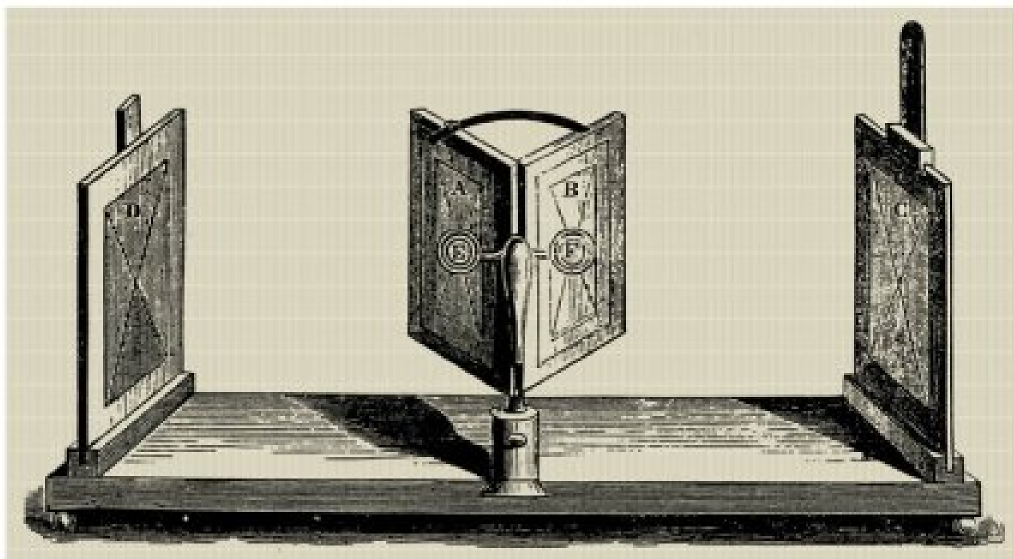


Figure 1.4: Wheatstone's mirror stereoscope (1838)

In a real-life scenario, eye-ball movements and the focus action of the eyes work together. The visual axes of the eyes intersect on the fixation distance and ciliary muscles of the eyes adjust the focus distance of the eye's crystalline lens to minimize the retinal blur simultaneously. Thus, accommodation and vergence are neurally coupled in natural vision [10]. This natural coupling is broken in conventional stereoscopic displays due to the mismatch between accommodation and convergence distance. The decoupling is called Vergence - Accommodation Conflict (VAC) as illustrated in Figure 1.5. Although accommodation depth is the distance between the observer and the display screen, the intersection of the visual axes of the eyes that stems from the eye-ball movement is at the appearing depth of the virtual object. Many recent studies show that VAC is one of the most significant causes of discomfort, visual fatigue and induced binocular stress in stereoscopic virtual and augmented reality displays [11, 12, 13, 14] and it might even lead to error in the perception of the scene geometry [15].

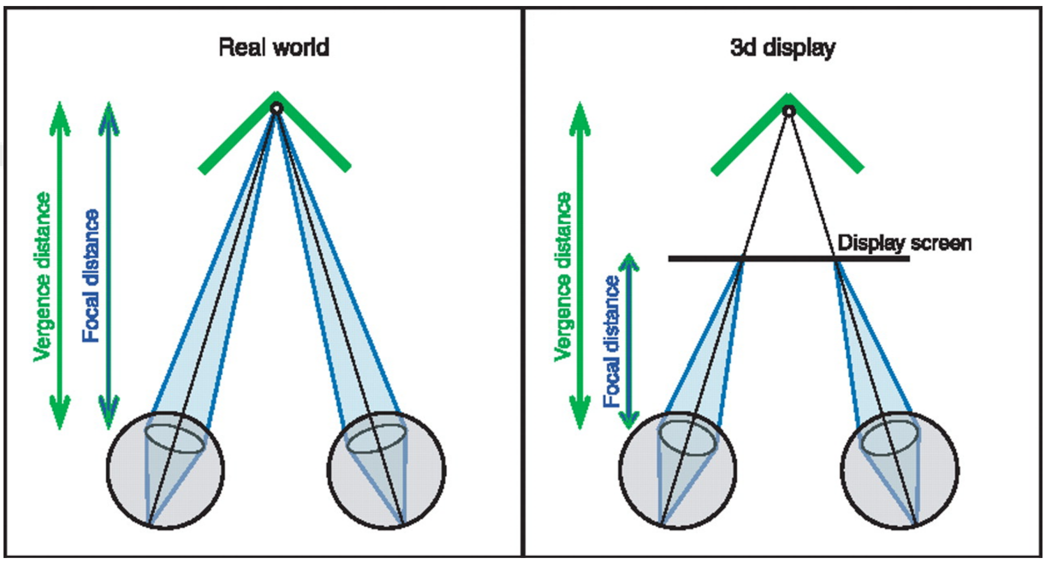


Figure 1.5: VAC in stereoscopic displays. While vergence and accomodation distance are matched in real life scenarios, decoupling occurs in stereoscopic 3D displays. [13]

1.3 Alleviating VAC in HMDs

As mentioned in the previous sections, VAC is a crucial problem in near-eye displays. Several solutions are provided in the literature to mitigate VAC problem in conventional HMDs. Mimicking natural vision requires rendering correct focus cues together with the natural blurring of images [16]. Hong Hua categorizes the methods attempting to minimize the visual discomfort associated with VAC through enabling correct focus cues as Maxwellian view displays, vari-focal plane displays, multi-focal plane displays, integral imaging-based displays as well as computational multi-layer displays [17]. In this mentioned study, distinct methods are reviewed in detail; but, I think adding computational holographic displays into this categorization would be useful since there are promising recent studies about this topic.

1.3.1 Maxwellian view displays

The basic idea of Maxwellian view displays is that rays emitted by a point source are collimated by a condenser lens and focused by an eyepiece lens onto the eye's pupil instead of retina [18]. Thus, the depth of field (DOF) is effectively extended without the need of an accommodation due to the pinhole effect since the diameter of the source's image on the pupil is much smaller than the pupil diameter itself. So, it is one of the well-known ways to mitigate the VAC problem in HMDs and many studies propose holographic displays utilizing Maxwellian approach [19, 20]. In these studies, a laser source is collimated and parallel rays pass through a spatial light modulator (SLM) and then focused by a holographic optical element (HOE) onto the pupil so that the pattern on the SLM is imaged sharply on the retina as illustrated in Figure 1.6.

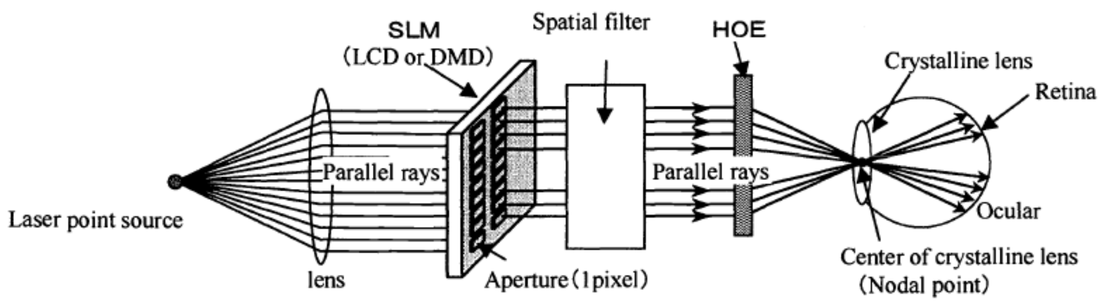


Figure 1.6: Holographic Maxwellian view display [20]

Main limitations of Maxwellian view displays are reduced spatial resolution and small entrance pupil diameter (EPD) which limits the eye-box of the system to the diameter of the eye's pupil.

1.3.2 Vari-focal plane displays

Vari-focal plane displays reduce VAC problem by dynamically altering the focal depth of the virtual plane through hardware adjustment. Thus, distance to the virtual plane matches with the convergence of the eyes if a focal change in the system is faster than the human depth perception. This can be achieved by mechanical adjustment in the display system i.e. changing the eye-relief distance of the eyepiece lens [21] or tuning the distance between microdisplay and eye-piece lens [22] to change the focal depth of the system. Alternatively, active optical elements such as focus-tunable lenses [23] and deformable mirrors [24] can be used.

Figure 1.7 illustrates the optical design of a near-eye display system that utilizes focus-tunable lenses to weaken the VAC effect. A magnifier lens is placed in front of the microdisplay for imaging. The focal length of the lens can be electrically controlled to adjust perceived retinal blur so that the system called accommodation-invariant display.

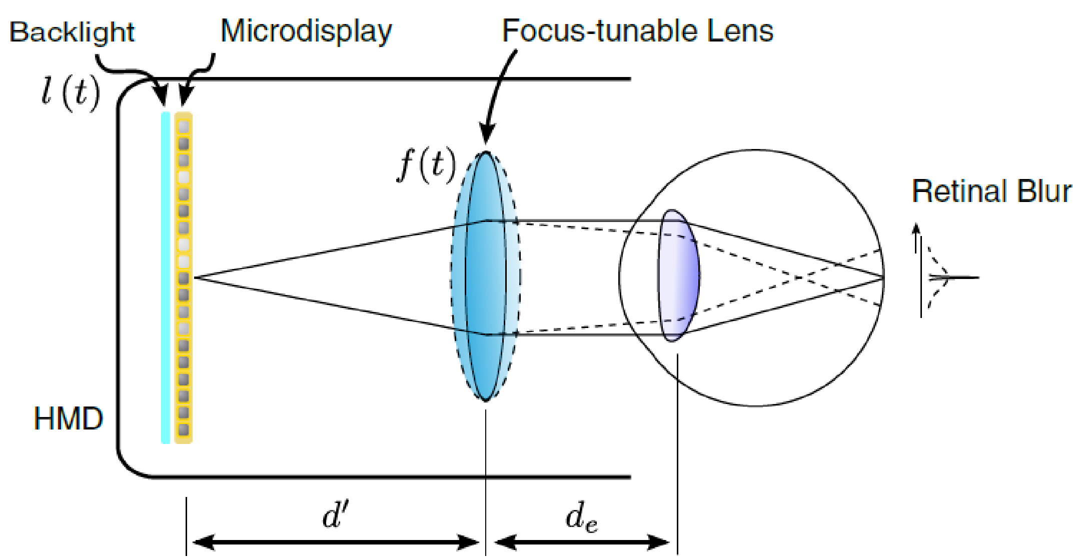


Figure 1.7: Optical layout of a vari-focal near-eye display with a focus-tunable lens [23]

In Figure 1.8 optical layout of another vari-focal display system adopting deformable membrane mirror is shown. An electronic-driven mechanism causing a uniform deflection of the membrane is deployed into the system. As a result of the deflection, a virtual image appears at distinct focal depth according to mirror curvature.

