eyeSelfie: Self Directed Eye Alignment using Reciprocal Eye Box Imaging

eyeSelfie: Self Directed Eye Alignment using Reciprocal Eye Box Imaging

Tristan Swedish* Karin Roesch Ik-Hyun Lee Krishna Rastogi Shoshana Bernstein Ramesh Raskar†
MIT Media Lab

Abstract

Eye alignment to the optical system is very critical in many modern devices, such as for biometrics, gaze tracking, head mounted displays, and health. We show alignment in the context of the most difficult challenge: retinal imaging. Alignment in retinal imaging, even conducted by a physician, is very challenging due to precise alignment requirements and lack of direct user eye gaze control. Self-imaging of the retina is nearly impossible.

We frame this problem as a user-interface (UI) challenge. We can create a better UI by controlling the eye box of a projected cue. Our key concept is to exploit the reciprocity, "If you see me, I see you", to develop near eye alignment displays. Two technical aspects are critical: a) tightness of the eye box and (b) the eye box discovery comfort. We demonstrate that previous pupil forming display architectures are not adequate to address alignment in depth. We then analyze two ray-based designs to determine efficacious fixation patterns. These ray based displays and a sequence of user steps allow lateral (x, y) and depth (z) wise alignment to deal with image centering and focus. We show a highly portable prototype and demonstrate the effectiveness through a user study.

CR Categories: I.3.3 [Computer Graphics]: Three-Dimensional Graphics and Realism—Display Algorithms H.5.2 [Information Systems]: User Interfaces—Screen Design

Keywords: retina, fundus photography, head mounted display, human computer interaction

1 Introduction

The importance of precise eye alignment to optical systems has been highlighted by the rise in popularity of consumer head mounted displays (HMDs). In these systems, we define the "eye box" as the orientation in space the eye can move while perceiving a displayed image. The eye box can be thought of as the extreme orientation (rotation and alignment) where the pupil is able to sample all desired ray angles. Many systems are designed to have an eye box that balances light efficiency and ease of alignment. However, exact alignment of HMDs cannot be guaranteed with current display methodologies. Improper alignment can lead to discomfort and nausea.

One of the most challenging alignment tasks is retinal imaging. Traditionally, acquiring retinal images involves complicated, difficult to use and expensive equipment. The devices are designed to be used by a trained operator, need securing of head position, and non-trivial mechanical controls to ensure precise alignment. Furthermore, most retinal imaging techniques require the use of dilation drops to obtain a sufficiently large field of view (FOV). It is the combination of previous factors that make self-imaging of the retina nearly impossible.

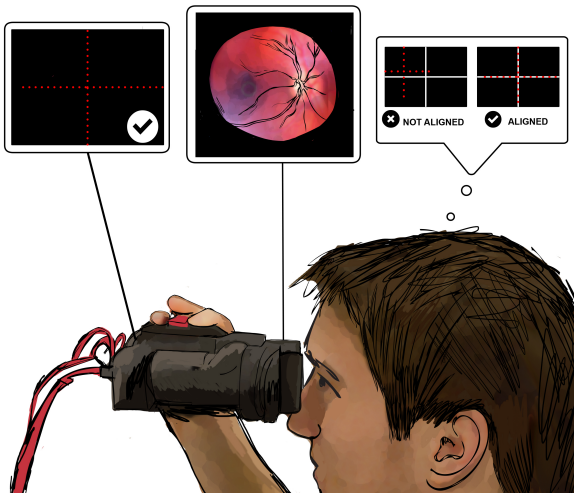

Figure 1: *Self-aligned, mobile, non-mydriatic Fundus Photography. The user is presented with an alignment dependent fixation cue on a ray-based display. Once correctly aligned, a self-acquired retinal image is captured. This retinal image can be used for health, security or HMD calibration.*

Our key concept to drastically simplify retinal imaging is to improve the user interface by developing a framework for tighter eye box control. We develop a near eye alignment display which enables self-alignment by the user, and explore interface methodologies which ensure the precise alignment necessary to obtain self-aligned retinal images. In particular, our approach is to spread different display angles over the area of the pupil so that pupil misalignment produces a different perceived image to the user.

1.1 Contributions

- Our primary contribution is the development of an interactive "ray-cone" approach to allow for self-alignment of eyes. This method enables lateral and axial alignment of the eye to an eyepiece by producing "virtual pinholes" at the user's pupil.
- We constructed a mathematical framework and measurement method for eye box size by evaluating the set of intersecting rays near the eye's pupil.
- We tested this concept by establishing the first interactive system for user aligned retinal photography employing an efficient and low cost hardware design. Our self-alignment strategy for fundus photography makes retinal imaging outside a clinical setting possible.

*e-mail:tswedish@media.mit.edu
†e-mail:raskar@media.mit.edu

We demonstrate that a combination of simple optics and an interactive user interface can be employed for self-imaging of the retina. To the best of our knowledge, this is the first time interactive self-imaging of the retina has been demonstrated.

Benefits and Limitations Our setup avoids many of the pitfalls found in typical fundus camera arrangements by providing a fixation cue that indicates to the user when they are correctly aligned. Our fixation displays use much less light than infrared illumination alignment and other focusing methodologies used in standard retinal photography.

Our alignment strategy assumes the user is able to focus on the fixation target. Our approach has not yet been confirmed for users with visual impairment. Consistently capturing well aligned images with the prototype requires a short (10 minute) training period. The stability of the handheld prototype may vary for different users, however, we attempt to minimize this effect using a short exposure time and user training.

2 Related Work

In recent years, computer graphics researchers have shown an increased interest in exploring the human visual system. Computational models describing cone distribution [Deering 2005], anatomy [Sagar et al. 1994], pupil light reflex [Pamplona et al. 2009], and light transport [W.Y. Lam and V.G. Baranoski 2006], have been formulated to improve the performance of various graphics and display applications. Modeling real-world objects and their corresponding reflective and refractive properties has especially been of interest. Rendering the eye has been widely explored for a great variety of applications, such as modeling realistic video game characters and surrounding objects, refractive errors and cataracts [Cui et al. 2004; Lefohn et al. 2003; Makthal and Ross 2005; Ritschel et al. 2009; Wecker et al. 2005; Pamplona et al. 2010; Pamplona et al. 2011].

Near Eye Displays The computer graphics community has had a sustained interest in near-eye display technologies, which have made virtual (VR) and augmented reality (AR) possible. The purpose of most near eye displays is to transfer the image of a scene onto the retina using the eye's optical system. The human eye can change orientation and accommodate to different focal depths, motivating display designs that can take advantage of the eye's fixation fidelity. Numerous display technologies have been demonstrated using simple optics, pinlights, and light fields [Kress and Starner 2013; Maimone et al. 2014; Lanman and Luebke 2013]. These systems have various levels of transparency for AR applications, depth of field reproduction, and FOV.

