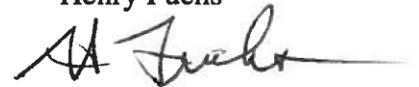# COMPUTATIONAL SEE-THROUGH NEAR-EYE DISPLAYS

Andrew S. Maimone

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.
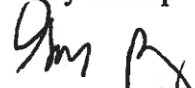
Chapel Hill
2015

Approved by:

Henry Fuchs

Gary Bishop

Jan-Michael Frahm

Douglas Lanman

David Luebke

ABSTRACT

Andrew S. Maimone: Computational See-Through Near-Eye Displays
(Under the direction of Henry Fuchs)

See-through near-eye displays with the form factor and field of view of eyeglasses are a natural choice for augmented reality systems: the non-encumbering size enables *casual and extended use* and large field of view enables *general-purpose* spatially registered applications. However, designing displays with these attributes is currently an open problem. Support for enhanced realism through mutual occlusion and the focal depth cues is also not found in eyeglasses-like displays.

This dissertation provides a new strategy for eyeglasses-like displays that follows the principles of computational displays, devices that rely on software as a fundamental part of image formation. Such devices allow more hardware simplicity and flexibility, showing greater promise of meeting form factor and field-of-view goals while enhancing realism. This computational approach is realized in two novel and complementary see-through near-eye display designs. The first *subtractive* approach filters omnidirectional light through a set of optimized patterns displayed on a stack of spatial light modulators, reproducing a light field corresponding to in-focus imagery. The design is thin and scales to wide fields of view; see-through operation is achieved with transparent components placed directly in front of the eye. Preliminary support for focal cues and environment occlusion is also demonstrated. The second *additive* approach uses *structured point light illumination* to form an image with a near minimal set of rays. Each of an array of defocused point light sources is modulated by a region of a spatial light modulator, essentially encoding an image in the focal blur. See-through operation is also achieved with transparent components, and thin form factors and wide fields of view ($\geq 100°$ diagonally) are demonstrated.

The designs are examined in theoretical terms, in simulation, and through prototype hardware with public demonstrations. This analysis shows that the proposed computational near-eye display designs offer a significantly different set of trade-offs than conventional optical designs. Several challenges remain to make the designs practical, most notably addressing diffraction limits.

To Wendy

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AR   Augmented Reality

CPD   Cycles per Degree

FOV   Field of View

GPU   Graphics Processing Unit

HMD   Head Mounted Display

IPS   In-Plane Switching

ISO   Industry Organization for Standardization

LCD   Liquid Crystal Display

LED   Light Emitting Diode

OLED   Organic Light Emitting Diode

PC   Personal Computer

PSNR   Peak Signal to Noise Ratio

RGB   Red, Green, and Blue

SLM   Spatial Light Modulator

TIR   Total Internal Reflection

VA   Vertically Aligned

VR   Virtual Reality

CHAPTER 1

Introduction

1.1   A Vision for Augmented Reality

Decades of augmented reality (AR) research paint an exciting portrait of the future. Computers and displays of all types – smartphones, tablets, laptops, desktops, TVs, theaters – are replaced with a persistent interface that travels with the user. Augmented objects are stored and retrieved at will, manipulated naturally with the user's hands and positioned anywhere – taking advantage of the user's spatial memory. These objects approach the visual fidelity of real objects and replace many real objects without mechanical function.

Users retrieve fragments of spatially and context-aware information throughout the day – the identity of people and objects, directions, warnings of imminent danger – that they would not ordinary have the time or capacity or awareness to access through conventional means. This stream of information helps the user learn about the world, and the computing platform learns the relevance to the user.

The appearance of the world is programmable, and can be customized by each user. It can be optimized for aesthetics or function, or can take on an entirely new appearances. The whole environment becomes a giant workspace or playfield. These spaces can be personal or shared with others – allowing multiple superimposed virtual worlds to be visited and manipulated by many. People and real spaces are brought together convincingly over great distances.

Visual capabilities are enhanced beyond innate ability: users see over great distances, over multiple spectra, and through inclement weather. They see through obstructions and unpleasant or distracting objects are erased from view. Visual problems are corrected and compensated for.

In summary, augmented reality promises to make *computer graphics and user interfaces more integrated with human vision and our perception of the world*, and less an external entity arranged in an artificial way.

1.2 Augmented Reality and Eyeglasses-Like Displays

The above vision of augmented reality relies on a set of technical refinements, among them: improved head and eye tracking, better environment mapping and localization, improved computation performance and power efficiency, high throughput wireless networking, and reduced system latency. However, among the most crucial aspects of such an augmented reality system is obtaining a suitable display. A see-through near-eye display seems a natural choice for an AR display that is used for *casual, general-purpose, and extended use* as described in the vision; it can be made mobile and remains in the user's field of view (FOV) at all times. Among near-eye displays, a display with an *eyeglasses-like* (or perhaps even a *contact lens-like*) form factor appears the most appealing: it covers much of the viewer's field of view, is non-encumbering and already used by much of the population, is socially acceptable, and can easily be taken on and off. Ideally, the eyeglasses-like display would also support the resolution and dynamic range of the human visual system, have the ability to selectively occlude real objects in the environment, and would support all of the major depth cues, including the focal depth cues.

Since Sutherland introduced the first see-through near-eye display system with computer generated imagery in the 1960s (Sutherland, 1968), such displays have progressed to form factors that approach ordinary eyeglasses (e.g. Lumus DK-32). However, state-of-the-art examples in this form factor are limited to the same $40°$ field of view used in the Sutherland system – much narrower than the field of view through ordinary eyeglasses, generally $\geq 100°$. It is believed that these $40°$ fields of view are much too narrow to be of much utility for general spatially registered content; objects will be cropped to a small viewing window, reducing realism and requiring active effort by the user to keep augmented objects within view. *The limited field of view available to displays with practically compact form factors is considered one of the major obstacles to obtaining a display suitable for general-purpose augmented reality*. Displays in eyeglasses-like form factors also do not have a selective occlusion capability or support the focal depth cues, which limit the realism of the displays. See Section 2.1 for more discussion of these factors.

In this dissertation, a new design strategy is proposed for see-through near-eye displays that relies on the principles of *computational displays*, optical devices in which a fundamental part of the image formation process takes place in software. Computational designs have a different set of tradeoffs than conventional optical approaches; notably a compact form factor and wide field of view are more readily obtainable. More elusive features, such as a selective occlusion capability and focal depth cues, are also

within reach. In this dissertation, two computational display designs are described that support these features, and performance is analyzed. Although the proposed displays have various limitations that prevent their immediate commercialization, they show promise in that they *offer a vastly different set of trade-offs* than conventional optical designs. It is hoped that the proposed designs will stimulate new ideas in the space of eyeglasses-like near-eye displays, and advance the field one step closer to the goal of ubiquitous augmented reality.