Note that, these systems require real-time eye-tracking to determine the user's 3D

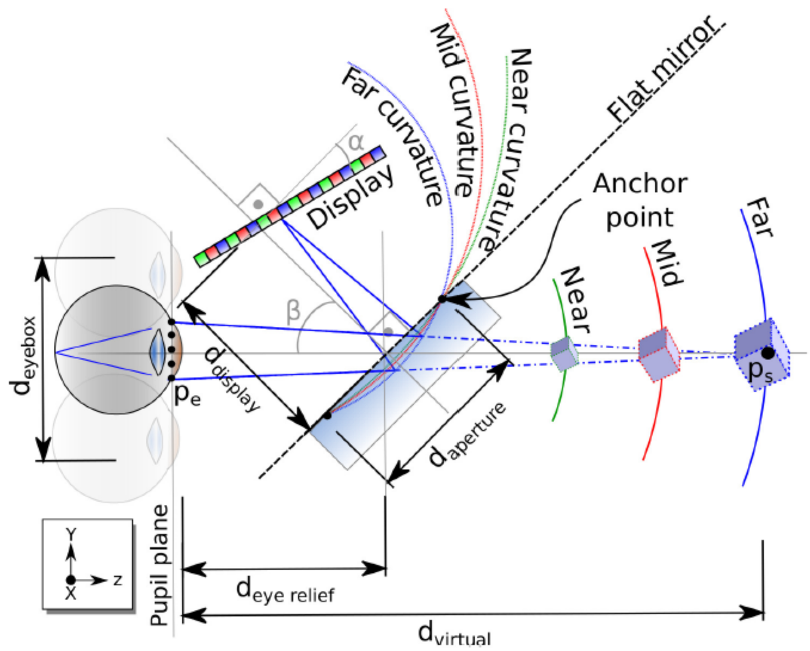


Figure 1.8: Optical layout of a vari-focal display system with a deformable membrane mirror [24]

gaze point in the scene. Although vari-focal display systems are capable to render correct focus cues through adjusting the focal depth of the system, objects at other depths than fixation distance still have incorrect focus cues.

1.3.3 Multi-focal plane displays

Producing multiple image planes in an HMD could be a remedy for matching convergence and accommodation. This approach can be implemented by either spatially multiplexing 2D displays [25, 26] or time-multiplexing method that involves very fast switching of the focal distance. [27, 28]. Projections of a 3D virtual scene into different discrete focal depths are rendered in multiplexed displays. Thus, the addition of these projections at the visual sight creates the virtual scene itself.

The principle of the spatially-multiplexed multi-focal approach is shown in Figure 1.9. This method utilizes a thick display system. Stacked transparent displays are able to adjust the depth of the displayed objects plane-by-plane and the virtual objects in the scene are distributed over focal depth planes. This method does not require any moving elements; however, it suffers from a lack of display technologies with high transmittance.

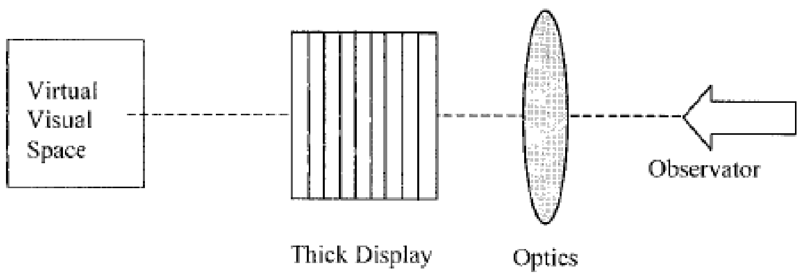


Figure 1.9: Schematic of spatially-multiplexed multi-focal planes HMDs [25]

The main idea behind the time-multiplexing approach is changing the focal depth of the virtual plane via moving elements rapidly i.e 120Hz. Thus, this approach requires stable and active optical elements and high computational power for a flicker-free experience.

1.3.4 Integral-imaging based displays

Integral-imaging based displays involve 3D sensing stage together with visualization stage. 2D projections of a 3D scene viewed from different eye positions can be recorded through a microlens array (MLA) and these images can be used to reconstruct the light fields of recorded 3D scene [29]. Image capturing stage is illustrated in Figure 1.10. MLA allows users to capture 2D images of the scene from different perspectives through a sensor. Each captured 2D image is called an elemental image (EI). To avoid overlapping between adjacent EIs, it is needed to place a barrier between MLAs. When the same MLA is used and EIs are rendered in the 2D display, the 3D scene can be obtained with correct focus cues as shown in Figure 1.11.
The main limitation of displays adopting this technique is its reduced spatial resolution which is proportional to the ratio of the micro-lens' focal length to the distance

of the display from the eye [31]. Thus, they suffer from a large pixel pitch size of available micro-displays.

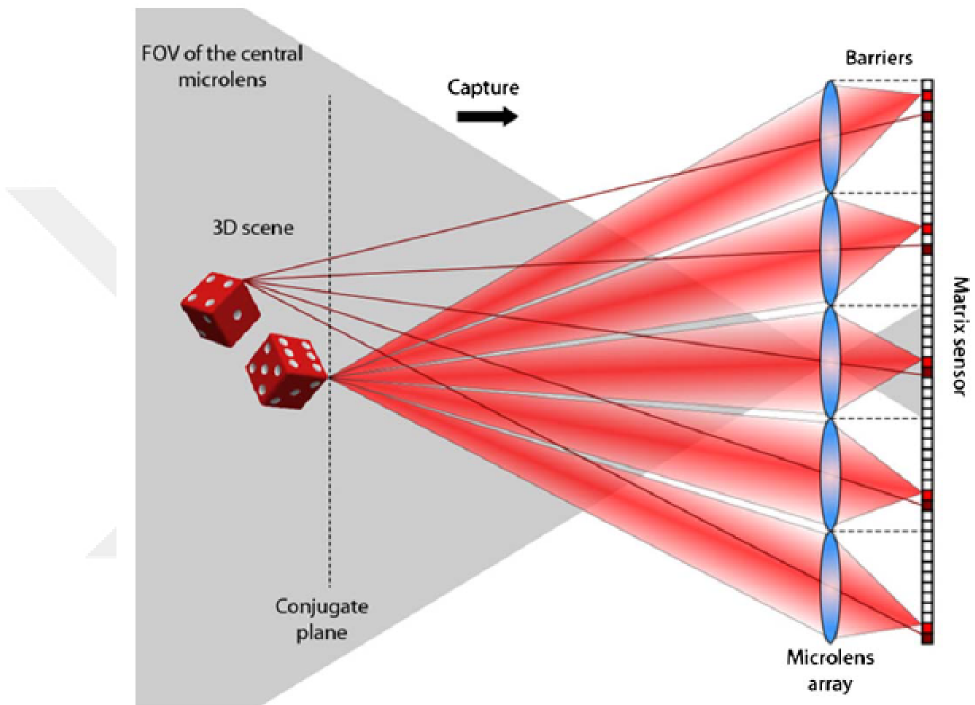


Figure 1.10: Image capture stage in integral imaging [30]

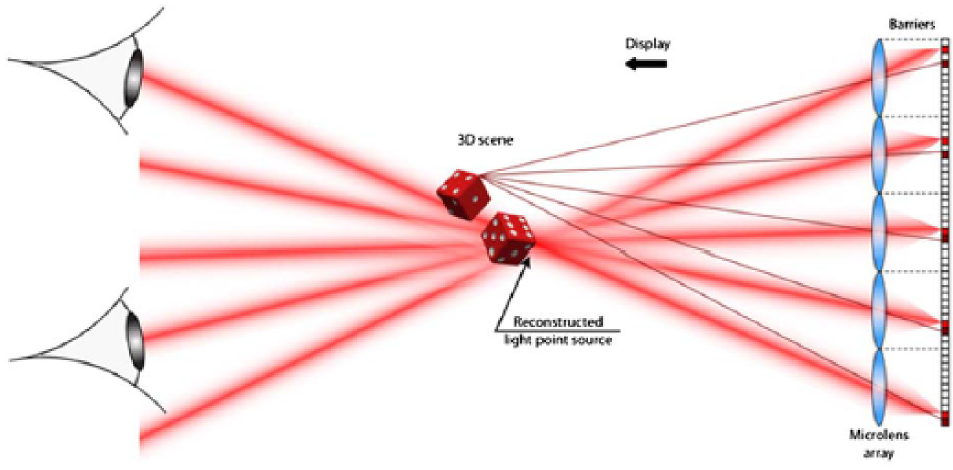


Figure 1.11: Display stage in integral imaging [30]

1.3.5 Computational multi-layer light-field displays

The computational multi-layer display method is a relatively new approach for mitigating VAC problem compared to other methods explained in previous subsections. In this method, multiple layers of transparent displays are illuminated through a uniform backlight to create light-fields of the 3D scene; thus, it utilizes a multiplicative fashion for imaging instead of addition method [32]. Recently, a study proposed a stack of two LCDs usage in a virtual reality application adopting computational light-field stereoscopy approach [33] as illustrated in Figure 1.12. In this study, a combined 4D light field to each eye of the viewer is presented instead of the magnified 2D image.