Many HMDs have an eye box that allows the user some movement, making alignment easier and accommodates a wide range of interpupillary distances (IPD) without mechanical binocular translation of the system. While advantageous in some applications by adding comfort and usability, a large eye box can sometimes be undesirable [Itoh and Klinker 2015; Plopski et al. 2015]. The large eye box of consumer HMDs allows the user to be aligned imperfectly, causing unwanted distortions in the viewed image and may contribute to user nausea. To date, very few displays are designed to indicate accurate optical alignment to the user. Holographic heads up displays have been used for alignment of firearms to targets [Upatnieks and Tai 1997], but our approach specifically addresses eye alignment to an optical system.

Retinal Biometrics Retinal scanners typically use compact laser scanning systems to measure the blood vessel pattern around the optic disc. Retinal scanners have largely ignored visible light photographic methods. Many of the alignment complexities found in eye imaging have restricted retinal scanners to access control in high security environments [Unar et al. 2014]. Improved eye alignment techniques would make these devices more accessible in other environments. Recent work has focused on making capture devices smaller and more compact [Woittennek et al. 2015] but have largely ignored the alignment problem.

Fixation Cues in Eye Imaging User fixation cues have been used to minimize user movements and to simplify the alignment of physician operated anterior segment and fundus imaging devices. Typically these cues are small targets at the focal plane of the user's eye, such as the tip of the physician's finger. Fixation cues can be separated into *Imaged Eye* or *Fellow Eye* configurations. *Imaged Eye* cues are presented to the eye being imaged, and *Fellow Eye* cues are provided to the eye not being imaged, in which bi-ocular coupling of the human visual system is being exploited to "steer" the eye being imaged [Samaniego et al. 2014; Lawson and Raskar 2014].

Ray based, pupil integrating displays have been used to examine how rays are refracted as they move through the eye's optical system [Pamplona et al. 2011] and for correction of a viewer's refractive error [Huang et al. 2014; Pamplona et al. 2012].

Fundus Photography Indirect ophthalmoscopy allows an observer to examine a subject's retina using handheld or wearable setups [Keeler 1997]. Indirect ophthalmoscopy uses a lens placed in front of the user's eye to magnify features, which are observed through a head-worn, angled-reflector setup [Goldfain et al. 2005].

As an extension of indirect ophthalmoscopy, fundus photography captures an image by using an additional lens that focuses an intermediate image onto a sensor. Today's state-of-the art imaging systems utilize SLR sensors [Ishihara and Kogawa 2008], as well as smaller CMOS sensors [Filar et al. 2011]. Trends in low cost imaging hope to make compact hardware more accessible [iEx 2010; Giardini et al. 2014]. The CavCam uses a consumer camera to enable handheld fundus photography [Tran et al. 2012].

While improvements to image quality has been an active area of research, precise eye alignment requirements remain largely the same. The reason for this is that Fundus Photography makes use of "Maxwellian" illumination to illuminate the retina. This approach uses a cone of light focused through the eye's pupil, the marginal rays of this illumination is seen as dotted red lines in Figure 2. If the center of the pupil moves even a few millimeters away from the illumination cone focus, rays are blocked or reflected from the cornea in undesirable ways.

3 System Overview

We start with the theoretical framework to design a display which ensures accurate self-alignment of the pupil and proper illumination of the retina in our imaging system. HMDs are divided into two distinct classes: pupil forming and non-pupil forming displays [Kress and Starner 2013]. In pupil forming designs, the eyepiece's eye box is determined by the system's exit pupil. These designs relay an image onto the retina through the observer's pupil. The optical design for retinal imaging systems can be thought of as the inverse, in which the image of the retina is relayed through the pupil and onto an image sensor. We develop our retinal imaging theory by inverting the ray model for pupil-forming head mounted displays using the thin lens approximation [Dodgson 2002; Lanman and Luebke 2013]. In short, we want to construct an eye box in the

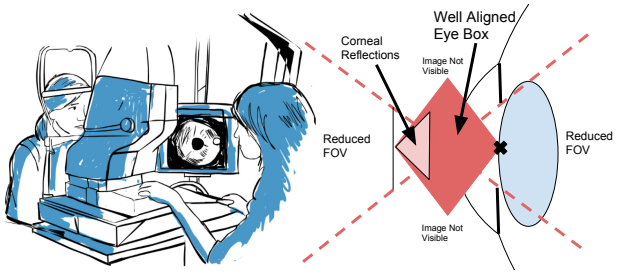

Figure 2: *Fundus Photography. A typical fundus camera is bulky and expensive. Much of the mechanical complexity is to ensure precise eye alignment. The eye box (red diamond) of both a pupil forming display or fundus photography can be defined by the volume the center of the pupil (black "x") is able to move in space without blocking the marginal rays of illumination (dotted red lines). Reflections from the curved surface of the cornea are more localized at the image plane as the eye moves closer to the imaging system (left of figure). If the pupil blocks both marginal rays by moving laterally, the image is blocked. If the pupil moves outside of the eye box volume axially, the FOV is reduced.*

display path that matches the eye box of the illumination path.

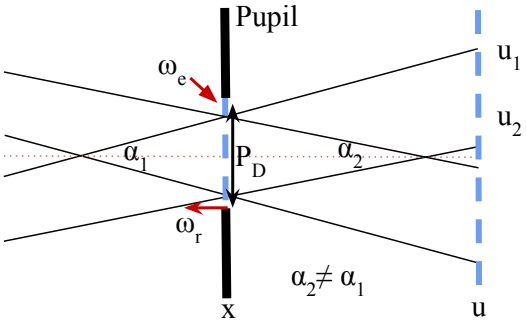

Figure 3: *Geometry of Rays at Pupil. We define a set of rays using the two plane parameterization of the light field at the pupil. The eye can move a certain distance laterally (ω_e), or axially (ω_r) relative to the imaging system. The x-plane at the pupil describes the location of rays in space, and the u-plane describes the direction the rays are headed. We observe the two virtual pinholes at the pupil plane where the generated rays intersect. The user's pupil integrates both of the images generated by these two pinholes when properly aligned.*

3.1 Imaging System

The retina is a great absorber of light, but it is only visible through a small aperture, the pupil. Furthermore, the region in front of the pupil, the cornea, produces strong specular reflections due to the index of refraction mismatch of ambient air and hydrated tissue. This situation motivates us to develop a cohesive illumination strategy to image the retina.

It is useful to model our system using geometrical optics and the thin lens approximation. At a plane along the optical axis, we parameterize the set of rays with two dimensions of space (x, y) and two dimensions of direction (u, v), where $u = \tan(\theta)$ and $v = \tan(\phi)$ are the conversions of the rays from spherical coordinates. The light field at the retinal plane $l(x, y, u, v)$ relates to the illumination path and imaging system by a coordinate transformation.