1.3 Thesis Statement

The use of a joint computational and optical approach to see-through near eye displays will allow displays to achieve higher performance, a new set of capabilities, and a different set of trade-offs as compared to purely optical approaches.

1.4 Contributions

The main contribution of this dissertation is to introduce computational design principles to the field of near-eye see-through displays. These designs offer some characteristics that have not been previously achieved in compact see-through near-eye designs: wide fields of view, occlusion support, and support for the focal depth cues. Specifically, two complementary display designs are proposed: a *subtractive* approach that filters an omnidirectional light source through a display stack, and an *additive* approach that uses a display panel with a highly directional backlight. Contributions specific to each of these approaches are listed in Sections 3.2 and 4.2.

1.5 Structure

Chapter 2 provides an introduction of pertinent factors in near eye displays and an overview of existing approaches. Chapter 3 describes a *subtractive, multilayer* approach to near-eye display design, and Chapter 4 describes a complementary *additive, point light source* approach. These two chapters generally follow two previous publications (Maimone and Fuchs, 2013; Maimone et al., 2014b). Chapter 5 provides a summary and conclusion.

CHAPTER 2

Background

2.1 Factors for See-Through Near-Eye Displays

This section lists an overview of factors to consider when designing a see-through near-eye display. The list is limited to factors directly related to the display hardware, the focus of this dissertation; the design of a complete mobile augmented reality system would include a broader range of topics including head tracking and latency, networking, power consumption, the integration of electronics, and other factors.

2.1.1 Spatial Resolution

Humans are very sensitive to spatial resolution and thus this factor is very important to preserve in a display that will provide high information density or realism. The acuity of human vision is typically measured by the highest spatial frequency that the visual system can resolve in terms of cycles per visual angle. The standard measure of "normal" or 20/20 vision is support for a spatial frequency of 30 cpd (cycles per degree), i.e. the ability to resolve a contrasting intensity cycle spanning two arc minutes (2/60 degree), or a feature spanning one arc minute.

This standard measure is conservative, however; higher resolution limits have been reported. Campbell and Green (1965) measured human visual acuity limits of about 60 cpd using a laser to form interference fringes directly on the retina, bypassing some of the optics of the eye. Curcio et al. (1990) arrived at a similar average figure of 66 cpd by measuring the peak density of cone cells on the retina (i.e in the foveal region). The human visual system can also determine the *relative* position between visual features with much higher accuracy than it can resolve two adjacent features, a phenomenon known as *hyperacuity*. Wang et al. (2009) measured a mean localization accuracy of 8.3 arc *seconds* in adults, greatly exceeding the visual system's standard 1 arc minute resolving power.

With the assumption that a contrasting cycle can be represented by two adjacent pixels, a near-eye display supporting a field of view of $100° \times 100°$ would need to provide a resolution of approximately 6000×6000 to 12000×12000 pixels to match the limits of the visual system described above (30-60 cpd). These high resolutions place significant rendering and bandwidth requirements in the display system, but these requirements can be significantly reduced if resolution is adjusted according to the varying resolution sensitivity across the retina. The human visual system perceives fine detail only in the central $5°$ of the retina (known as the *fovea*), with sensitivity falling steeply outside this region. Guenter et al. (2012) estimate that a $70°$ wide display would benefit from a $100\times$ speedup in rendering performance if the display were segmented into four zones of decreasing detail toward the periphery matched to human resolution sensitivity; greater speedup factors are predicted with increasing fields of views and numbers of detail zones.

2.1.2   Field of View

The human eye has a static field of view of approximately $151°$ wide, and the combined field of view between both eyes is about $190°$. With eye motion, this field of view expands to approximately $290°$ (Howard and Rogers, 1996). While not restricting the FOV in the far periphery is important to maintain balance and locomotion (Assaiante and Amblard, 1992), the narrower field of view supported by corrective eyeglasses ($\geq 100°$ horizontally) suggests a pragmatic target for imagery requiring the viewer's active attention. Thus, an ideal near-eye see-through display should support augmented imagery over a $\geq 100°$ wide FOV while not occluding the viewer's peripheral sight. Wider fields of view are especially important for displaying imagery *spatially registered* to the world, which may appear anywhere in the viewer's field of view according to the viewing position; restricting this field of view may result in imagery disappearing or being cropped to the viewing window. Such cropping limits the region of the world that can be instantaneously augmented, reduces the realism of the augmented display, and requires active user effort to keep spatially registered content within the supported field of view. An example measure of a field of view is shown in Figure 2.1.

2.1.3   Eyebox

Another important consideration for any near-eye display is the *eyebox*, the viewing volume in which the eye can move relative to the display while the user still sees the complete, intended image. The eyebox is measured as the size of the cross section of this viewing volume at the intended viewing distance as shown in Figure 2.1. Outside of the eyebox, the user may see vignetting (gradual darking of the image toward the
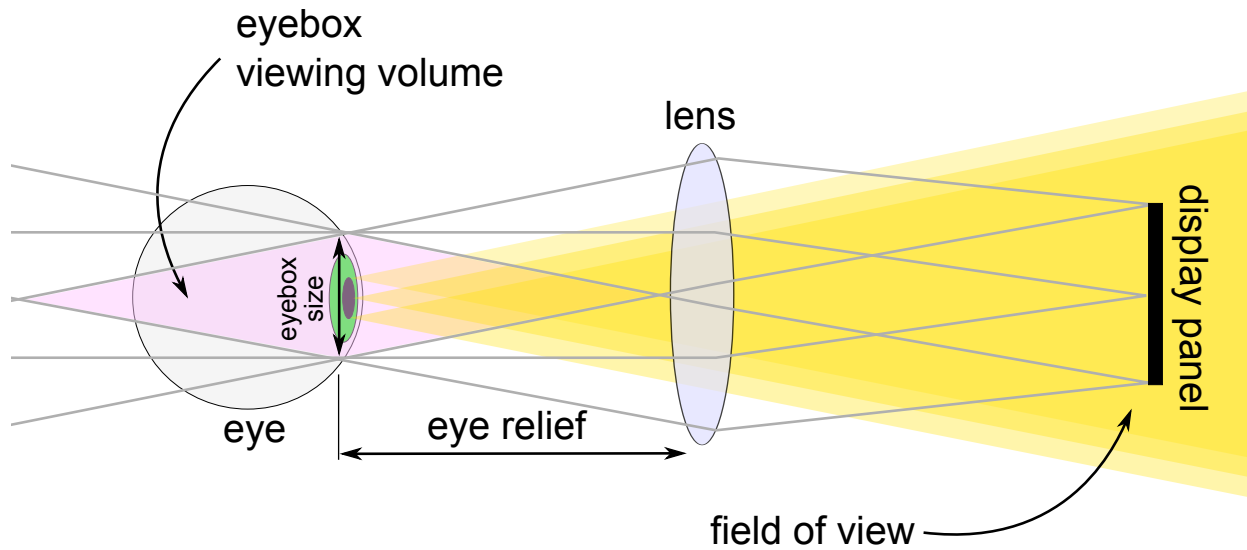
Figure 2.1: Near-eye display parameters. The eyebox, eye relief, and field of view are identified for a near-eye display consisting of a simple magnifying lens in front of a display panel that produces an image at optical infinity.