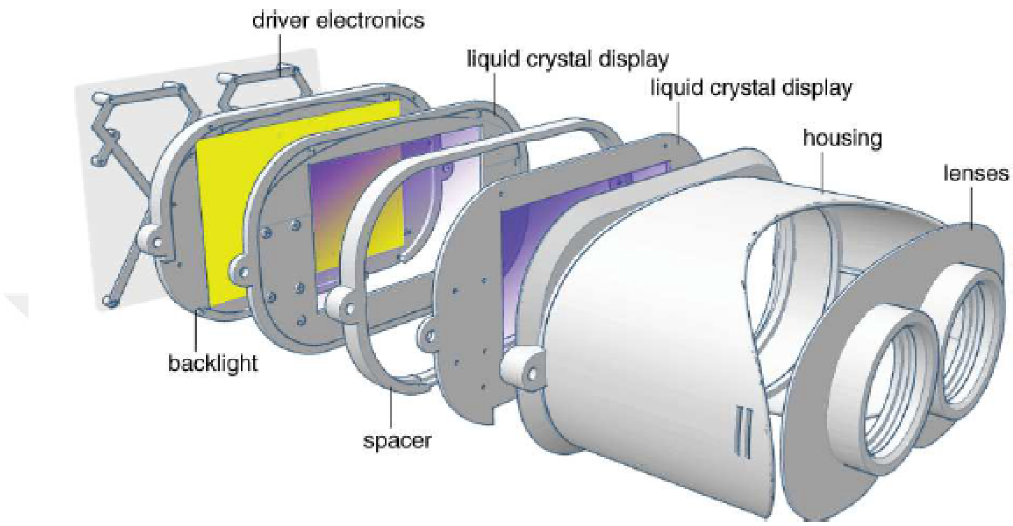


Figure 1.12: Schematic layout of a near-eye computational multi-layer light field display [33]

Whereas this method provides focus cues in addition to binocular disparity, it still suffers from reduced spatial resolution due to the diffraction stems from small pixel size of layers and the proposed method is computationally very expensive as well.

1.3.6 Computational holographic displays

Holographic displays can reproduce a 3D scene with all natural depth cues. SLMs combined with computer-generated holograms (CGHs) synthesize wavefronts that are indistinguishable from those emanated by physical objects [34]. The possibility of achieving the holographic near-eye-display (HNED) is justified by the basic operational principles illustrated in Figure 1.13.

The point light source generates a narrowband, spatially coherent diverging wave that illuminates a reflective and semi-transparent SLM. The SLM, by virtue of the properly computed CGH loaded on it, acts as an optical mask that transforms the illumination wave directly to the true light wave that would emanate from the virtual objects if they actually existed at their apparent locations. This latter wave, along with the

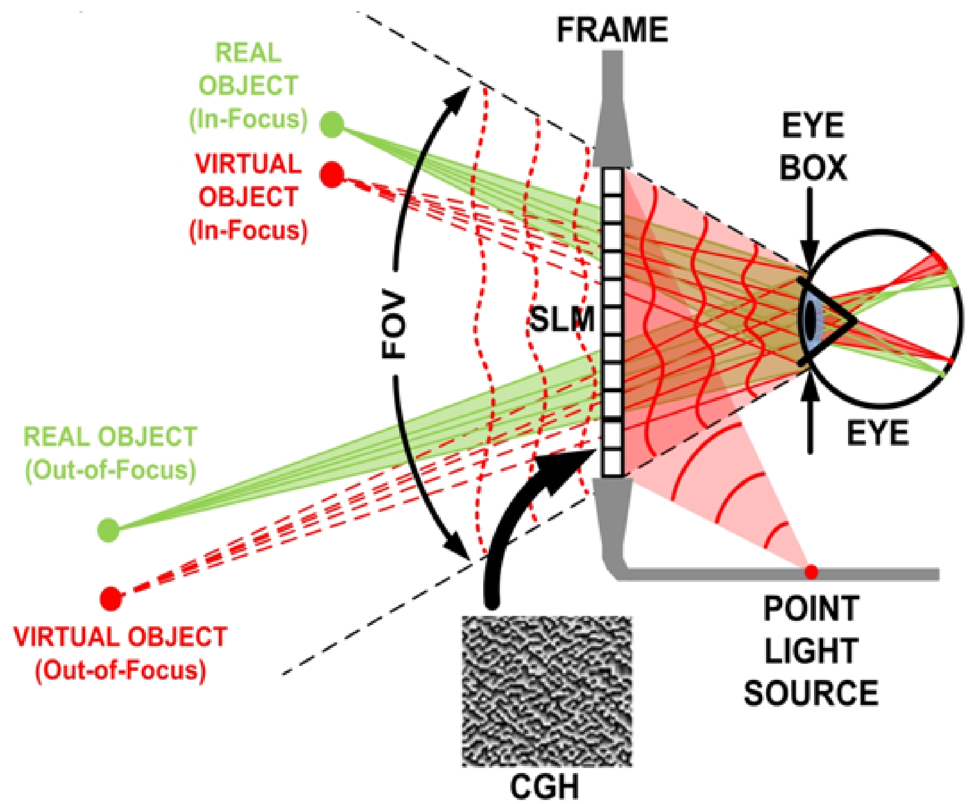


Figure 1.13: Schematic top view of a user wearing the HNED (left eye only). The CGH on the SLM converts the diverging wave from the point light source directly to the true light wave (with correct ray angles) from virtual objects. That wave goes through the eye pupil along with the light from real objects and forms a retinal image. The eye is focused on the nearby objects

wave emerging from the real objects, goes through the eye pupil and forms an image on the retina. Since the wave generated by the SLM is already shaped to possess the correct ray angles, the eye can readily form a focused image, without requiring the aid of relay lenses. The correct ray angles also lead to correct accommodation cues and remove the visual fatigue intrinsic to all stereoscopic displays. Virtual objects whose waves are delivered within the entire eye box are viewed at the retinal resolution. Always-in-focus objects can be displayed by utilizing a smaller portion (around 1 mm) of the eye pupil. The main drawbacks of computational holographic displays is high computational power requirement and large pixel pitch size of available SLM technologies.

1.4 Contributions of the Thesis

Multi-focus displays method is an important research area as they can reduce the vergence and accommodation conflict that is present in all stereoscopic 3D displays. This thesis demonstrates a novel biocular dual-focal plane see-through display system using a simple architecture. Our system has an advantage over fixed focus AR systems since two focal planes are simultaneously present. Our system does not have motion parallax or binocular disparity. Despite, it still gives very good immersive 3D feeling as both eyes simultaneously see the entire content at two different depths. The virtual depth planes appear at 25 cm (4 diopters) and 50 cm (2 diopters) in order to allow interaction at arms-length.

A gaze tracker was developed using a single camera. The camera captures both eyes and detects pupil centers accurately by fitting an ellipse to the radial edges of the iris through a coordinate transformation technique on the captured frames. The gaze tracker provides a binary gaze depth information based on instantaneous inter-pupillary distance by using the constructed gaze geometry.

The proposed system provides a unique and novel user experience as the translucent virtual objects are allowed to overlap in the line of sight but are still distinguishable by changing the accommodation of the eye. Important distinguishing features of our system are wide field-of-view (60x36degrees), large eye-box (100mm), and integrated gaze tracking system using single camera, which provides gaze-contingent information display without any infra-red illumination.

Chapter 2

WIDE FOV DUAL-FOCAL-PLANE SEE-THROUGH DISPLAY

To eliminate the error in the perception of the scene geometry in displays, correct focus cues must be enabled in different depth levels. The depth resolution of a human is 1/7 diopters due to the limited depth of field of the eye [35]. Thus, adjacent depth planes must be separated 1/7 D for a continuous depth of field and the range of human accommodation is at most 4 diopters for a typical near point. In this case, a display with full accommodative range requires 28 image planes [36]. However, such an implementation is not feasible due to lack of available transparent display technologies and high computational power requirement. Furthermore, stacking a high number of displays might suffer from bad form factor. Therefore, I decided to create a dual-focal-plane see-through display since it allows users to distinguish virtual objects in different depths through accommodation action of the eye. Whereas it causes a discontinuity in the virtual scene, it is a feasible application in terms of form factor and it presents a novel experience for users. In today's consumer HMD applications, interaction in the arms-length without VAC is a challenging task. Thus, distances to the focal depth planes are determined as 25 cm (2D) and 50 cm (4D).

2.1 Design

In this section, I discuss the design process of a spatially-multiplexed dual-focal plane biocular (both eyes see the same content) see-through near-eye display system prototype. The system provides correct focus cues in two focal planes with natural blur effect. Although the proposed display does not require focus tunable lenses and dynamic components, depth perception for the dual-focal plane is easily observable as it is shown in the dual-focal-plane interactive AR display chapter.

The dual-focal plane is acquired by using two separate planar LCDs placed at different optical distances with respect to the single magnifier Fresnel lens. Optical distances are optimized in ZEMAX to minimize the spot radius of different field points of the image. Virtual content is separated into two sections based on their depth information. The closer section to the viewer's eye appears at 25 cm (4D). The rear section is displayed at 50 cm (2D). Images are superimposed into real word through a planar beam splitter.

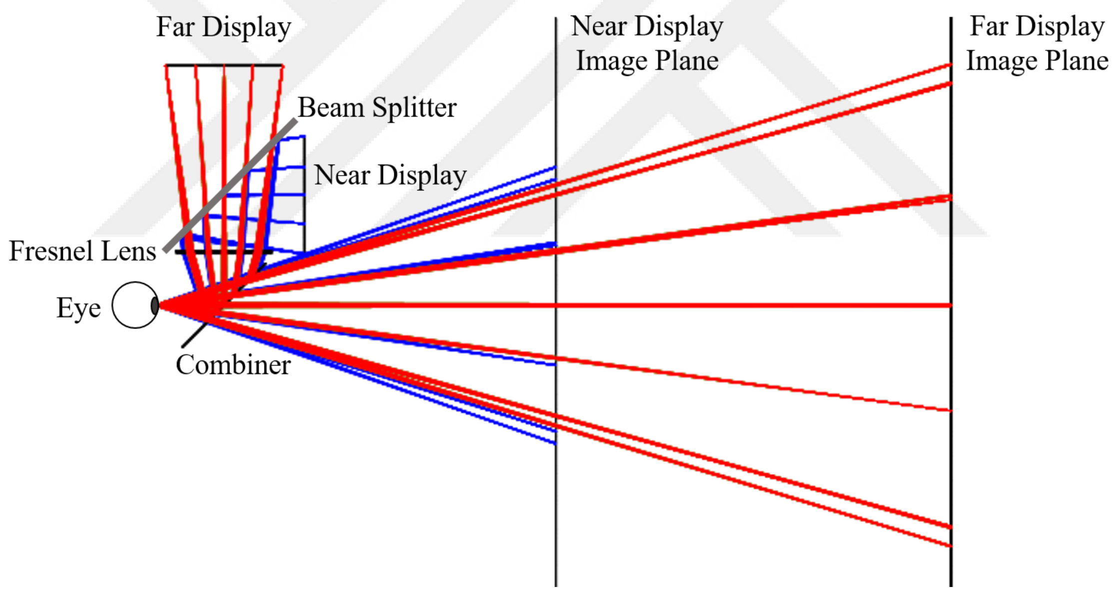


Figure 2.1: 3D Optical Layout in ZEMAX

2.2 Optical simulations

In this section, I evaluate several attributes of our display by simulations. All simulations are performed in ZEMAX optical design software. In all simulations, I used a ZEMAX model of the used Fresnel lens from Edmunds Optics official website [37]. The 3D optical layout of the setup is shown in Figure 2.1. In Figure 2.1, the blue solid lines represent the rays emanating from the near display. The rays emanating from the far display is illustrated by solid red colors, as evident in Figure 2.1.