To simplify our analysis, we will use one spacial dimension and consider the 2D light field (Figure 3), $l(x, u)$, moving through our system. This leads to a simple set of operations which we can perform using basic optical components. First, the light field at some distance, d, along the optical axis can be described by the free space propagation operator:

$$S_d = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \tag{1}$$

This matrix describes a shear in the 2D light field in 1D space, where rays in free space do not change direction and continue the direction they were headed. We can next describe the spherical thin lens operator. This describes another shear operation in u:

$$L_f = \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix} \tag{2}$$

Finally, we notice that an aperture that blocks rays in an area in space, Ω, extinguishes rays over the entire u dimension. We describe this using the pupil operator, P.

$$P_\Omega(l) \rightarrow l(x_\Omega, u) = 0 \tag{3}$$

We can then use this notation to describe light entering the eye and illuminating the retina. The light field at the cornea, l_c, is refracted by the cornea, stopped by the pupil, further refracted by the lens, and travels to the retina. We recover the observed irradiance on the retina by summing over all the incident angles at each point in space.

$$I_r(x) = \sum_u S_e L_e P_\Omega (S_c L_c l_c) \tag{4}$$

We can use this model to understand how light is transformed by the eye in order to construct a system to image it.

Corneal Reflections The reflections of the illumination engine back into the imaging system can produce undesirable out of focus reflections, reducing the noise floor of acquired images. We model corneal reflections as the image produced by a convex spherical mirror with a radius of curvature of R=7.2mm. We calculate the rays which will reflect back into our imaging system using Equations 1 and 2. The 2D light field after reflecting off the cornea and reaching the objective lens is:

$$\begin{bmatrix} x' \\ u' \end{bmatrix} = \begin{bmatrix} x(1 - \frac{f_L}{f_c}) + f_L u \\ -\frac{x}{f_c} + u \end{bmatrix} \tag{5}$$

where f_L is the focal length of the objective lens and f_c is the effective focal length of the reflective cornea defined by $f_c = -R/2$. We then determine the light field at the image plane behind the objective.

$$\begin{bmatrix} x'' \\ u'' \end{bmatrix} = \begin{bmatrix} f_L(-\frac{x}{f_c} + u) \\ x(\frac{1}{f_c} - \frac{2}{f_L}) + u(1 + f_L) \end{bmatrix} \tag{6}$$

We note that in order for the rays to stay in the system, $|x'| < D/2$, where D is the diameter of the objective lens. Similarly, $|x''|$ must

be within the FOV of the camera. We can thus write the bound on u in the light field as a function of x using Equations 5 and 6 to solve the inequality imposed by the objective lens and camera apertures.

$$-\frac{D}{2f_L}+\frac{x}{f_c}<u<\frac{D}{2f_L}+\frac{x}{f_c} \tag{7}$$

$$-\frac{D}{2f_L}-x\left(\frac{1}{f_L}-\frac{1}{f_c}\right)<u<\frac{D}{2f_L}-x\left(\frac{1}{f_L}-\frac{1}{f_c}\right) \tag{8}$$

We take the intersection of these two regions to determine the area of the light field at the cornea, which will reflect back into the system.

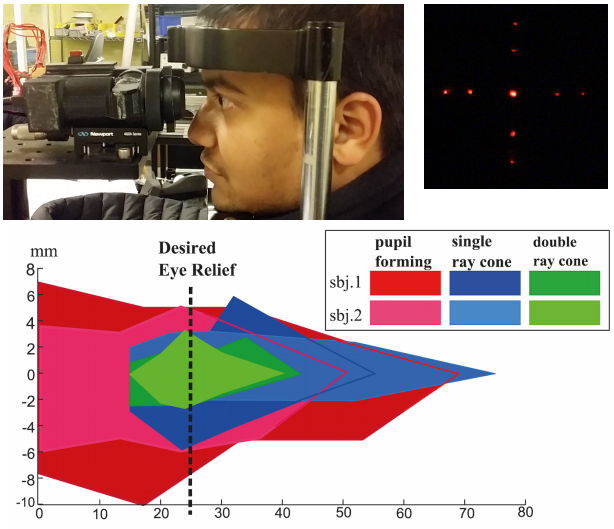

Figure 4: *Measuring Eye Alignment. Top: The subject was asked to report when the perceived image (top right) was disrupted after moving the device on the translation stage. Bottom: The actual eye box was measured using three different display methodologies for lateral (vertical axis) and axial (horizontal axis) translation of the observing eye relative to the imaging system. The system was moved relative to the subjects stationary head position. Each subject started 100mm from the imaging system and was asked to report when they could view the display and when it disappeared upon moving too close. The subjects were then asked to report the lateral offset in which occlusion of the display occurred for five evenly spaced axial distances from these two extreme points. The dotted line is the desired eye relief, the distance from the imaging system which would allow the marginal rays of the illumination to maximally clear the pupil.*

We observe that converging rays towards the center of the imaging axis (x and u opposite signs and small) account for most reflections back into the system. This analysis explains the motivation for the annular type illumination popular in mydriatic fundus cameras. Non-mydriatic systems such as ours must still deliver much of the light through a smaller pupil, therefore cross polarized illumination is used to reduce reflections at the image plane. Furthermore, this analysis suggests that the eye should be further than the focus of converging illumination to avoid localized reflections, which we confirm with experimental observation in Section 4.3 (see Figure 5).

Retinal Illumination The pupil restricts the spatial extent of incident light that makes it to the retina. The pupil diameter contracts with increased ambient light. The relaxed pupil size varies from person to person. Quality retinal images require even illumination across the fundus. It is this series of factors which suggest a converging beam focused onto the pupil as the best illumination strategy. This is called Maxwellian illumination, where the full-cone angle is denoted as α. One large drawback to Maxwellian illumination is a greatly reduced eye box. We solve this problem by introducing an alignment dependent display and by establishing a stimulus to direct the user into the illumination eye box as accurately as possible.

3.2 Pupil Alignment

Correct positioning of the eye in space is critical and needs to address six degrees of freedom: three translational, two rotational and one focal. Eye alignment could be achieved by creating an image which is only visible to the user when they are in the illumination eye box. Our approach is to create a display which matches our illumination eye box. Furthermore, the display eye box should be restricted to ensure that the user does not move closer than the illumination focus, which would cause unwanted corneal reflections.