edges), a corrupt image, a partial image, or no image at all. The eyebox should be appropriately sized to allow for eye rotation, shifting of the near-eye display on the head, and variation in human anatomy. An eyebox of 10 mm or larger is generally recommended for head worn displays (Cakmakci and Rolland, 2006). However, note that the effective range of pupil motion within the eyebox depends on the pupil size; for a given eyebox size, a larger pupil has less freedom of movement. The eyebox position for a given display may be fixed, may be translated mechanically (e.g. by centering optical components over the eyes), or may be adjusted in software (Lanman and Luebke, 2013). In particular, it is useful to adjust the eyebox position to account for varying interpupillary distances, which vary between $45 - 80$ mm for virtually all adults (Dodgson, 2004). Larger eyeboxes are preferred, but eyebox sizes are usually balanced with other factors such as size (e.g. a larger imaging lens will generally produce a larger eyebox) and image quality (e.g. increasing eyebox size may increase the number of constraints on an optimized optical surface). In the computational designs proposed in this dissertation, eyebox size may be traded for non-traditional factors, such as resolution.

2.1.4 Eye Relief

Eye relief is the intended viewing distance of the display, measured from viewing center to the nearest display component. A suitably long eye relief is needed to allow space for viewer eyeglasses behind the display

and to improve viewer comfort; an eye relief of 20 mm or more is recommended (Cakmakci and Rolland, 2006). Eye relief is generally traded off among other factors in a display; Figure 2.1 shows how eyebox size varies directly with viewer eye relief in a simple near-eye display design. In optical configurations where the field of view is limited by the size of the optical element nearest the eye, the field of view of the display also decreases with increasing eye relief. Incorporating corrective eyewear into the display itself (e.g. as with Google Glass Frames[1]) offers an opportunity to reduce the necessary eye relief.

2.1.5 Depth Cues

The human visual system perceives depth through a variety of monocular and binocular cues which can be reproduced by near-eye display hardware. The binocular cues consist of *retinal disparity* (the eyes seeing the scene from different perspectives) and *convergence* (the eyes rotating inward when fixating on an object). These are readily handled by most near-eye displays by placing a display element in front of each eye and using appropriate left and right eye view rendering. Among the monocular cues, *motion parallax* – the apparent relative motion between objects based on their depth – is readily incorporated into a near-eye display by changing the rendered viewpoint according to head tracking data. Other monocular cues, such as perspective (perceived distance from apparent size), can be reproduced by sending an appropriately rendered image to the display. These cues have been supported since the early head mounted display systems (Sutherland, 1968).

Additional monocular cues are related to depth of focus. *Accommodation* is the cue stimulated when the ciliary muscles of the eyes expand or contract to change the shape of the lens and thus its focal length. *Retinal blur* is the sensed magnitude of blur for out of focus objects. Most near-eye displays support only a fixed focal position, often infinity. However, inclusion of focal cues in display systems improves consistency between augmented object depth cues (e.g. accommodation and convergence) and improves consistency between augmented object and real world object cues. Retinal blur can also provide a more precise depth cue than binocular vergence for objects away from the point of visual fixation (Held et al., 2012). Proposed approaches for including these cues in see-through near-eye displays include modulating focus through deformable mirrors (Hu and Hua, 2014) or liquid lenses (Liu et al., 2010). Focus control places heavy constraints on the hardware design; these current approaches are bulky and have limited fields of view. In this dissertation,

---

[1]https://www.google.com/glass/help/frames/

7

some progress is made in this area by demonstrating adjustable focus cues in a compact see-through display by generating a high angular resolution near-eye light field (see Chapter 3). Similar progress has also been recently demonstrated in an opaque near-eye display design (Lanman and Luebke, 2013).

2.1.6 Form Factor and Mobility

An attractive package for a general-purpose augmented reality display would resemble ordinary eyeglasses, a form factor that has been widely accepted for extended and everyday use for centuries. It is unlikely that a device significantly larger would be accepted for all-day use, but may find acceptance for specialized tasks within the home or workplace. Likewise, even smaller or more unusual form factors, such as the small offset display of Google Glass, have had difficulty gaining social acceptance (Gross, 2014).

See-through displays in form factors approaching eyeglasses have been demonstrated using freeform optics and waveguide designs by Epson, Lumus, Optinvent, Augmented Vision and others (see Section 2.2). Although these designs show impressive progress in form factor over those used in early see-through near-eye display systems (Sutherland, 1968), even the most advanced designs (e.g. Lumus DK-32) are limited to the same $40°$ field of view[2]. Section 2.2 describes how field of view and form factor are coupled in these designs, making it difficult to expand the field of view while retaining the same size. In this dissertation, progress is made in this area; two see-though display designs are proposed which have competitively slim form factors and fields of view of $100°$ diagonally or more.

2.1.7 Real-World Occlusion

A truly convincing augmented reality display should present augmented objects that appear opaque, and thus the display should be able to block light from the environment that coincides with light from augmented objects. Several optical see-through displays (Kiyokawa et al., 2003; Cakmakci et al., 2004; Santos et al., 2008) have been designed with such a selective occlusion capability by focusing the light from the world to a plane, attenuating it with a spatial light modulator, and combining it with the augmented image optical path. The ability of the display to provide occlusion is the primary design goal of such displays and places heavy constraints on the hardware design, making the devices quite bulky; the only known commercial

---

[2]Sutherland (1968) does not identify if the field of view is measured horizontally or diagonally; in either case, it is as wide or wider than the $40°$ diagonal Lumus DK-32 design.

device (Santos et al., 2008) has a form factor that resembles binoculars. In this dissertation, some progress has been made by proposing a thin light field occlusion mask (see Chapter 3).