2.2.1 Field-of-view

The proposed see-through display achieves an overlapping biocular 60.0 degrees horizontal and 35.5 degrees vertical FOV for both near and far displays, where the inter-pupillary distance (IPD) is set to 65 mm in the ZEMAX.

2.2.2 Eye-box

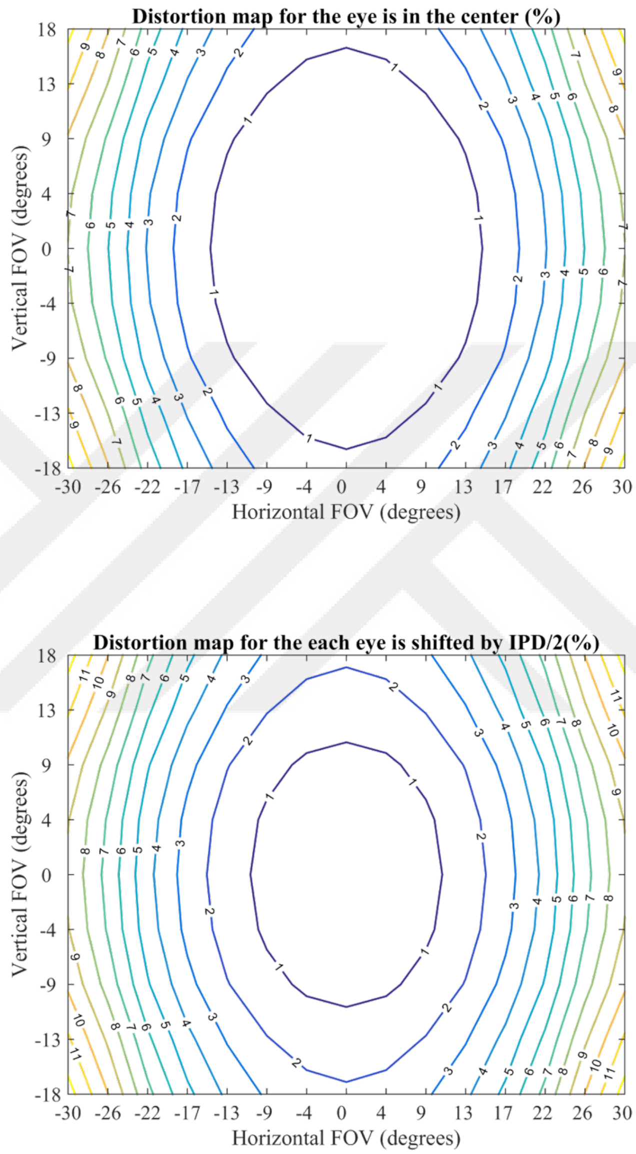
Eye-box size is defined as a space in which the user's head is decentered in this space and still, a horizontal and a vertical FOV of at least 10 degrees is available for both eyes [38]. Based on this definition, the designed see-through display has an eye-box of 105 mm in the horizontal axis and 66 mm in the vertical axis.

2.2.3 Optical distortion

For analyzing the optical distortion of the system, I assume that one of the users eyes is in the center of the eye-box. Distortion value of a single field point in percent determined as

$$Distortion(\%) = 100 \times \frac{y_{chief} - y_{ref}}{y_{ref}} , \tag{2.1}$$

where y_{chief} is chief ray height and y_{ref} is reference ray height for the undistorted ray as provided by ZEMAX. Distortion values are recorded for distinct points in the available instantaneous FOV. Then, I assume eyes are in the natural position which means each eye is shifted by IPD/2 or -IPD/2 in the horizontal axis. The simulation results of the optical distortion are shown in Figure 2.2. Figure 2.2(a) illustrates the distortion values of different field points, while the eye is located in the center of eye-box. In Figure 2.2(b), the eyes are shifted by 32.5 mm and the distortion values of the different field points along the FOV is reported for each eye. It is evident that the image distortion does not exceed 11% in the entire FOV when eyes are in their natural position. The distortion value increases at the edges of the eye-box due to the limited clear aperture of the Fresnel lens.



(a)

(b)

Figure 2.2: a) The distortion values of different field points, while the eye is in the center of eye-box. b) The distortion values of different field points, while eyes are in the natural position (i.e. off-centered by IPD/2=32.5mm)

The spot diagram of the system in Figure 2.3 shows the spot size variation for red, green, and blue wavelengths at 9 different field points at the center, edges, and corner points to cover the entire FOV. The RMS spot radius varies from $367\mu m$ to $667\mu m$ while the diffraction limited airy spot diameter is $164\mu m$ for the 25 cm virtual image distance. This corresponds to 1.5-3.0 cycles/degrees angular resolution. Image quality can be improved by using multiple lenses and glass optics.

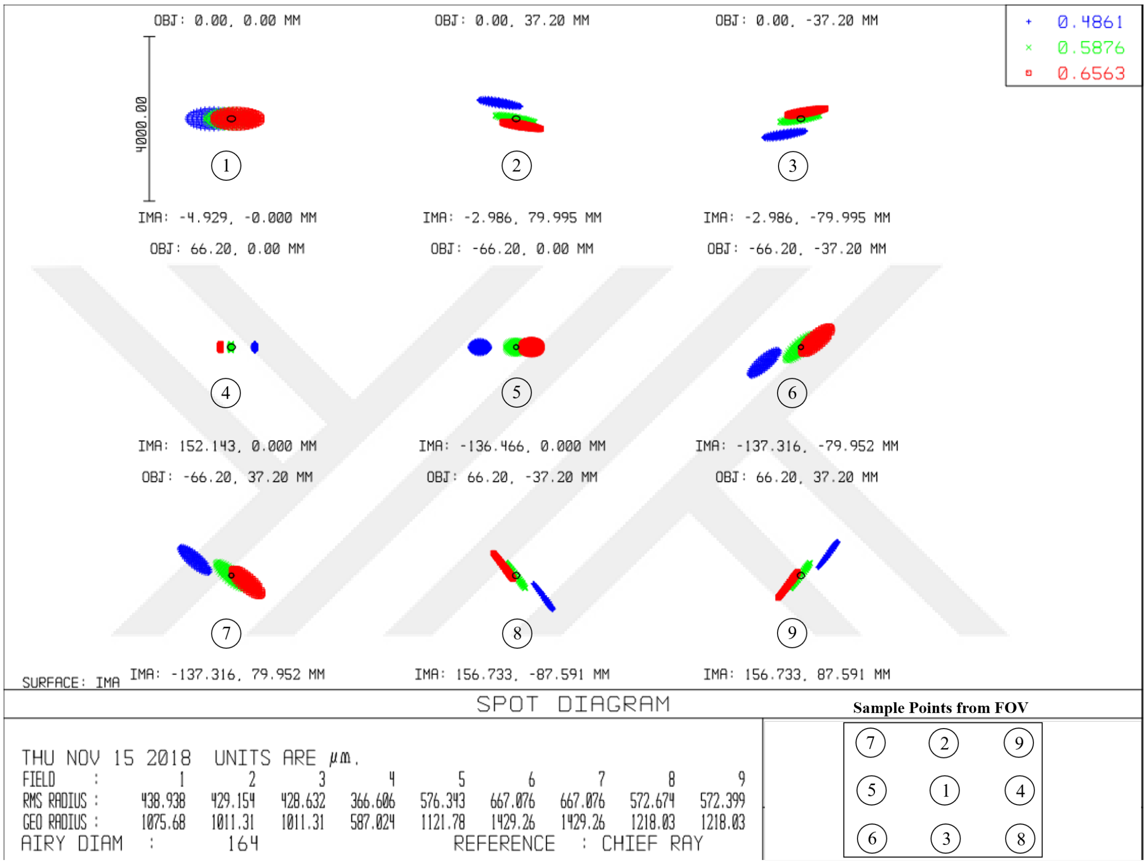


Figure 2.3: The spot radius sizes for 9 sample points, which are selected as the center, edge and corner points of the display to cover the entire FOV. In this ZEMAX simulation, the eye is shifted by IPD/2 and the wavelengths are chosen to support the visible light spectrum (red color with 656.3 nm, green color with 587.6 nm and blue color with 486.1 nm). Positions of the field points in the FOV are shown in bottom-right.

As observed in Figure 2.2 and Figure 2.3, the proposed system suffers from optical distortion and chromatic aberration. Since the full images on both LCDs are seen by both eyes, distortion or chromatic aberration compensation is not performed, i.e any correction that improves the right eye image quality will result in an equal degree of degradation in the left eye image quality in the horizontal axis. While the angular resolution is 10x below the retinal resolution limit and maximum optical distortion is 11%, the experimental setup has good image quality and effectively demonstrates the natural blurring effect as discussed below.

Chapter 3

SINGLE CAMERA GAZE TRACKER WITHOUT INFRA-RED ILLUMINATION

A gaze tracker is also added to the system to interact with the virtual objects. The system uses a single camera (Goldmaster V-52 webcam) with the resolution of 640x480 pixels and it does not require any infra-red illumination. Furthermore, I aimed an easy calibration process. The calibration process assumes a fixed head position. The output of the algorithm is a single digit output, indicating the display at which the user converges for the central region in the virtual scene.