Pupil forming displays have the ability to control the relative rotation given a fixed location in space. They address rotational and focal degrees of freedom of the eye with limited ability of the viewer to know where their eye is located in space. In order to address spatial degrees of freedom for alignment we must use a ray based approach, which we call a **single ray cone** design. These kind of displays create a ray of cones along the marginal rays of a focused beam such as our Maxwellian illumination source (see Figure 2). This is very effective for constraining lateral movement (x, y), but it does not fully address the problem of alignment in depth (z). In order to address this final degree of freedom, we introduce another set of converging rays, which when combined with a single ray cone becomes a double ray cone.

The marginal rays that describe the ray cone have an angle of extent, α. The eye box for this illumination source, ω_e, can be found as a function of z:

$$\omega_e(z)=P_D-2\left|z-z_0\right|u_\alpha \tag{9}$$

where for an extended angle, α, $u=\tan(\frac{\alpha}{2})$ and P_D is the diameter of the subject's pupil. We then can find the maximum axial distance the eye, z_r, can be from the ray intersection at z_0 by solving for when $\omega_e=0$, or:

$$z_r=\frac{P_D}{2u_\alpha} \tag{10}$$

We observe that in order to achieve a full FOV, the marginal rays of the illumination beam must pass the pupil. The marginal rays must be distinct from the rest of the rays entering the pupil to indicate to the user that they have made it through. The simplest solution is to produce only the marginal rays of the illumination. If the user sees them, then we can assume the system is well aligned. Ideally, we want to further constrict the eye box of the display so that we avoid situations where the marginal rays may scatter off the edge of the pupil. To this end, we want to pick a set of rays which produces an eye box that is smaller than the illumination eye box, ω_e. The set of rays is shown in Figure 3. We also note that when the eye is focused to infinity, u must be unique for each ray in order to avoid

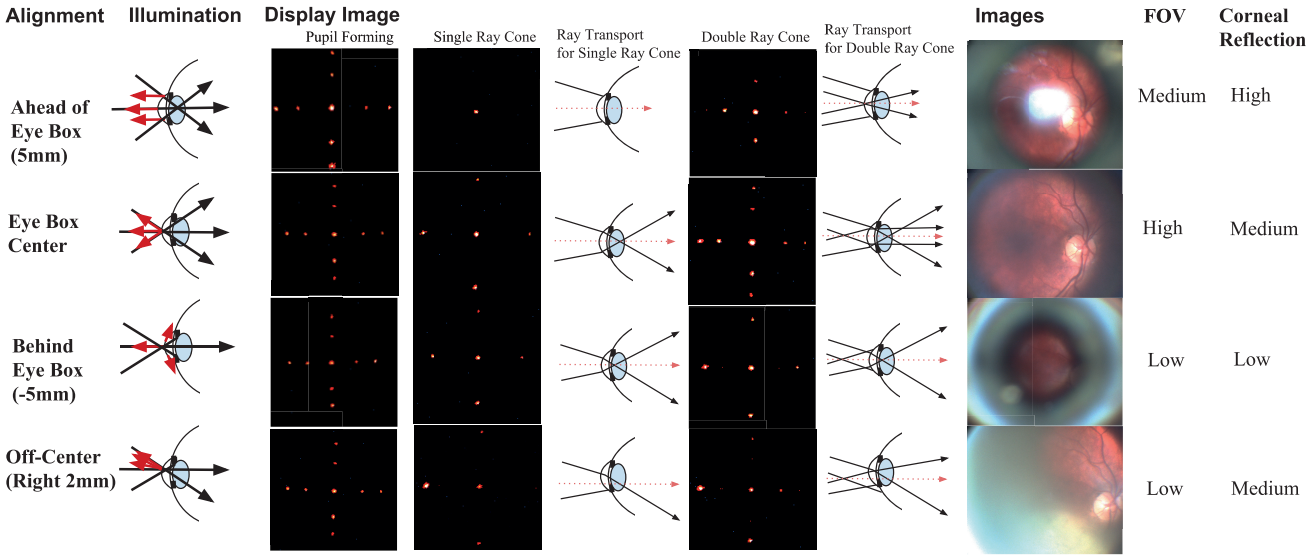

Figure 5: *Physical Alignment and Perceived Display. Starting from the left. Alignment: There are four primary locations the eye can be relative to the eye box center; well aligned, too close and too far from the instrument, and off center. Illumination: A collimated illumination source is focused by the objective lens to a point near the eye pupil, reflected light behaves differently depending on the location of the eye relative to the illumination focus. Display Image: Three alignment strategies are used to match the display to the illumination eye box. Each position produces a distinct type of retinal image for the two ray cone pattern. Images: Reflections from the cornea appear as localized bright spots due to the cornea re-collimating reflected rays. The pupil blocks marginal rays that are too far or off center. FOV and Corneal Reflection: The FOV and reflection pattern are highly specific to the eye location relative to the instrument.*

ambiguity, since an eye focused to infinity will map each ray with the same u to the same location on the retina. Our set of rays is:

$$l_c = \begin{pmatrix} \frac{P_D}{2} & \frac{P_D}{2} & -\frac{P_D}{2} & -\frac{P_D}{2} \\ -u_1 & u_2 & u_1 & -u_2 \end{pmatrix} \tag{11}$$

Where P_D is the user's pupil diameter and (u_1, u_2) are two sufficiently distinct ray angles that make it through the imaging system.

The set of rays defined by equation 11 is shown in Figure 3, representing a **double ray cone** pattern with ray intersections at different distances from the eyepiece. As such, to define the eye box of this design, we simply take the intersection of equation 9 applied to u_1 and u_2.

We now describe our display optics using a combination of equations 1 and 2 for each optical element, which produces a 2×2 ray transfer matrix T for the 2D light field case.

$$T = S_1 L_1 S_2 L_2 ... S_n L_n \tag{12}$$

The light field at the display, l_d, propagates through our optical system, and produces a light field at the eye pupil $l_c = T\ l_d$. We can then solve this equation for the desired light field at our display, ensuring that the resulting light field clears the system's exit pupil P_E.

$$l_d = T^{-1} l_c \tag{13}$$

We observe that the double cone approach addresses all the degrees of freedom in eye alignment. We use this approach to generate our desired light field at the pupil. Figure 3 illustrates this light field and Figure 5 shows the ray transport for the four distinct eye alignment scenarios. This ray pattern could be interpreted as five virtual pinholes at the eye's pupil plane, allowing for spacial coding of the perceived image as the user's pupil samples different combinations of these virtual pinholes. We construct the light field in our display using two masks 12mm apart, where the mask closest to the magnifier lens is one focal length away in order to provide a focus cue for the user. Our analysis could be applied to other light field displays such as those formed by lenslet arrays, but we elected to use a pair of pinhole masks to allow for finer angular control of the generated light field.

4 Implementation

A simple retinal imaging device, eyeSelfie (eye self imaging), was built to explore the advantages of using alignment dependent fixation targets in fundus photography. We created our prototype using CAD software and 3D printed the optical and electronics housing.