2.1.8   See-Through Capability

The defining aspect of an augmented reality display is a "see-through" capability: the ability to see synthetic objects overlaid upon one's natural environment. Such a capability can be provided in two forms: a *video see-through* capability, in which the user views a camera feed of the real environment on an opaque display, and a *optical see-through* capability, in which the user views the environment directly through a transparent or semi-transparent display. As outlined by Rolland and Fuchs (2000), there are various tradeoffs between the designs that are summarized here. A video see-through display is better able to handle tracking latency because the video signal can be delayed and synthetic content can be registered to the world view in software. It is also trivial to provide an occlusion capability in a video see-through display and less difficult to support a wide field of view for synthetic objects. However, video see-through displays currently provide a significantly degraded view of the real environment: there is lag between viewer movement and video update, spatial resolution and field of view are reduced, and depth of field is limited. Although video see-through devices may prove useful for specific applications (such as the medical applications considered by Rolland and Fuchs), the degradation of their see-through view is currently too severe to consider for displays targeted for general-purpose and extended everyday use, as is the focus of this dissertation. Therefore, only optical see-through designs are considered in this dissertation, and in this document "see-through" will be treated to be synonymous for "optical see-through".

As described later in Section 2.2 and illustrated in Figure 2.2, the general strategy for *optical* see-through designs is to place opaque optical components (e.g. display panels) away from the eye and to relay light into the eye using reflective optical components. The disadvantage of this approach is that the need to relay light adds extra complexity and bulk. In this dissertation, a new approach is proposed in which relay components are eliminated by designing the display so that all components are transparent and can be placed directly in front of the eye.

9
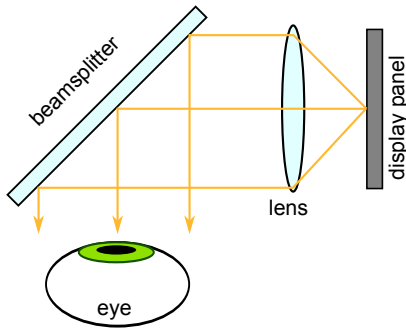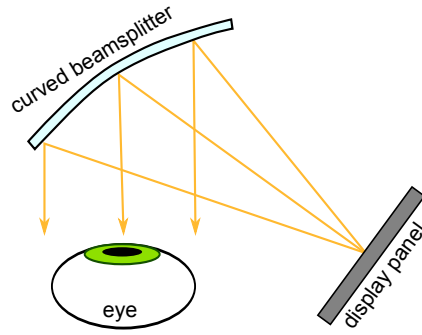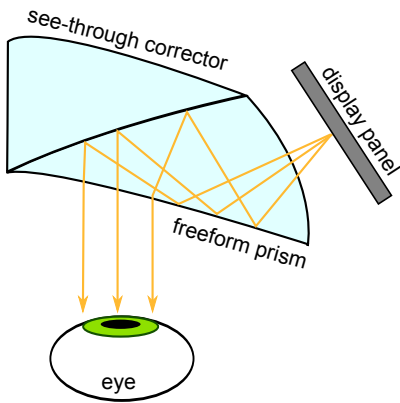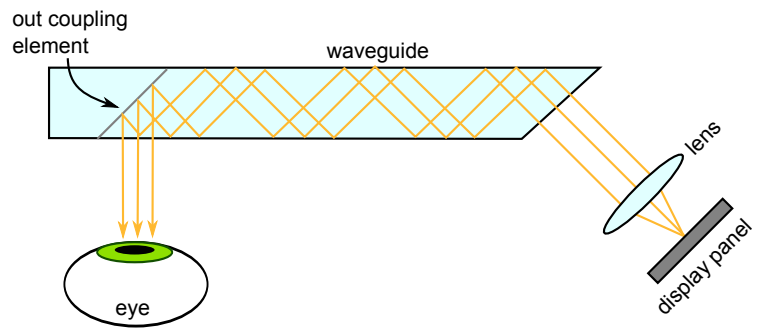
Figure 2.2: *A-D:* Common optical designs for see-through near-eye displays. Each diagram shows how the light from one pixel in a display panel propagates through the display optics to create an apparently distant point (here at optical infinity).

Figure 2.3: Example of simple combiner head-mounted display used in a system by Sutherland (1968).

2.2 Related Work

A near-eye see-through display must perform three primary functions: to bring into *focus* (and usually *magnify*) a synthetic image source placed near the eye, to *relay* the image into the eye, and to *combine* the optical paths of the synthetic image source and the see-through view. In this section, various existing approaches to achieve this functionality are discussed, with emphasis on common, general-purpose designs. Work in specialty designs focused on the focal depth cues and real-world occlusion support were listed along with the introduction of these factors in Sections 2.1.5 and 2.1.7, and more design variations are covered in a recent survey by Kress and Starner (2013). Related works that are relevant to the specific display designs proposed in this dissertation but are not directly related to near-eye displays are introduced along with each of the designs in Sections 3.4 and 4.4.

2.2.1 Simple Combiners

Perhaps the simplest way to achieve the three primary objectives of a see-through near eye display is to start with a simple opaque near eye display (e.g. a simple magnifier placed over a display panel) and reflect the display into the eye with an angled partially reflective mirror. This design is illustrated in Figure 2.2A and is similar to one used by Sutherland (1968) in one of the earliest near-eye display systems (see Figure 2.3). The disadvantage of this approach is that the display becomes very bulky as the field of view increases; the display

Figure 2.4: Example of combined reflective design head-mounted display, the Laster SeeThru. Photo by Laster.

used by Sutherland (1968) supported a moderate $40°$ field of view while retaining an acceptable form factor. However, variations of this simple design may prove successful for small, unobtrusive information displays.

2.2.2 Combined Reflective Designs

An improvement on the simple combiner design is to combine the functions of the lens and beamsplitter into a single *curved* reflective combiner as shown in Figure 2.2B. The partial reflectivity of the combiner allows the light from an opaque display panel to be reflected into the eye, while the combiner's curvature focuses the light rays so that the image appears within the viewer's accommodation range. An image source placed off-axis (e.g. near the side of the head) allows the display to be miniatured. An optimized aspheric (rather than spherical) surface may also be used to reduce spherical and comatic aberrations. However, as a single optical surface has limited degrees of freedom (i.e. each point on the surface can only reflect light in one direction), such designs are limited in field of view and eyebox size. A state of the art example, the Laster SeeThru[3] (see Figure 2.4), provides a compact form factor and a field of view of $25°$ diagonally. Other announced products of Laster provide fields of view of up to $40 \times 30°$.