3.1 Gaze Geometry

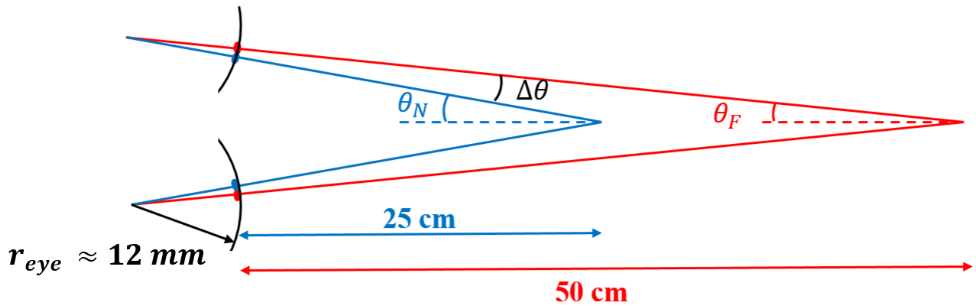
To clarify the assumptions in the gaze tracking algorithm, two schematic drawings are provided in Figure 3.1. A geometrical model is developed to estimate the convergence depth of a user, as shown in Figure 3.1(a). Assuming the IPD of the user is 65 mm and using the geometry shown in Figure 3.1(a), the rotation angle difference of each eye is computed as:

$$\Delta\theta = \theta_N - \theta_F = tan^{-1}(\frac{\text{IPD/2}}{\text{Near depth}}) - tan^{-1}(\frac{\text{IPD/2}}{\text{Far depth}}) = 3.7^\circ = 0.065 \text{ rad}, \quad (3.1)$$

where θ_N and θ_F are the angles of rotation when the user gazes at the near and far display, respectively.

Figure 3.1(b) shows the IPD shift between the near and far focus planes. Ellipses represent the left and the right eyes. The eye pupil positions are color coded. The blue dots indicate the position of pupils when the user converges at the near display.

(a)



(b)

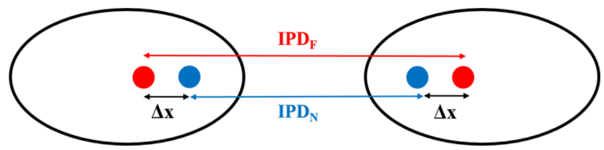


Figure 3.1: a) A geometrical model is developed to estimate the convergence depth of a user. b) Pupil centers illustrated on the left and right eyes. The blue and red dots respectively indicate the positions of pupils when user converges at the near display and far display

Locations of eye centers while gazing at far display are illustrated by red dots. Based on the model in Figure 3.1(a), IPD difference will be:

$$\Delta IPD = 2\Delta x \approx 2 \times r_{eye} \times \Delta\theta = 1.56mm \,, \tag{3.2}$$

where the radius of the eye is assumed 12 mm. Horizontal 640 pixels across the camera correspond to 128 mm for the fixed head and camera position in our setup, therefore the amount of shift ΔIPD corresponds to approximately 8 pixels. Since 2 diopters shift from 50 cm to 25 cm correspond to 8 pixels, camera limits the minimum measurable gaze to 0.25 diopters. The real limit comes from the noise and the algorithms. The accuracy can be improved by using a camera with a smaller FOV, larger pixel count, and frame averaging.

3.2 Calibration Process and Gaze Tracking Algorithm

The main steps of the calibration, gaze tracking, and the interaction process are shown in Figure 3.2(a). The user looks at the near display for a while. The tracking algorithm finds the locations of the eye center. The average of the eye-center locations is recorded and IPD_N is extracted in terms of pixels. Afterward, the same process is repeated for the far display and IPD_F is recorded for the user. After acquiring near and far IPDs, a threshold is determined for robust decision making in the gaze tracking system. Determining convergence depth requires robust pupil center estimation for each eye and accurate IPD calculation in each frame. The details of iris tracking and estimating the accurate location of the eye center algorithm are shown in Figure 3.2(b).

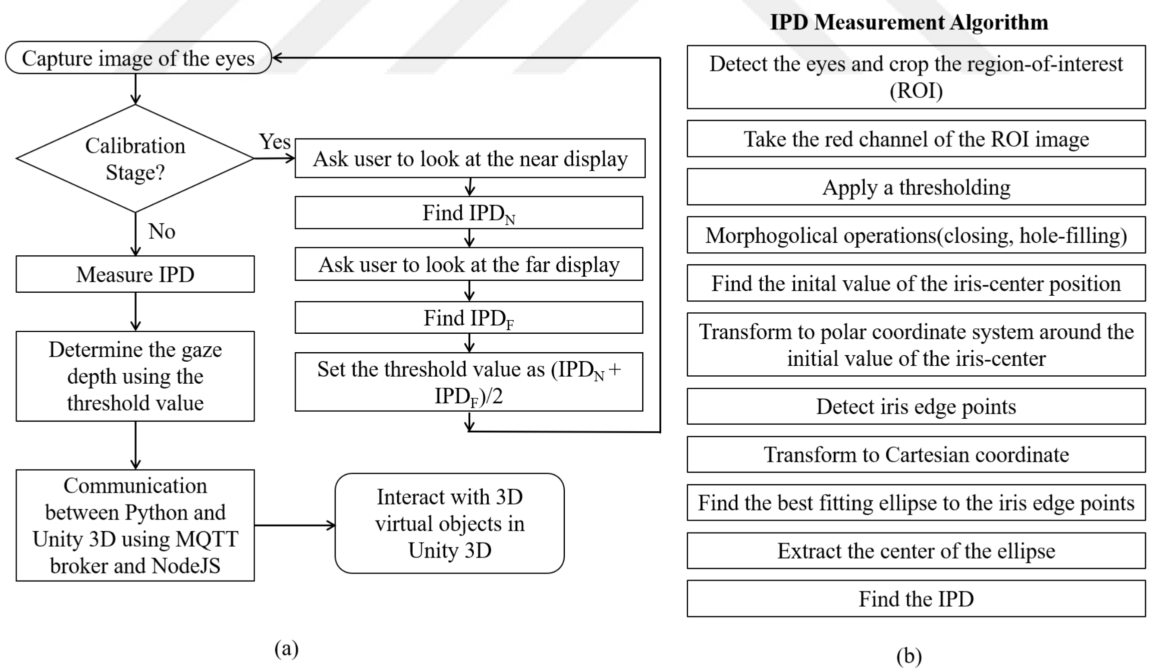


Figure 3.2: a) Calibration and tracking process. b) Details of the eye center and IPD estimation algorithm

Detailed steps of the finding accurate location of the eye center are shown in Figure 3.3. For each frame, I extract a region of interest (ROI) for each eye by using the cascade classifiers as illustrated in Figure 3.3(a). Only the red channel of the RGB image of ROI is used instead of the gray-scale image. The red-channel image, which is presented in Figure 3.3(b), is segmented into two classes - iris, and background - through thresholding via the optimized threshold value. After morphological operations on the segmented image such as hole-filling and closing, rough eye center position is determined and presented in Figure 3.3(c). The resulting image is transformed into polar coordinates around the initial value of the eye center which as shown in Figure 3.3(d). In this case, the iris fills the left-side of the polar transformed image. Strong vertical edge points, indicated in the region between dashed red lines in Fig. 10(d), correspond the accurate radial edges of the iris. Figure 3.3(e) demonstrates the transformed image in the cartesian coordinates. In this figure, white points correspond the accurate radial edges [39] of the iris found in the previous step. The best-fitting ellipse and its center are acquired through a direct least-squares method [40] which is shown in Figure 3.3(e). Ellipse's center corresponds to the accurate eye center position in 2D and they are projected into the original frame as illustrated in Figure 3.3(f).

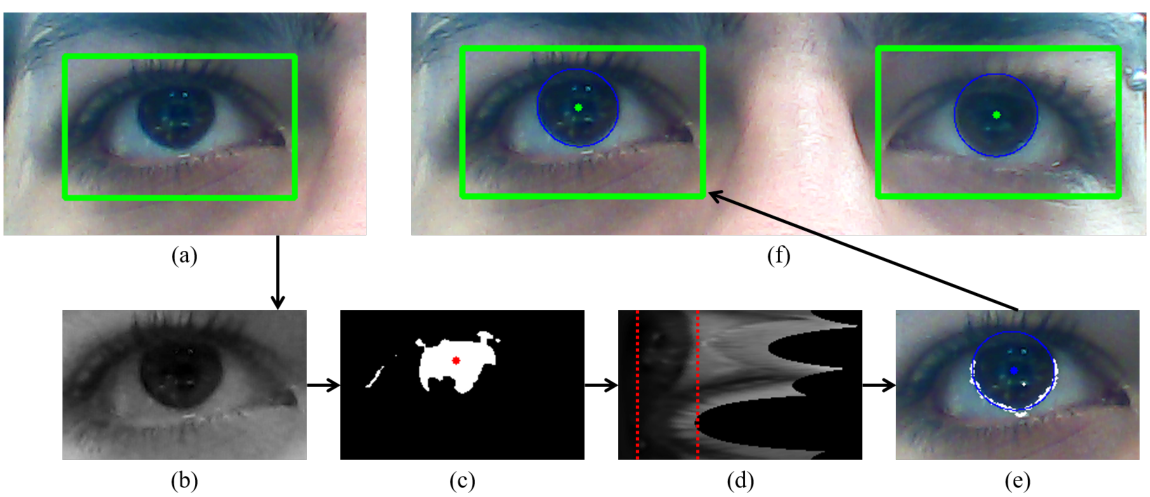


Figure 3.3: a) Region of interest for one eye b) Red-channel image c) Thresholded and segmented image d) Polar transformed image e) Radial edge points, the best-fitting ellipse, and its center (illustrated as blue dot) f) Final result on both eyes (computed pupil center illustrated as a green dot)

3.3 Experimental Results for Different Users

The gaze tracking system is tried by 4 different users. IPD distribution for convergence in the near and far display for the users is presented in Figure 3.4(a). Distribution in each interval is represented by a box by taking the data extracted from 180 frames into consideration. On each box, the central red line indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points and the outliers are plotted individually using the red + symbol [41]. Thresholds are determined in a way that system can decide which display user is looking at. The threshold value changes between users automatically since the users have different nominal IPDs. Then, I repeated the gaze tracking and decision-making experiment 10 times by asking the User-4 to look at the near and the far display consequently. I recorded the IPD value at each frame while the user is converged in the near or far display and each interval consists of 60 frames. Results are shown in Figure 3.4(b). A pre-determined IPD threshold value, found in the calibration process, is illustrated by the dashed red line. It can be clearly observed that I acquired a robust gaze tracker system with only a few outliers. Our implemented gaze tracker is optimized for a few people in a certain illumination condition. Thus, the results might differ in different illumination conditions and for the people with extreme IPDs or different eye shapes. Furthermore, robustness of the system might suffer from the head movement.