4.1 Hardware

Our optical design is similar to a mydriatic add-on to a point and shoot camera [Tran et al. 2012] that uses collimated cross-polarized illumination and an ophthalmic objective lens to focus light through the pupil and form an image of the retina. Our additions include a beamsplitter to produce a pupil-forming display in the optical axis (see Figure 7), and the use of a point source instead of an annular pattern to make imaging without dilation drops possible. Our illumination and imaging electronics are synchronized by a program written in C++ running on a laptop connected via USB.

We use a 40D ophthalmic lens from Volk Optical Inc. as the objective lens at the front of the device. This lens relays the retinal plane from within the eye to a plane behind it. Following Figure 7, we then separate the optical layout of our device into three sections.

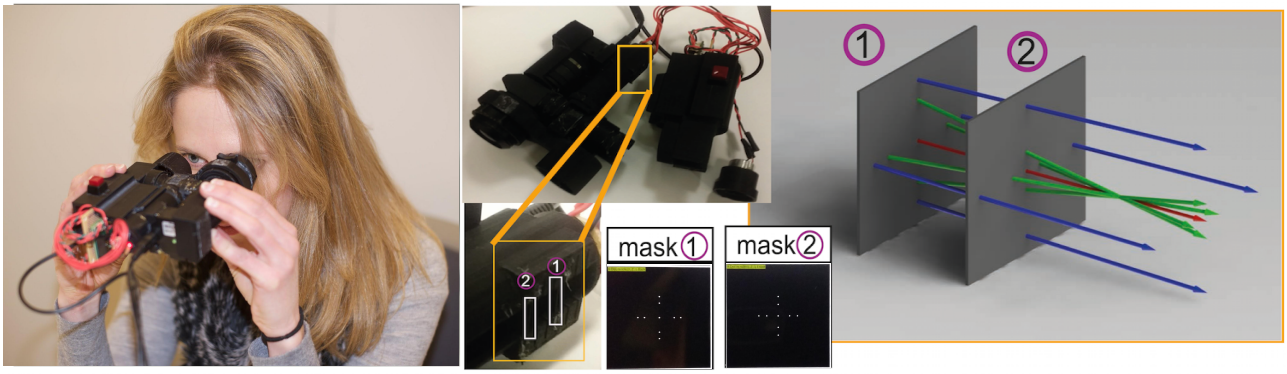

Figure 6: *Implementing Display in Hardware. Left: eyeSelfie in use being aligned while seated at a table. Center: Upon opening the device, mask placement and calculated mask pattern can be illustrated. Right: Ray diagram of two masks restricting the light from a near lambertian light box to a set of rays which make up the light field at the display.*

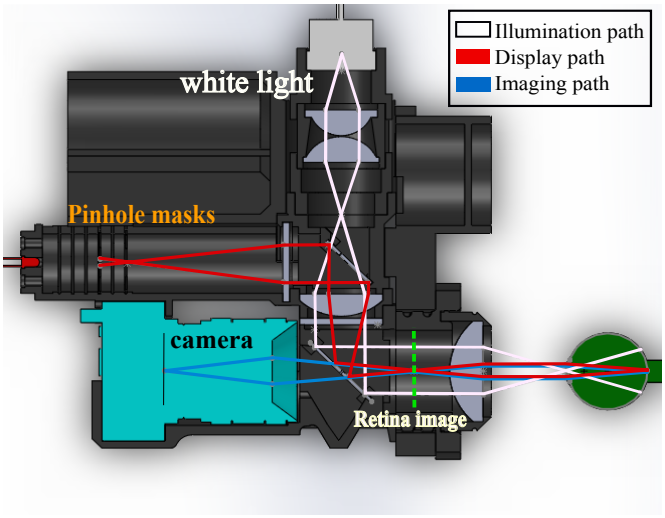

Figure 7: *Optical Layout. eyeSelfie consists of three primary optical paths. Illumination: A white light source is triggered by the user after they align with the display. Display: A display is created by a pair of pinhole masks. Imaging Path: The optical paths are aligned so that if the display is visible, the camera is well aligned. The image of the retina captured by the camera is inside the instrument.*

Illumination A 4000K white LED is controlled by a flash driver (Texas Instruments LM3644) connected to an Arduino Micro using I^{2}C. The Arduino Micro is connected to a laptop, and our C++ controller program sends flash trigger signals over USB. In normal operation, the flash driver is set to produce one 10ms flash per exposure with an illumination energy of $81 \times 10^{-6} \frac{\text{Joules}}{\text{cm}^2}$ incident on an aphakic eye in the worst case, which is over 3 orders of magnitude below the maximum permissible exposure (MPE) limits for Maxwellian illumination sources (ISO 15004-2:2007 and ISO 10940:2009).

The LED is placed behind a pair of lenses that relay its image one focal length away from a 25mm focal length plano-convex condenser lens (Edmund Optics #45-097). The collimated beam is then inserted into the imaging path with a polarizing beamsplitter. A linear polarizer is placed in front of the light source and is configured so that the polarizing beam splitter only allows light with a different polarization state back through the system. This cross polarized configuration reduces undesirable specular reflections.

Display The display path is inserted into the illumination path right before the condenser lens by a 90T/10R plate beamsplitter. The beamsplitter is followed by a 50mm magnifier lens which produces a plane conjugate to the retina. The 90T/10R beamsplitter behaves as the exit pupil of the system, defining the eye box of the display plane. A pair of printed transparency masks are placed 12mm apart, with the closest mask one focal length from the magnifier lens.

Imaging A USB camera (Point Grey FMVU-13S2C) with a 16mm c-mount lens is placed behind the polarizing beamsplitter and focused on the retinal image plane. A C++ program running on the connected laptop synchronizes the shutter and illumination triggers.

4.2 Software

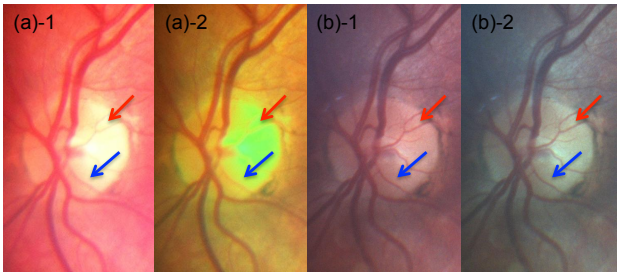

Figure 8: *Sample images acquired by Pictor and prototype: (a) Pictor images, (a)-1 normal color image, (a)-2 color corrected image with normalized white balance and brightness, (b) prototype images, (b)-1 normal color image, (b)-2 white balanced image; blue and red arrows indicate blood vessels at the optic disc.*

In order to assess our image quality, we acquire images with eyeSelfie and the commercially available digital ophthalmic imager Volk Pictor [pic 2014], shown in Figure 8. The Pictor provides two image correction modes; a raw image (Fig. 8(a-1)) and a color corrected image with normalized white balance and brightness (Fig.