2.2.3 Combined Refractive Designs

Optical functions can also be combined in a single *refractive* design; *freeform prisms* combine all optical functions of a see-through near-eye display into a single optic with multiple optimized aspheric surfaces, allowing more degrees of freedom. In such designs, illustrated in Figure 2.2C, the image source can be placed

---

[3] http://laster.fr/products/seethru/

Figure 2.5: Example of combined refractive design head-mounted display by Cakmakci et al. (2010)
.

closer to the optics, resulting in a more compact design. Form factors approaching eyeglasses or sunglasses have been reported (Cakmakci et al., 2010; Gao et al., 2012), albeit with narrower fields of view ($20 - 30°$ diagonally); see Figure 2.5. Wide field of view freeform designs (approaching $120°$ horizontally) have also been proposed by tiling multiple larger prisms (Cheng et al., 2011), but the optics become proportionately heavy and thick (1.7 cm thick in the cited design).

2.2.4 Waveguides

A further evolution is the use of waveguides, thin optical slabs that relay light through total internal reflection (TIR). A waveguide can be used to relay a remotely placed image source (e.g. at the temples or ears) so that it appears in front of the eyes when the viewer is in the eyebox as shown in Figure 2.2D. Waveguide approaches are distinguished by the composition of their out-coupling (and associated in-coupling) elements, which cause light to stop propagation by TIR for redirection to the eye. The simplest out-coupling design is a single angled reflective element embedded in the waveguide, as shown in Figure 2.2D. However, this simple design suffers from a narrow field of view that is directly related to the thickness of the waveguide. One such commercial design, the Epson Moverio BT-100/BT-200[4], retains a moderately compact form factor, but the field of view is limited to $23°$ diagonally. More complex designs such as the Lumus DK-32[5] and

---

[4]The Epson BT-200 uses a curved reflector, rather than the flat angled reflector in the BT-100, but supports the same field of view.

[5]http://www.lumus-optical.com/index.php?option=com_content&task=view&id=9&Itemid=15

Figure 2.6: Left: Example of waveguide design head-mounted display, the Lumus DK-32. Right: Diagram showing Lumus optical components including the LOE (Light-guide Optical Element), i.e. a waveguide, and the POD, which projects the image source into the waveguide. Photo and diagram by Lumus.

.

the Optinvent ORA-S[6] use *cascaded arrays* of such angled reflectors to reduce size and expand the field of view. Holographic elements can also be used as out-coupling elements (Levola, 2006) to achieve similar size reductions and have been used in designs from Vuzix and Sony. Fields of view of up to $40°$ diagonally have been demonstrated commercially (Lumus DK-32, see Figure 2.6). Wider fields of view are possible by increasing the TIR angle supported by the waveguide and the supported angles of the in- and out-coupling elements, but increasingly difficult due to the need for more exotic components such as high-index glass and gratings with high efficiencies at wider angles.

2.2.5 Contact Lenses

In this dissertation, it is generally assumed that the most desirable augmented reality display matches the form factor and field of view of a pair of ordinary eyeglasses. However, an alternate goal is to incorporate a similar display into contact lenses, a form factor preferred by many requiring corrective lenses. Some early attempts have been reported (Lingley et al., 2011; De Smet et al., 2012) (see Figure 2.7), but are in the very early stages of development: current displays consist of only a single or small number of pixels and do not address the issue of the eye focusing on the display.

An alternative approach combines both contact lenses and eyeglasses in simple forms. The recently-announced Innovega iOptik display system[7] consists of a contact lens containing a small inset lens that

---

[6]http://optinvent.com/see-through-glasses-ORA

[7]http://www.innovega-inc.com/new-architecture.php
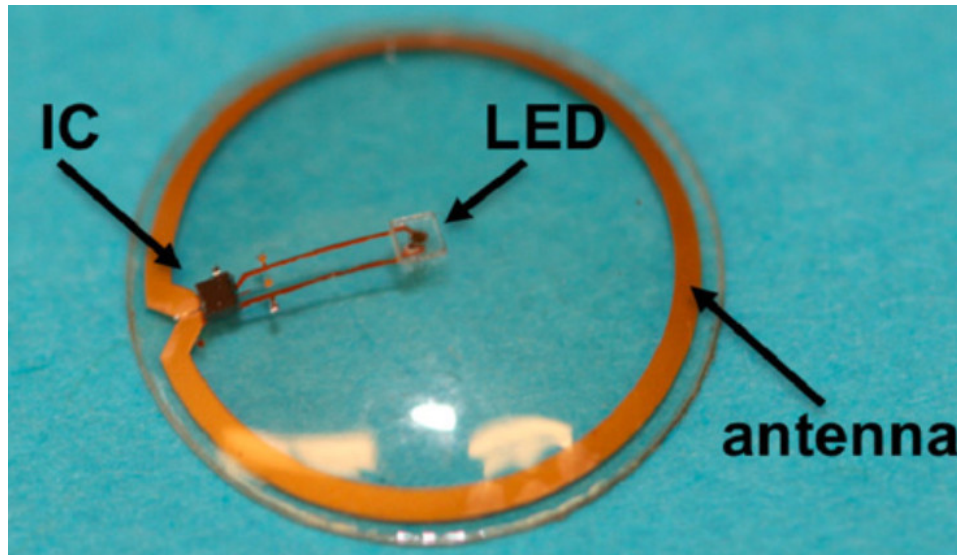
Figure 2.7: Example of contact lens display by Lingley et al. (2011).

allows the user to focus on a see-through near-eye display (e.g. a semi-transparent display panel or projection screen) while allowing light from the environment to enter the remainder of the pupil. Filters are used to separate light between the two regions. The manufacturer claims the design can ultimately achieve a $100°$ diagonal field of view (Innovega, 2012). This interesting design may prove to be a viable competitor to eyeglasses-based designs, although it remains to be seen whether wearing both contact lenses and eyeglasses provides an improved and acceptable level of encumbrance.

2.2.6 Summary

There is currently a fundamental tension between form factor and field of view among all of the common see-through near eye display designs. In the following two chapters, two alternative designs are described which have a more favorable balance between form factor and field of view. This is achieved by eliminating most of the optical components present in these prior designs and instead relying on computation to achieve some of the image formation process.