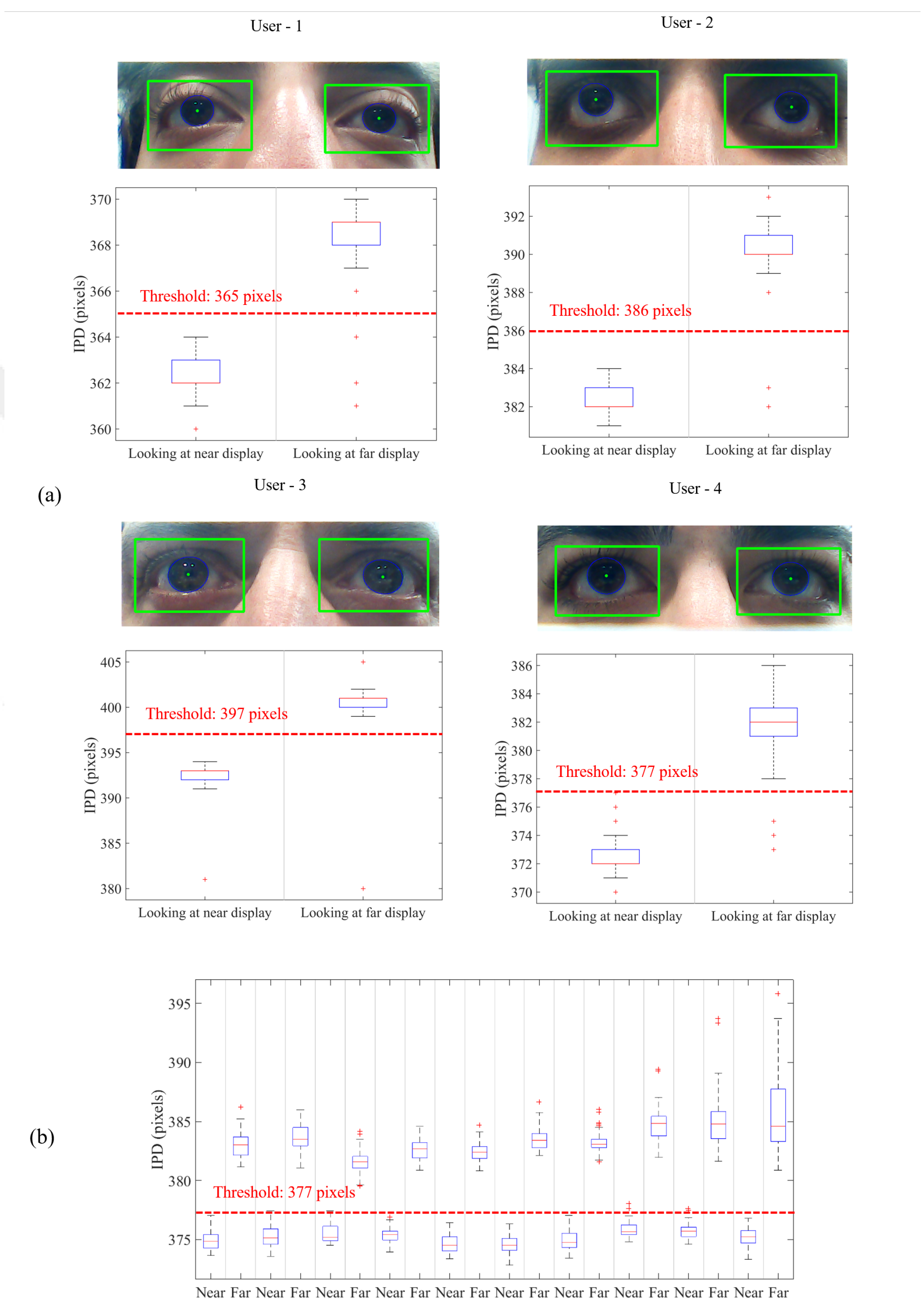


Figure 3.4: a) Measured IPD distribution of 4 users while looking at the near display and far display subsequently during the calibration process. Data extracted from 180 frames in each interval. Red dashed line shows the selected threshold value for each user. b) Repeated gaze tracking results with User-4 when the user repeatedly focuses on near and then far display. User-4. Each box consists of IPD measurement extracted from 60 frames. The threshold value found in the calibration process is illustrated by the dashed line.

Chapter 4

DUAL-FOCAL-PLANE INTERACTIVE AR DISPLAY

4.1 Display Prototype

Figure 4.1 illustrates the schematic drawing of the dual-focal-plane interactive AR display system. In Figure 4.1, the far display (sketched with a solid red color) is placed at 125 mm from the lens. The image of the far display (shown with a transparent red color) is formed 50 cm away from the viewer's eye. The physical distance between the near display (solid blue colored display in Figure 4.1) and the lens is 86 mm. The distance of the image of the near display to the user's eye is 25 cm. The distance between the eyes and the combiner is 40 mm. The combiner is placed 33 mm below the lens. Furthermore, a camera is placed towards to eye to track depth of user's convergence depth for interaction with the 3D content.

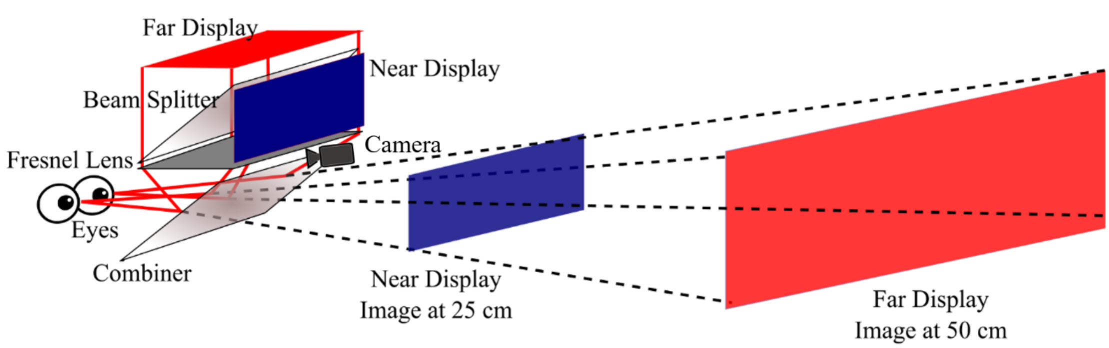


Figure 4.1: Schematic illustration of the optical layout

I use two Topfoison TF60010A 1440x2560 6.0" TFT LCD panel with a refresh rate of 60Hz to display rendered images. These displays can be driven by a single portable computer simultaneously. As a magnifier, a 6" (152.4 mm) focal length acrylic Fresnel lens with 1.5 mm thickness from Edmund Optics is used. I used a 50/50 flat combiner in front of the user's eyes with a thickness of 2 mm. I intentionally chose a very thin beam-splitter to eliminate the ghost artifacts and augment the contents displayed on the LCDs. Furthermore, a chin-rest is placed to fix the head position for the gaze tracking process. A photo of the implemented bench-top prototype hardware is illustrated in Figure 4.2.

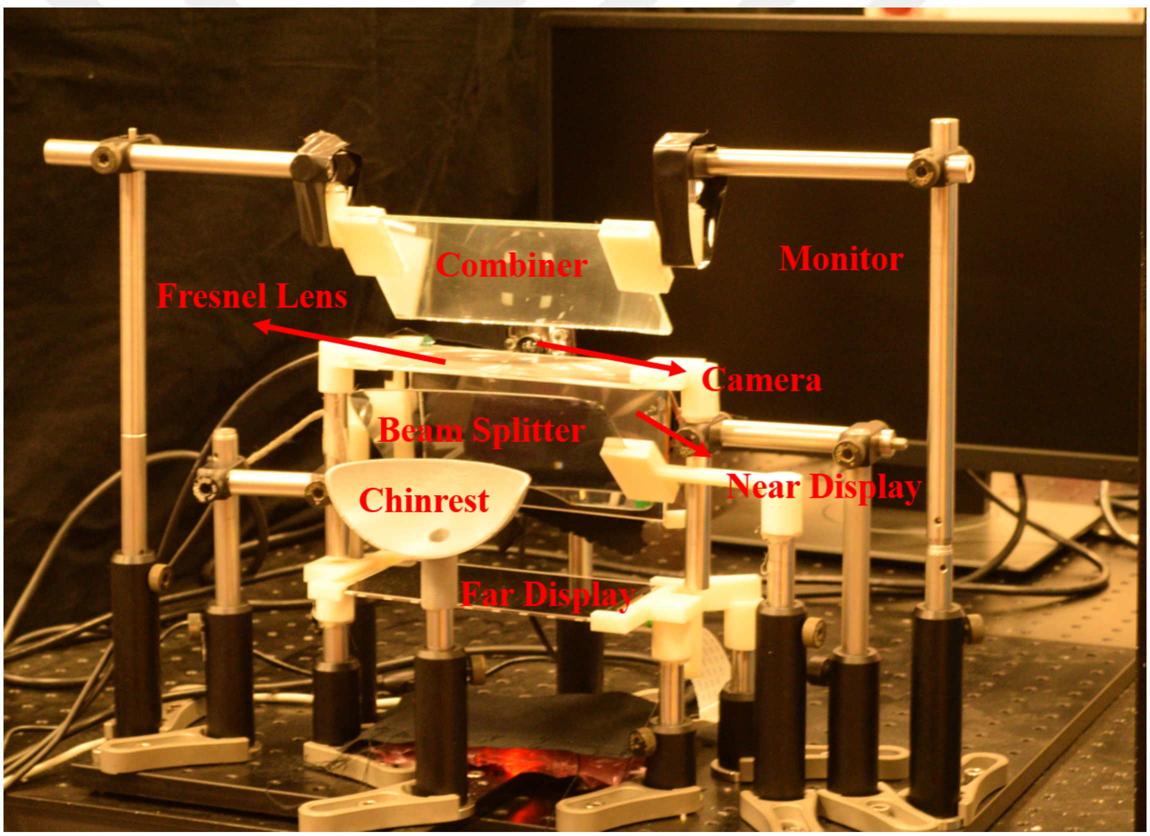


Figure 4.2: The prototype hardware including 2 LCDs for two focal planes and camera for simultaneously tracking two eyes and computing the gaze distance

4.2 Content Generation and Experimental Results

In this section, I present the content generation process and experimental results of the proposed system. I use Unity 3D game engine and *C#* programming language to render the virtual scene. Two virtual cameras are placed in the same virtual 3D position inside Unity 3D with different culling masks. Each virtual camera renders the scene with the associated culling mask and displays the rendered content on the corresponding LCD. Virtual objects that are supposed to be closer than 3 diopters rendered in the near display and virtual objects with a depth of more than 3 diopters are rendered in the far display. In this way, the rendered objects are distributed over one of the available fixed virtual depth planes. Then, we designed a road scene as a convenient use-case of the proposed see-through display. A large LCD monitor (Dell 21.5" 1920x1080 pixels at a distance of 50cm (2D) is used as a background scene display. The virtual scene which is rendered in Unity 3D is augmented on the background scene which is displayed in the monitor. As a result, I get three video outputs simultaneously from Unity 3D and display them in two LCDs together with the monitor synchronously.