8(a-2)). We processed the prototype images in a similar fashion, producing a normal (Fig. 8(b-1)) and color corrected image (Fig. 8(b-2)) for each acquisition. The prototype produces good images of the retinal blood vessels at the optic disc, while the Pictor saturates these blood vessel in the normal image. In addition, blood vessels cannot be seen in the color corrected image of Pictor due to the saturation.

We measured the image sharpness using image focus [Lee et al. 2013b; Lee et al. 2013a]. Focus quality is assessed using 10 different focus measures (FM). The 10 FM used are: Brenner's focus measure (MIS2), spatial frequency (MIS8), thresholded absolute gradient (GRA3), histogram entropy (STA7), energy of Laplacian (LAP1), modified Laplician (LAP2), diagonal Laplacian (LAP3), varian of Laplacian (LAP4), sum of wavelet coefficients (WAV1), variance of wavelet coefficients (WAV2) [Pertuz et al. 2013]. We compare both the normal and corrected images of eyeSelfie and the Pictor. Our prototype device provides the best image quality except 2 FMs (GRA3, STA7) out of 10.

4.3 Prototype Results

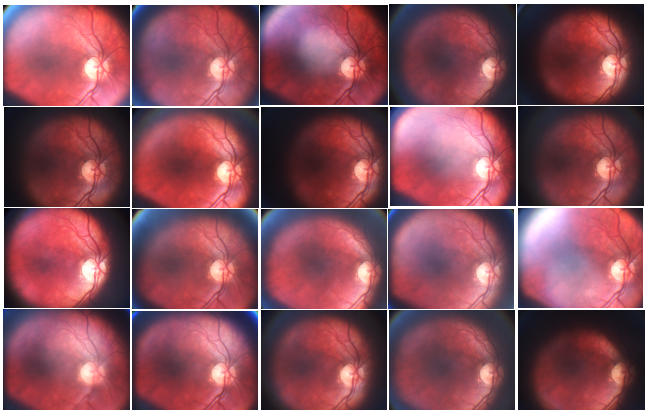

Figure 9: *Consistent Image Capture. Twenty consecutive images taken by a familiarized user approximately every 30 minutes. Every image produced by eyeSelfie during this time period is shown, demonstrating the consistency at which a familiar user can self-perform precise eye alignment tasks using the techniques developed.*

We demonstrate the effectiveness of our alignment setup by conducting a simple alignment task for a group of 10 volunteers. Each individual was asked to sit down and hold the device like a pair of binoculars while resting their elbows on a table. They were then asked to move the device relative to their head until they could see 9 points oriented in a cross shape. Once the volunteer was confident they could see all 9 points, they were asked to fixate on the point that appeared closest to their nose, and push the capture button. This nasal fixation was designed to center the frame so that the macula and optic disc could be observed in the captured images. The capture button triggered the 10ms flash and shutter of the camera. The short flash interval and device operation instructions were designed to compensate for the reduced stability of a hand held device.

We consulted three independent ophthalmologists to assess the quality of our alignment from the images generated by the prototype. We asked the ophthalmologists to rate images captured by familiar users, familiar users with refractive errors, and novice users without refractive errors for alignment and focus. We define "full alignment" as a full FOV image with no localized corneal reflection, and "partial alignment" as some retinal features visible even if occluded by corneal reflections.

Novice Users When first introduced to the device, some users had difficulty understanding and carrying out the directions (discussed in Section 5). We define novice users as those who have taken less than five images of their own eye. Delivering directions and capturing five images takes approximately 10 minutes. After each image was captured, novice users were asked to observe the resulting image. We found this user feedback to be essential in familiarizing a user. We found that 46% of novice users achieved full alignment and 79% achieved at least partial alignment on a per capture basis. Novice users seemed to have difficulty focusing on the fixation target, 13% of the images were in focus.

Familiar Users More regular users of the device quickly overcame initial challenges, Figures 10 and 9 show consecutive images captured for three users familiar with the device. Alignment can be performed reliably in consecutive images taken by a familiarized user. We see that most images are of relatively high quality with a FOV ranging from 15 to 30 degrees. familiar users with uncorrected refractive errors underperformed, they achieved 22% full alignment and 72% with at least partial alignment, while 11% of the images were in focus. Future work could correct this disparity by allowing adjustment of the display accommodation. In contrast, familiar users with normal vision captured images that were fully aligned 75% of the time with all images achieving at least partial alignment. 83% of the images captured by familiar users were in focus.

In some cases, users were able to achieve good alignment laterally, but had difficulty with the depth-dependent two ray cone pattern. We noticed that users with refractive errors would consistently move too close to the device, producing corneal reflections in the captured images. Furthermore, users with larger pupils have greater freedom of movement and were not able to achieve as precise alignment because our fixation target was designed for a 2mm pupil. A programmable display would allow Equation 11 to be calculated for every individual pupil size and enable the construction of the ideal ray pattern.

This study confirms that highly accurate alignment can be achieved using this prototype display. These alignment cues are not specific for our prototype alone, they could be applied to other optical instruments that require precise eye alignment.

5 Discussion

There are two aspects which underlie eye imaging. The display eye box must match the illumination eye box while at the same time being comfortable to find. Even though initially easier for users, large eye boxes have the disadvantage of being less light efficient and allowing distortions in the intended image for off axis viewing. Our methodology uses the concept of *overlapping* eye boxes to indicate to the user how to correct their misalignment.

Some challenges still exist for our system. We hope that future authors in the graphics community could help facilitate a better experience for first time "novice" users. A small amount (10mins) of training time is required to familiarize users with the device. In particular, users need to be coached to reduce squinting, to avoid involuntary shifting of the device during image capture, and need help with easing first time use anxiety. After becoming familiar with eyeSelfie, users become adept at obtaining consistent pictures even with lapses of time between image captures.

There are other vision related issues such as cataracts, severe refractive errors, and loss of accommodation that remain untested.