CHAPTER 3

Subtractive Approach: Multilayer Display

3.1 Introduction

In conventional near-eye see-through displays, the displayed image begins as a set of point sources, each representing a pixel in the image. The point sources are typically emitted over a wide angle from a small display panel (e.g. an LCD microdisplay) to be collected by the eye – the wide emission angle allows the eye some freedom of movement while still seeing the image. To appear in focus, the rays emitted from each of these point sources must appear to intersect at the focal plane of the eye so that the lens of the eye brings them into a focused spot on the retina. The rays can be adjusted to appear to intersect at a more distant location than the panel using refractive, diffractive, or reflective optical elements. Any opaque or distorting optical components must be placed outside of the field of view and reflected back into the eye or must be corrected with additional components, creating a see-through display. In summary, the job of the optics of a conventional see-through near-eye display is to *redirect* the light rays from a small image source into those that appear to form a distant transparent image in front of the viewer. The primary disadvantage of such optics is that they overly restrict the field of view or add unacceptable bulk, as described in Section 2.2.

In this chapter, a new see-through multilayer display design is described that takes a fundamentally different approach. Rather than start with the desired set of rays and use optics to redirect them into the eye, the multilayer design starts with the set of all possible light rays that could be emitted by the display (i.e. a white omni-directional area light source) and *filters* them through a set of optimized patterns displayed on a stack spatial light modulators (SLM) so that only the desired rays remain, those that represent a distant, focused image. In this sense, it is a *subtractive* approach. As described in this chapter, the optics for such a display can be made very compact while still supporting partial transparency and a wide field of view.
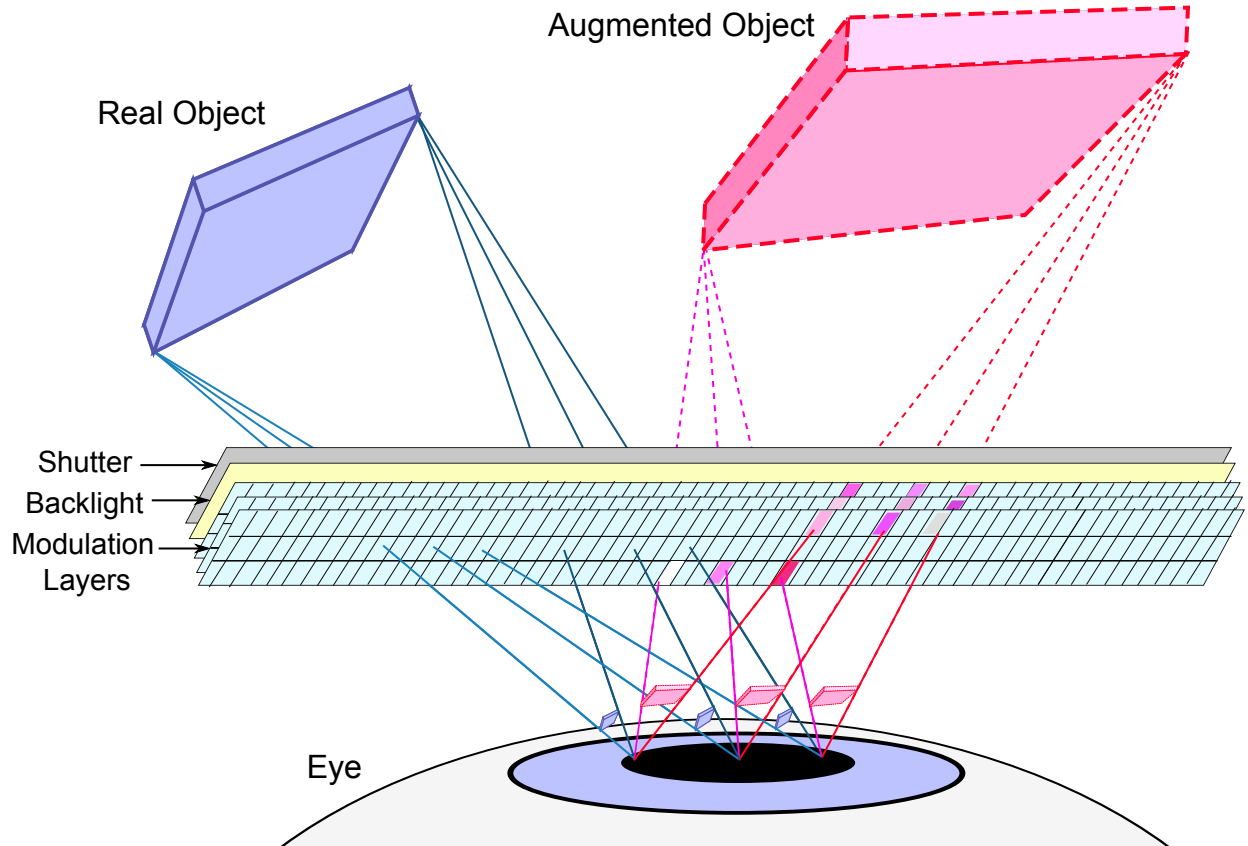
Figure 3.1: Components and ray formation. The display consists of a shutter and a backlit set of modulation layers; all components are transparent. When the shutter is closed, each augmented image ray is reproduced as a white light ray from the backlight that is attenuated across the layers to produce the desired color. The final ray color is the product of the intensities of the intersected pixels across all layers. Multiple regions over the pupil are considered when setting the pixel intensities so that a distant focused image can be formed using layers placed very near the eye, out of focus. When the shutter is open, light rays from real objects may reach the eye or may be selectively occluded by the layers.

3.2   Contributions

The multilayer design is a novel approach to see-through near-eye displays. Specific contributions include:

1. The analysis of multilayer display architectures for a new use case: a *partially transparent, near-eye* display that creates images that appear *far from the physical display layers*

2. The proposed use of the display layers both to form images and to provide *per-ray, variable transparency occlusion* of the real scene, in a time-multiplexed fashion

3. Revised multilayer light field constraint and optimization criteria critical to achieve improved performance in near-eye display configurations

3.3   Benefits and Limitations

The multilayer design offers the following benefits:

- Simple hardware architecture

- Support for wide field of view in a compact form factor

- Support for multiple simultaneous focal depths

- Support for selective occlusion of the real environment

The multilayer design suffers from the following limitations:

- Compressive approach results in image degradation

- Layer optimization process is computationally expensive

- Practical use requires spatial light modulations that exceed current levels of speed and transparency

- Practical use requires mitigation of diffraction effects through display layers

3.4 Related Techniques

An emerging class of multilayer computational displays produce multi-view imagery by displaying image patterns on a stack of display layers (Gotoda, 2010; Lanman et al., 2010; Wetzstein et al., 2011; Lanman et al., 2011; Wetzstein et al., 2012). When the display stack is viewed from different positions, parallax causes different pixels on each layer to align and form the perceived image, providing motion parallax and stereo depth cues. The modulation state of each layer (e.g. attenuation, polarization, etc.) is optimized to produce the most numerically accurate images for the desired set of viewpoints. Pattern time-multiplexing over the flicker fusion threshold can be used to obtain additional degrees of freedom for layer optimization, improving image quality. Directional backlights may be used with the display layers to increase angular resolution (Wetzstein et al., 2012).