While Figure 4.3 illustrates the contents shown in both LCDs that correspond to two frames, Figure 4.4 shows the experimental results captured with a camera. The frames shown in Figure 4.3 are displayed by the proposed system and a background scene is displayed using a monitor. Scenes, provided by the display, are captured with Nikon D5300 DSLR camera. The camera aperture was set to f/10 with an exposure time of 1/800 seconds.

In Figure 4.4(a), when the camera is focused in the near-focal-plane; speedometer (bottom-left), warning signs (upper-left), navigation map (upper-right) and an animation showing car's technical situation (bottom-right) are sharp, and they are displayed in the near display as shown in Figure 4.3(a). In this case; blue arrow signs, hotel sign, and yellow bounding box are blurred, and they are displayed in the far display as illustrated in Figure 4.3(b). The experimental results show that the implemented display provides correct depth cues. The user can distinguish the super-imposed virtual objects, even when they are in the same line of sight due to the clear depth perception.

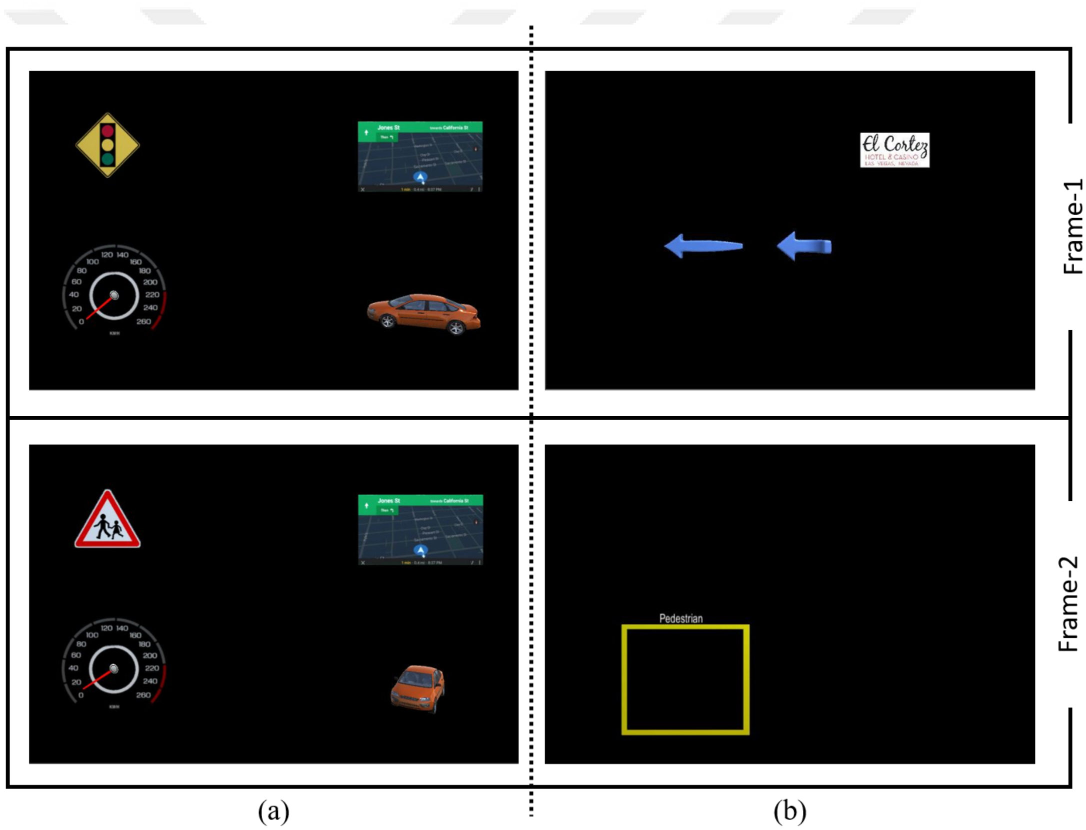


Figure 4.3: a) Content appearing in the near display b) Content appearing in the far display

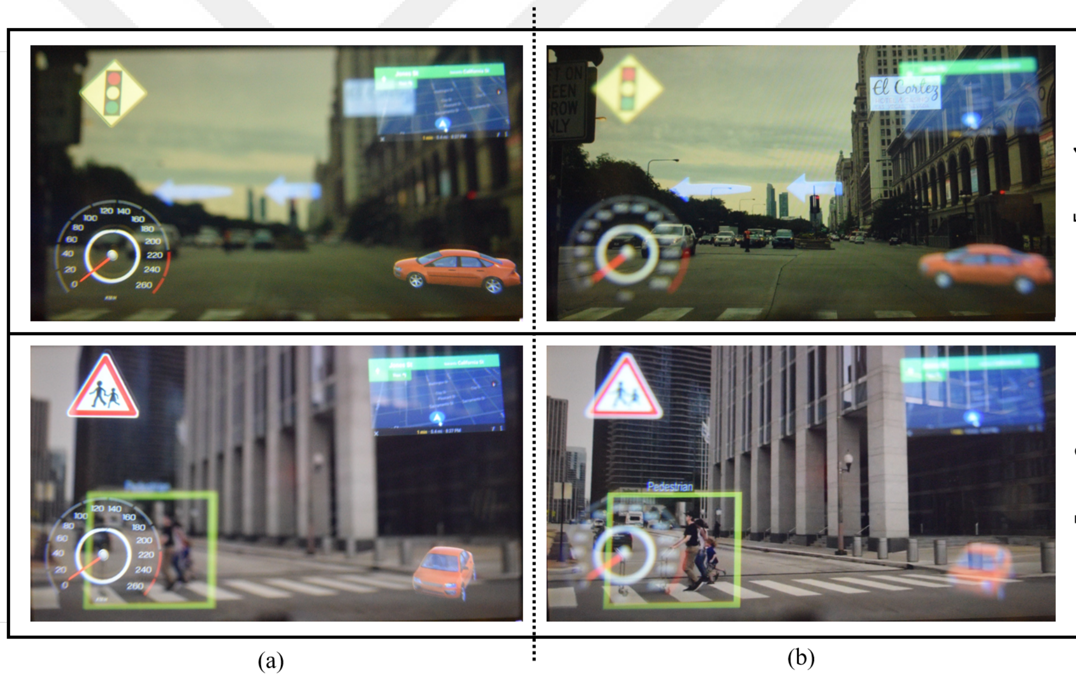


Figure 4.4: Experimental captures of the virtual scene. a) The camera is focused in 4 diopters (25 cm). b) The camera is focused in 2 diopters (50 cm). FOV is 60.0 by 35.5 degrees and the image can be seen with both eyes simultaneously.

4.3 Interaction with Gaze Tracker

The gaze position of the user is estimated at each frame. I used the information of the gaze depth plane to interact with 3D virtual objects placed on the different planes of convergence. To demonstrate the interaction, I created a proof-of-concept game using the Unity3D game engine [42]. The gaze tracker script is developed in Python programming language utilizing OpenCV library [43]. To establish communication between Unity3D game engine and the gaze tracker script, I used the NodeJS Mosca library of the MQTT broker [44]. I send binary gaze depth information via a specific topic to the broker. Then, the broker sets the gaze state flag in Unity3D. The gaze state flag is used in *C#* script to change the virtual content on the proper display. The screen-shots of the game together with acquired IPD values are illustrated in Figure 4.5. As evident in Figure 4.5, two texts are displayed at different depths (or displays). User changes the color of the near object into the red by converging at the near display, as shown in Figure 4.5(a). The color of the far object is turned into red in Figure 4.5(b), when the user converges at the far display.

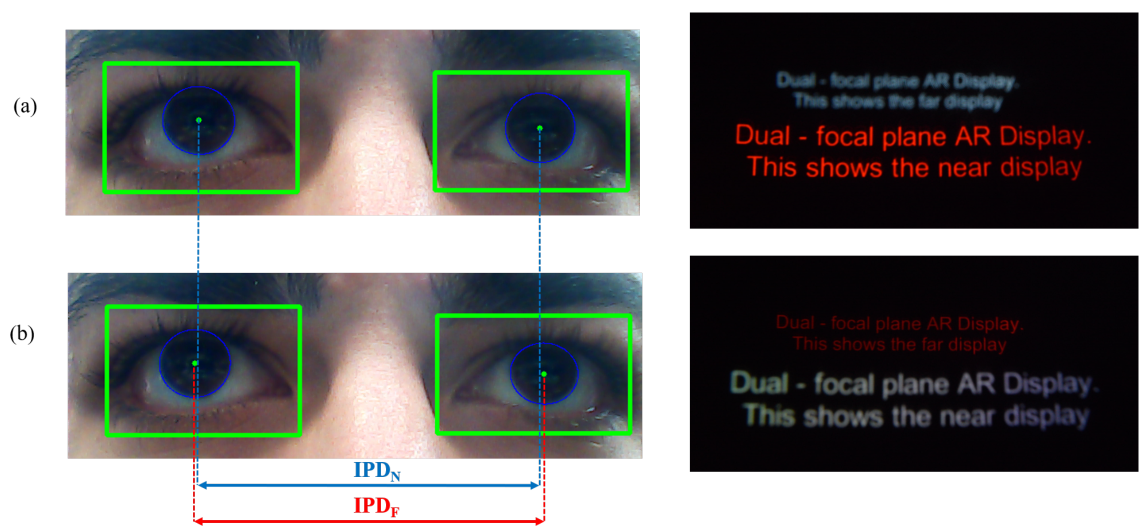


Figure 4.5: Interaction with gaze tracking. User's gaze is at the a) near display and b) far display. Based on the user's gaze, text color changes automatically.