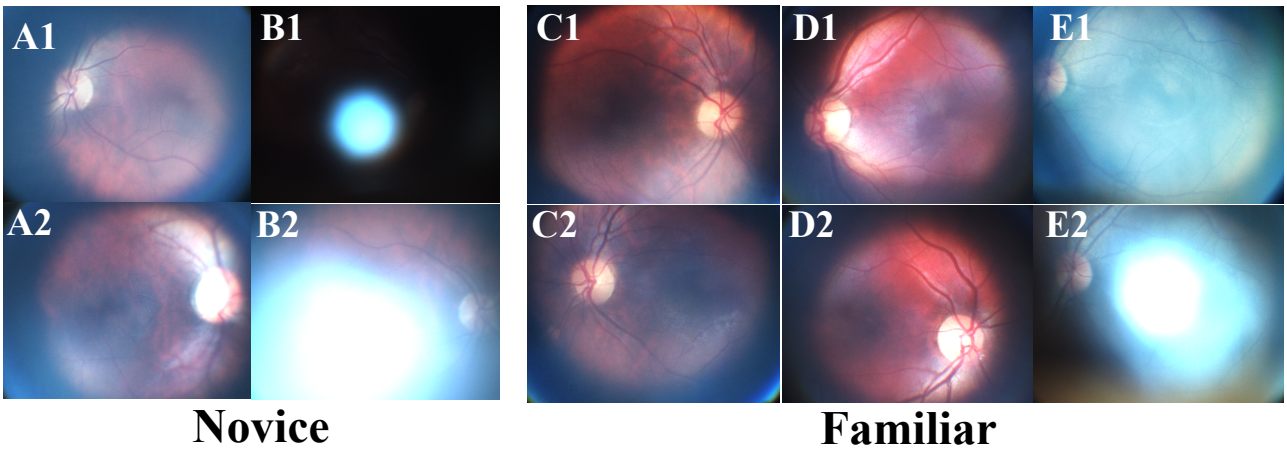

Figure 10: *Raw Images from User Study. Images captured by novice (A-B) and familiarized users (C-E). Some novice users achieved good alignment (A), however a large portion of these images were out of focus (A2). Novice users sometimes were poorly aligned (B). Familiar users were much more capable of capturing well aligned and in focus images (C-D). Users with refractive errors (E) tend to align closer to the device than desired, producing localized corneal reflections (E2).*

We observed that some familiar users with refractive errors consistently aligned themselves ahead of the eye box. We hypothesize this is a result of the individual's refractive error.

5.1 Future Directions

A user aligned fundus camera opens the possibility for studies that allow more regular measurements of the retina and increases its availability as a diagnostic tool.

Changes in vasculature morphology are indicative of health status. Fundus photography is essential in the diagnosis of blindness causing diseases such as glaucoma, age-related macular degeneration, and diabetic retinopathy. Furthermore, recent research has suggested that fundus images can be used to identify risk factors for heart disease, stroke, Alzheimers disease, and multiple sclerosis. However, many people currently have limited or no access to standard retinal imaging devices. Regular screening is infeasible. In order to realize the potential of the retina to indicate the quality of our health, accessible imaging tools are necessary.

An image of the retina with a large FOV could be constructed as a mosaic of images captured with some set of angular displacements. Figure 11 demonstrates a retinal image mosaic made from images captured while the user was fixating on the four points at the edges of the two ray cone display. Detailed description of the mosaicking process can be found in [Lee and Choi 2013]. Greater angular displacements could be achieved by increasing the FOV of the display. The virtual pinholes formed at the pupil plane of the eye ensure that if the display is visible, the illumination and imaging path will be able to capture images with large angular displacements. As such, even if the internal display has a limited FOV, as long as the displayed image is visible to the user at the periphery of their vision, an image at that angular displacement can be captured. Exploiting this property is a topic for future work.

Once alignment has been achieved, corneal reflections and other artifacts can be removed through novel sensors and computational techniques [Kadambi et al. 2013].

The alignment demonstrated by eyeSelfie also has biometric applications. By solving alignment tasks using view dependent displays, secure identity verification could be made more widely available.

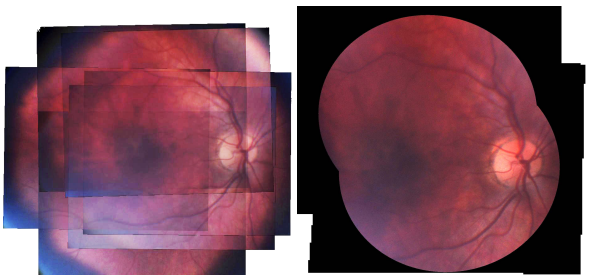

Figure 11: *Image Mosaic to improve FOV. Stitched image mosaic using eyeSelfie images made using a number of images captured while the user is looking at the edge of the display. Combining multiple images can allow the creation of an image with a larger FOV.*

6 Conclusion

We present a user aligned retinal imaging system that uses an alignment dependent display on the same axis as the imaging system. This enables users to capture high resolution images of their retina without the help of a trained operator. Users familiar with the device show great consistency in the results obtained.

User-aligned fundus photography is now possible. This will allow new paradigms of care and lead to more regular monitoring of systemic disease and vasculature health.

We hope our work will broaden access to retinal imaging and have a significant impact on global health. We envision our technology to be connected to the cloud, which allows for integration of retinal imaging data with other health data. Our technology will enable retinal imaging to occur in a more frequent manner and outside of the clinical setting. "Interval-based" retinal monitoring will allow creation of baselines and trends, and can provide new opportunities

for detection, diagnosis and treatment of ocular and systemic diseases. Our user-friendly high-throughput instrumentation will not only allow the collection of patterns for one individual, but also enable large-scale data collection and comparison of patterns across subpopulations. Big data analytics (machine learning) on longitudinal and cross sectional studies will not only allow for a comparison to a "static" baseline (e.g. gender, age) but to a profile with many metrics (ethnicity, genotype, geography).

We hope the SIGGRAPH community will be motivated to develop this approach further.

Acknowledgements

We would like to thank our volunteers for their participation in the study, R. Daniel Ferguson at Physical Sciences Inc. for advice and insightful discussions on optics development, the members of the Camera Culture Group at the MIT Media Lab for their support, our clinical collaborators (PS, TM, JP) for rating images and assessing the viability of our design; and the reviewers for their valuable feedback. Illustrations are drawn by Laura Piraino. This work is supported by the Vodafone Americas Foundation, the MIT Deshpande Center for Technological Innovation, and funding from the US Army Research Laboratory's Army Research Office.

References

CUI, J., WANG, Y., HUANG, J., TAN, T., AND SUN, Z. 2004. An iris image synthesis method based on pca and super-resolution. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 4, IEEE, 471–474.

DEERING, M. F. 2005. A photon accurate model of the human eye. *ACM Trans. Graph. 24*, 3, 649–658.

DODGSON, N. A. 2002. Analysis of the viewing zone of multiview autostereoscopic displays. In *Electronic Imaging 2002*, vol. 4660, International Society for Optics and Photonics, 254 – 265.

FILAR, P., ET AL., 2011. Apparatus for photographing the anterior segment and retina of the eye through the use of a camera attachment. US Patent 7,883,210.

GIARDINI, M., LIVINGSTONE, I., JORDAN, S., BOLSTER, N., PETO, T., BURTON, M., AND BASTAWROUS, A. 2014. A smartphone based ophthalmoscope. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, 2177–2180.