To date, multilayer designs have been limited to desktop 3D displays designed to be viewed at a distance and within the accomodation range of the viewer. The present work builds on these existing displays, while exploring several new design aspects: near eye displays with layers placed closer than the typical eye accommodation distance, sight through the display with selective occlusion, and revised optimization criteria for near-eye viewing.

Related work by the author (Maimone et al., 2013) in multilayer displays has also shown the possibility of stimulating nearly correct eye accommodation by synthesizing narrow, high density light fields over the areas of the viewer's pupils. In the multilayer design proposed here, such light fields are also synthesized, but with the primary goal of creating a focused image nearer than the eyes can accommodate, rather than presenting a scene with multiple focal depths on a distant display. A different hardware approach is used to gain a thin form factor and see-through ability: light is filtered through a set of transparent spatial light modulators rather than focused by a lens. The proposed multilayer approach also uses additional constraints on the input light field and considers the image formed on the retina during the optimization process to improve light field compression performance.

3.5 Display Design and Analysis

The proposed physical design in essentially a set of stacked transparent displays placed directly in front of the eyes, closer than the typical eye accommodation distance, and worn as eyeglasses. To create a

focused augmented image, multilayer optimization techniques are used (see Sections 3.5.4, 3.5.5) rather than conventional optical components such as lenses and beamsplitters. To provide a see-through capability, the displays are simply set to a transparent state, or are programmed to selectively occlude parts of the environment.

3.5.1 Components

The proposed design relies on the following hardware components:

1. Two or more thin, high-speed *transmissive* spatial light modulators that control the intensity of passing light through attenuation, e.g. transparent LCDs.

2. A thin and transparent backlight that uniformly distributes light over its face and can be rapidly modulated, e.g. a transparent OLED lighting panel.

3. Optionally, a thin high-speed shutter that can be switched from a globally transparent to a globally opaque state, e.g. a single cell liquid crystal shutter.

All components are sandwiched together with a small spacing between the spatial light modulator layers, as illustrated in Figure 3.1.

3.5.2 Physical Attributes

The proposed design supports a compact form factor and wide field of view that approaches ordinary eyeglasses. The thickness of the device is determined by the thickness of the modulation layers and spacing (typically 3 mm to 10 mm) and the thickness of the integrated backlight and shutter (typically ≤1 mm each). Performance generally improves as the layer separation is increased, as described in Section 3.5.5.1. The field of view of the device is limited only by the size of the spatial light modulator layers, the distance between the eye and layers, and the supported viewing angle of the modulation layers[1]. The specific configuration used for a prototype display is discussed in Section 3.6.
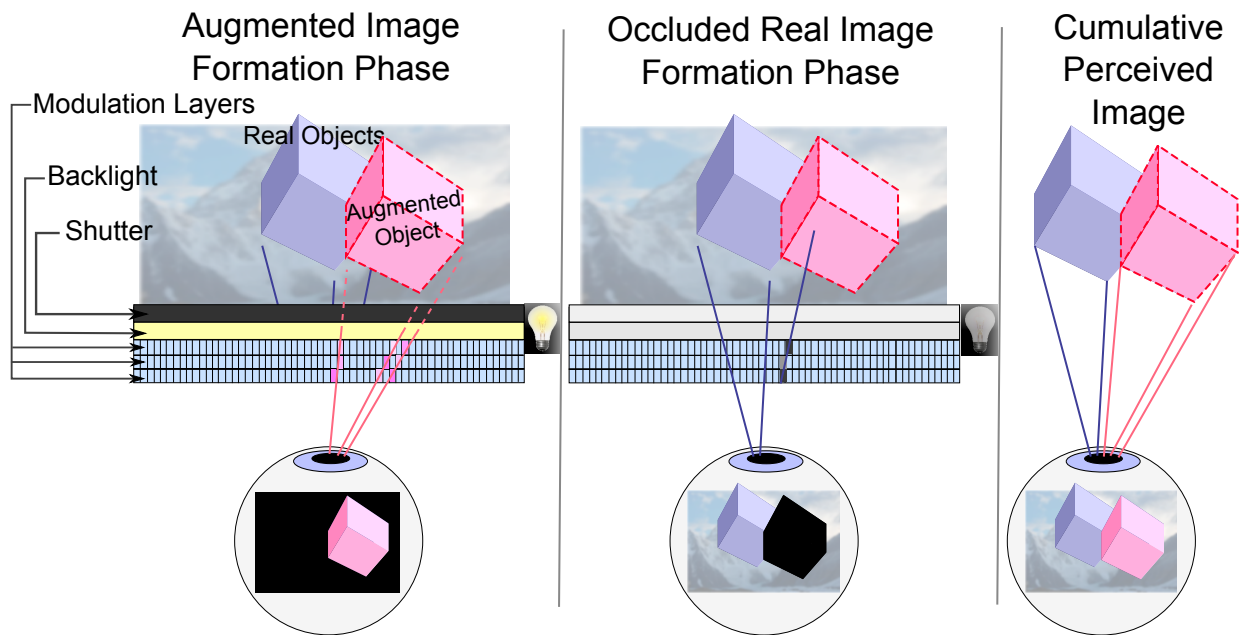
Figure 3.2: Time-sequential operation. *Left:* In the first phase, an augmented image is formed through the backlight and spatial light modulation layers while the real scene is blocked by the shutter. *Center:* In the second phase, an occluded real image is formed by opening the shutter and blocking light rays from real objects that coincide with augmented objects. *Right:* The two phases alternated in rapid sequence form a perceived AR scene with opaque augmented objects.

3.5.3 Time-sequential Operation

If the occlusion of real[2] objects is desired, the proposed device operates in two alternating phases, as illustrated in Figure 3.2. In the *augmented image formation* phase, light from the backlight passes through optimized patterns displayed on the spatial light modulation layers (see Section 3.5.4) to form an augmented image over the viewer's pupil while light from the real scene is blocked with the shutter. In the *occluded real image formation* phase, the shutter is opened and the backlight is disabled. Light from real objects enters the display and rays are selectively permitted to reach the eye or are blocked by the spatial light modulator layers. Real image light rays are blocked if they coincide with augmented image light rays for augmented objects that are intended to appear opaque. These two phases are rapidly alternated beyond the flicker fusion threshold so that the user perceives the superimposed result – a real scene with opaque augmented objects.