Chapter 5

CONCLUSION

In this dissertation, I proposed and demonstrated a simple dual-focal plane AR display prototype. I implemented a fast and accurate gaze tracker using a single camera to find whether the user is gazing at the near or far display. The presented display reduces the VAC by letting the user perceive objects in two different depths. I achieved overlapping biocular 60.0 degrees horizontal and 35.5 degrees vertical FOV with the eye-box size of 105 mm in the horizontal axis and 66 mm in the vertical axis. Optical distortion analysis realized in ZEMAX shows that maximum distortion does not exceed 11% for the natural position of the eyes at the corners of the image. Depth perception and natural blur effect can be clearly observed in the experiments for two different focal planes. Furthermore, gaze tracker is integrated with the rendering system to automatically change the content based on the gaze distance. The experimental results show that the proposed AR display system is promising to study wide FOV and gaze-tracker based interactive AR displays. While the system is not head-worn, it can be used as a near-eye display system for AR and vision research and for various medical and other training tasks with the guidance of experts. The system can be miniaturized for use as head-worn display or modified to provide larger eye-relief for use as a head-up display.

BIBLIOGRAPHY

[1] Dodgson, N. A. (2004). Variation and extrema of human interpupillary distance. *Stereoscopic Displays and Virtual Reality Systems XI.* doi:10.1117/12.529999

[2] Goodman, J. (1996). *Introduction to fourier optics (2nd ed.).* New York, NY: McGraw-Hill Higher Education

[3] Cholewiak, S. A., Love, G. D., & Banks, M. S. (2018). Creating correct blur and its effect on accommodation. *Journal of Vision*, 18(9), 1. doi:10.1167/18.9.1

[4] Sulai, Y., Geng, M., Luanava, S., Gao, W., Gollier, J., Wheelwright, B. M., & Choi, S. (2018). Field of view: Not just a number. *Digital Optics for Immersive Displays.* doi:10.1117/12.2307303

[5] Ebenholtz, S. M. (2001). *Oculomotor systems and perception.* New York: Cambridge University Press.

[6] Geng, J. (2013). Three-dimensional display technologies. *Advances in Optics and Photonics*, 5(4), 456. doi:10.1364/aop.5.000456

[7] Qian, N. (1997). Binocular Disparity and the Perception of Depth. *Neuron*, 18(3), 359-368. doi:10.1016/s0896-6273(00)81238-6

[8] Held, R. T., Cooper, E. A., Obrien, J. F., & Banks, M. S. (2010). Using blur to affect perceived distance and size. *ACM Transactions on Graphics*, 29(2), 1-16. doi:10.1145/1731047.1731057

[9] Wheatstone, C. (1852). Contributions to the physiology of vision.Part II.On some remarkable, and hitherto unobserved, phenomena of binocular vision. A Bakerian lecture. *Journal of the Franklin Institute*, 54(3), 196-199. doi:10.1016/0016-0032(52)90777-1

[10] Howard, I. P., & Rogers, B. J. (1996). Development and pathology of binocular vision. *Binocular Vision and Stereopsis*, 603-644. doi:10.1093/acprof:oso/9780195084764.003.0015

[11] Fincham, E. F., & Walton, J. (1957). The reciprocal actions of accommodation and convergence. *The Journal of Physiology*, 137(3), 488-508. doi:10.1113/jphysiol.1957.sp005829

[12] Lambooij, M., Ijsselsteijn, W., Fortuin, M., & Heynderickx, I. (2009). Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. *Journal of Imaging Science and Technology*, 53(3), 030201. doi:10.2352/j.imagingsci.technol.2009.53.3.030201

[13] Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergenceaccommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3), 33. doi:10.1167/8.3.33

[14] Shibata, T., Kim, J., Hoffman, D. M., & Banks, M. S. (2011). The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision*, 11(8), 11-11. doi:10.1167/11.8.11

[15] Mather, G., & Smith, D. R. (2000). Depth cue integration: Stereopsis and image blur. *Vision Research*, 40(25), 3501-3506. doi:10.1016/s0042-6989(00)00178-4

[16] Banks, M. S., Cholewiak, S. A., Love, G. D., Srinivasan, P., & Ng, R. (2017). ChromaBlur: Rendering Chromatic Eye Aberration Improves Accommoda-

tion and Realism in HMDs. *Imaging and Applied Optics 2017 (3D, AIO, COSI, IS, MATH, PcAOP)*. doi:10.1364/3d.2017.dm2f.1

[17] Hua, H. (2017). Enabling focus cues in head-mounted displays. *Imaging and Applied Optics 2017 (3D, AIO, COSI, IS, MATH, PcAOP)*. doi:10.1364/3d.2017.jtu1f.2

[18] Westheimer, G. (1966). The maxwellian view. *Vision Research*, 6(11-12), 669-682. doi:10.1016/0042-6989(66)90078-2

[19] Ando, T., Yamasaki, K., Okamoto, M., Matsumoto, T., & Shimizu, E. (2000). Retinal projection display using holographic optical element. *Practical Holography XIV and Holographic Materials VI*. doi:10.1117/12.379998

[20] Ando, T., & Shimizu, E. (2001). Head-mounted display using holographic optical element. *Three-Dimensional Video and Display: Devices and Systems: A Critical Review*. doi:10.1117/12.419789

[21] Shiwa, S., Omura, K., & Kishino, F. (1996). Proposal for a 3-D display with accommodative compensation: 3DDAC. *Journal of the Society for Information Display*, 4(4), 255. doi:10.1889/1.1987395

[22] Shibata, T., Kawai, T., Ohta, K., Otsuki, M., Miyake, N., Yoshihara, Y., & Iwasaki, T. (2005). Stereoscopic 3-D display with optical correction for the reduction of the discrepancy between accommodation and convergence. *Journal of the Society for Information Display*, 13(8), 665. doi:10.1889/1.2039295

[23] Konrad, R., Padmanaban, N., Molner, K., Cooper, E. A., & Wetzstein, G. (2017). Accommodation-invariant computational near-eye displays. *ACM Transactions on Graphics*, 36(4), 1-12. doi:10.1145/3072959.3073594

[24] Dunn, D., Tippets, C., Torell, K., Kellnhofer, P., Aksit, K., Didyk, P., & Fuchs, H. (2017). Wide Field Of View Varifocal Near-Eye Display Using See-Through Deformable Membrane Mirrors. *IEEE Transactions on Visualization and Computer Graphics*, 23(4), 1322-1331. doi:10.1109/tvcg.2017.2657058

[25] Rolland, J. P., Krueger, M. W., & Goon, A. (2000). Multifocal planes head-mounted displays. *Applied Optics*, 39(19), 3209. doi:10.1364/ao.39.003209

[26] Schowengerdt, B. T., Murari, M., & Seibel, E. J. (2010). Volumetric Display using Scanned Fiber Array. *SID Symposium Digest of Technical Papers*, 41(1), 653. doi:10.1889/1.3500550

[27] Liu, S., & Hua, H. (2009). Time-multiplexed dual-focal plane head-mounted display with a liquid lens. *Optics Letters*, 34(11), 1642. doi:10.1364/ol.34.001642

[28] Love, G. D., Hoffman, D. M., Hands, P. J., Gao, J., Kirby, A. K., & Banks, M. S. (2009). High-speed switchable lens enables the development of a volumetric stereoscopic display. *Optics Express*, 17(18), 15716. doi:10.1364/oe.17.015716

[29] Lippmann, G. (1908). preuves rversibles donnant la sensation du relief. *Journal De Physique Thorique Et Applique*, 7(1), 821-825. doi:10.1051/jphystap:019080070082100

[30] Xiao, X., Javidi, B., Martinez-Corral, M., & Stern, A. (2013). Advances in three-dimensional integral imaging: Sensing, display, and applications [Invited]. *Applied Optics*, 52(4), 546. doi:10.1364/ao.52.000546

[31] Lanman, D., & Luebke, D. (2013). Near-eye light field displays. *ACM SIGGRAPH 2013 Talks on - SIGGRAPH 13*. doi:10.1145/2504459.2504472

[32] Maimone, A., & Fuchs, H. (2013). Computational augmented reality eyeglasses. *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. doi:10.1109/ismar.2013.6671761

[33] Huang, F., Chen, K., & Wetzstein, G. (2015). The light field stereoscope. *ACM Transactions on Graphics, 34(4). doi:10.1145/2766922*

[34] Kazempourradi, S., Hedili, M., K., Soner, B., Cem, A., Ulusoy, E. & Urey, H. (2018). Full-Color Holographic Near-Eye Display with Natural Depth Cues. *in Proc. of the 18th International Meeting of Information Display (IMID)*

[35] Rolland, J. P., Krueger, M. W., & Goon, A. A. (1999). Dynamic focusing in head-mounted displays. *Stereoscopic Displays and Virtual Reality Systems VI.* doi:10.1117/12.349412

[36] Akeley, K., Watt, S. J., Girshick, A. R., & Banks, M. S. (2004). A stereo display prototype with multiple focal distances. *ACM SIGGRAPH 2004 Papers on - SIGGRAPH 04*. doi:10.1145/1186562.1015804

[37] Optics - Imaging - Photonics - Optomechanics - Lasers — Edmund Optics. (n.d.). Retrieved from http://www.edmundoptics.com/

[38] Spitzer, C. R., Ferrell, T., & Ferrell, U. (2014). *Digital Avionics Handbook, Third Edition*. CRC Press.

[39] Wood, E., & Bulling, A. (2014). EyeTab. *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA 14*. doi:10.1145/2578153.2578185

[40] Fitzgibbon, A., Pilu, M., & Fisher, R. (1996). Direct least squares fitting of ellipses. *Proceedings of 13th International Conference on Pattern Recognition.* doi:10.1109/icpr.1996.546029

[41] MATLAB. Retrieved from http://www.mathworks.com/help/stats/boxplot.html

[42] Unity. Retrieved from https://unity3d.com/

[43] OpenCV library. Retrieved from http://www.opencv.org/

[44] Mosca. Retrieved from http://www.mosca.io/