GOLDFAIN, E., LAGERWAY, W., ROBERTS, C., SLAWSON, S., AND KRAUTER, A., 2005. Eye viewing device for large field viewing. US Patent 6,939,006.

HUANG, F., WETZSTEIN, G., BARSKY, B., AND RASKAR, R. 2014. Eyeglasses-free Display: Towards Correcting Visual Aberrations with Computational Light Field Displays. *ACM Trans. Graph. 33*, 4, 1–12.

2010. iExaminer for Panoptic. `http://www.iexam.com/`.

ISHIHARA, M., AND KOGAWA, T., 2008. Fundus camera. US Patent 7,377,642.

ITOH, Y., AND KLINKER, G. 2015. Light-field correction for spatial calibration of optical see-through head-mounted displays. *Visualization and Computer Graphics, IEEE Transactions on 21*, 4, 471–480.

KADAMBI, A., WHYTE, R., BHANDARI, A., STREETER, L., BARSI, C., DORRINGTON, A., AND RASKAR, R. 2013. Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM Trans. Graph. 32*, 6, 167.

KEELER, C. 1997. 150 years since babbage's ophthalmoscope. *Archives of ophthalmology 115*, 11, 1456.

KRESS, B., AND STARNER, T. 2013. A review of head-mounted displays (hmd) technologies and applications for consumer electronics. vol. 8720, SPIE, 87200A–87200A–13.

LANMAN, D., AND LUEBKE, D. 2013. Near-eye light field displays. *ACM Trans. Graph. 32*, 6.

LAWSON, E. M., AND RASKAR, R. 2014. Smart phone administered fundus imaging without additional imaging optics. *Investigative Ophtalmology and Visual Science 55*, 5, 1609.

LEE, I.-H., AND CHOI, T.-S. 2013. Accurate registration using adaptive block processing for multispectral images. *Circuits and Systems for Video Technology, IEEE Transactions on 23*, 9, 1491–1501.

LEE, I.-H., SHIM, S.-O., AND CHOI, T.-S. 2013. Improving focus measurement via variable window shape on surface radiance distribution for 3d shape reconstruction. *Optics and Lasers in Engineering 51*, 5, 520–526.

LEE, I., TARIQ MAHMOOD, M., AND CHOI, T.-S. 2013. Adaptive window selection for 3d shape recovery from image focus. *Optics & Laser Technology 45*, 21–31.

LEFOHN, A., BUDGE, B., SHIRLEY, P., CARUSO, R., AND REINHARD, E. 2003. An ocularist's approach to human iris synthesis. *Computer Graphics and Applications, IEEE 23*, 6, 70–75.

MAIMONE, A., LANMAN, D., RATHINAVEL, K., KELLER, K., LUEBKE, D., AND FUCHS, H. 2014. Pinlight displays: Wide field of view augmented reality eyeglasses using defocused point light sources. *ACM Trans. Graph. 33*, 4, 89:1–89:11.

MAKTHAL, S., AND ROSS, A. 2005. Synthesis of iris images using markov random fields. In *Proc. 13th European Signal Processing Conf*, Citeseer.

PAMPLONA, V., OLIVEIRA, M., AND BARANOSKI, G. 2009. Photorealistic models for pupil light reflex and iridal pattern deformation. *ACM Trans. Graph. 28*, 4, 106.

PAMPLONA, V., MOHAN, A., OLIVEIRA, M., AND RASKAR, R. 2010. Netra: interactive display for estimating refractive errors and focal range. *ACM Trans. Graph. 29*, 4, 77.

PAMPLONA, V., PASSOS, E., ZIZKA, J., OLIVEIRA, M., LAWSON, E., CLUA, E., AND RASKAR, R. 2011. Catra: interactive measuring and modeling of cataracts. *ACM Trans. Graph. 30*, 4, 47.

PAMPLONA, V. F., OLIVEIRA, M. M., ALIAGA, D. G., AND RASKAR, R. 2012. Tailored displays to compensate for visual aberrations. *ACM Trans. Graph. 31*, 4 (July), 81:1–81:12.

PERTUZ, S., PUIG, D., AND GARCIA, M. A. 2013. Analysis of focus measure operators for shape-from-focus. *Pattern Recognition 46*, 5, 1415–1432.

2014. Volk Pictor Plus Portable Retinal Camera. `http://www.volk.com/`.

PLOPSKI, A., ITOH, Y., NITSCHKE, C., KIYOKAWA, K., KLINKER, G., AND TAKEMURA, H. 2015. Corneal-imaging

calibration for optical see-through head-mounted displays. *Visualization and Computer Graphics, IEEE Transactions on 21*, 4, 481–490.

RITSCHEL, T., IHRKE, M., FRISVAD, J., COPPENS, J., MYSZKOWSKI, K., AND SEIDEL, H. 2009. Temporal glare: Real-time dynamic simulation of the scattering in the human eye. In *Computer Graphics Forum*, vol. 28, 183–192.

SAGAR, M., BULLIVANT, D., MALLINSON, G., AND HUNTER, P. 1994. A virtual environment and model of the eye for surgical simulation. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, ACM, 205–212.

SAMANIEGO, A., BOOMINATHAN, V., SABHARWAL, A., AND VEERARAGHAVAN, A. 2014. mobilevision: A face-mounted, voice-activated, non-mydriatic lucky ophthalmoscope. In *Proceedings of the Wireless Health 2014 on National Institutes of Health*, ACM, 1–8.

TRAN, K., MENDEL, T. A., HOLBROOK, K. L., AND YATES, P. A. 2012. Construction of an inexpensive, hand-held fundus camera through modification of a consumer "point-and-shoot" camerapoint-and-shoot hand-held fundus camera. *IOVS 53*, 12, 7600.

UNAR, J. A., SENG, W. C., AND ABBASI, A. 2014. A review of biometric technology along with trends and prospects. *Pattern Recognition 47*, 8 (8), 2673–2688.

UPATNIEKS, J., AND TAI, A. M. 1997. Development of the holographic sight. vol. 2968, SPIE, 272–281.

WECKER, L., SAMAVATI, F., AND GAVRILOVA, M. 2005. Iris synthesis: a reverse subdivision application. In *Proceedings of the 3rd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, ACM, 121–125.

WOITTENNEK, F., KNOBBE, J., PUGNER, T., DALLMANN, H.-G., SCHELINSKI, U., AND GRUGER, H. 2015. Mems scanner mirror based system for retina scanning and in eye projection. In *Proceedings of the SPIE*, vol. 9375, 937506.

W.Y. LAM, M., AND V.G. BARANOSKI, G. 2006. A predictive light transport model for the human iris. *Computer Graphics Forum 25*, 3, 359–368.