Alternate Configuration Alternatively, the augmented image formation and occluded real image phases can occur simultaneously if an additional set of modulation layers are installed behind the backlight (i.e. farthest from the eye) for dedicated occluded real image formation. Augmented images are formed on the existing set of modulation layers, while leaving display regions that do not correspond to augmented images transparent. In this case, the backlight should provide selective light control over the display area (e.g. as with a transparent OLED display) so that only display regions corresponding to augmented objects are illuminated. The alternative approach offers a higher frame rate at the expense of a thicker and likely dimmer display, due to multiplicative light loss through each of the additional layers.

3.5.4 General Multilayer Optimization for Light Field Displays

In this section, a general approach to multilayer optimization that has been applied to past displays (Lanman et al., 2010) is summarized. In Section 3.5.5, additional innovations are discussed which open the approach to near-eye displays.

Introduction Each ray emitted by a multilayer light field display begins as a white light ray in a backlight that is attenuated by a series spatial light modulator layers as it travels toward the eye (see Figure 3.1).

---

[1]Human anatomy and form factor constraints may limit the field of view in practice. For example, with an eye relief of 20 mm and interpupillary distance of 64 mm, the maximum field of view on the nasal side is $\approx 58°$ horizontally before the left and right eye display elements reach each other.

[2]In this chapter, the term "real" is refers to physicality of an object, image, or scene, not the optical quality in which light rays actually converge at their apparent position.

The final color of the ray entering the eye is the product of the attenuation values assigned to each of the intersected pixels across the layers. Thus, to generate a light ray, the corresponding pixel intensities on each layer should be set so that their product equals the desired ray color.

The multiplicative constraints placed on each layer pixel from all intersecting rays must be considered to determine the attenuation values for each layer. In general, there will be many more constraints than modulatable pixels, and thus all constraints cannot be satisfied exactly for a non-trivial light field. Additional degrees of freedom can be obtained by time-multiplexing a set of layer patterns such that the sum of their emitted light fields approximates the target light field. This set of patterns is displayed in rapid sequence beyond the flicker fusion threshold to give the impression of a persistent image. However, for a practical number of time-multiplexed patterns, it is likely that the system is still overdetermined and the layer patterns must be *optimized* so that they generate a light field that best approximates the target light field. In this case, the optimal layer patterns will take advantage of redundancies in the light field, essentially *compressing* it.

Optimization Process Content-Adaptive Parallax Barriers (Lanman et al., 2010) provides a method for generating an optimized set of time-multiplexed layer patterns for a two layer display which is briefly summarized here. Using a two plane parameterization, the target 4D light field $\mathbf{L}$ to emit from such a display can be expressed as the sum of the T time-multiplexed matrix products of the 2D patterns displayed on a backlit pair of attenuation layers $\mathbf{f}$ and $\mathbf{g}$:

$$\mathbf{L}[i,j,k,l] = \sum_{t=1}^{T} \mathbf{f}_t[i,j] \otimes \mathbf{g}_t[k,l] \tag{3.1}$$

The N pixels of the 2D pattern for each of the T time-multiplexed patterns can be reordered as a $N \times T$ and $T \times N$ matrix for $\mathbf{f}$ and $\mathbf{g}$ respectively. Similarly, the 4D light field $\mathbf{L}$ can be reordered as $N \times N$ matrix to obtain the equivalent matrix product:

$$\mathbf{L} = \mathbf{FG} \tag{3.2}$$

The optimal time-multiplexed layer patterns $\mathbf{F}$ and $\mathbf{G}$ can then be obtained by solving the following optimization problem:

$$\arg\min_{\mathbf{F},\mathbf{G}} \frac{1}{2} \|\beta\mathbf{L} - \mathbf{FG}\|_{\mathbf{W}}^2, \text{ for } 0 \leq \mathbf{F}, \mathbf{G} \leq 1,$$

$$\frac{1}{2} \|\beta\mathbf{L} - \mathbf{FG}\|_{\mathbf{W}}^2 = \sum_{i \in W,L,FG} [\mathbf{W} \circ (\mathbf{L} - \mathbf{FG}) \circ (\mathbf{L} - \mathbf{FG})]_i, \quad (3.3)$$

where $\mathbf{W}$ is a binary valued weight matrix that is used to select which of all possible emitted rays should be constrained (which determines the field of view of the display), $\circ$ is the Hadamard (element-wise) product, and β is a scaling factor used to trade brightness for image fidelity. For the best image quality, $\mathbf{W}$ should be set to select the fewest number of rays as possible to place the fewest constraints on the optimization. Ideally, only the set of rays that enter the viewers' eyes should be constrained; this set of rays can be determined with eye tracking (Maimone et al., 2013, 2014a). Note that $\mathbf{L}$ must be decomposed into $\mathbf{F}$ and $\mathbf{G}$ using non-negative values since elements represent light attenuation values. One such decomposition method is the weighted variant (Blondel et al., 2007) of the iterative update rules by Lee and Seung (1999):

$$\mathbf{F} \leftarrow \mathbf{F} \circ \frac{[(\mathbf{W} \circ \mathbf{L})\mathbf{G}^\mathsf{T}]}{[(\mathbf{W} \circ (\mathbf{FG}))\mathbf{G}^\mathsf{T}]}, \mathbf{G} \leftarrow \mathbf{G} \circ \frac{[\mathbf{F}^\mathsf{T}(\mathbf{W} \circ \mathbf{L})]}{[\mathbf{F}^\mathsf{T}(\mathbf{W} \circ (\mathbf{FG}))]} \quad (3.4)$$

$\mathbf{F}$ and $\mathbf{G}$ can be initialized with random noise and will converge to a local stationary point (not necessarily the globally optimum) (Lee and Seung, 1999). The result of this procedure is that an approximation of the target light field $\mathbf{L}$ is "compressed" into to a into a set of T time-multiplexed pairs of layer patterns $\mathbf{F}$ and $\mathbf{G}$. Current optimization performance is sub-realtime; optimization of a light field using the implementation of Tensor Displays (Wetzstein et al., 2012) takes a few minutes under the configurations described in this paper, but faster methods have been proposed (Heide et al., 2013). Figure 3.3 shows example layer patterns and a corresponding reconstructed view of the light field.

For additional information, see Tensor Displays (Wetzstein et al., 2012), which extends this formulation to three or more modulation layers and directional backlighting. These additional features were not explored in this work.

3.5.5 Multilayer Optimization for Near-Eye Displays

In this section enhanced multilayer optimization techniques are discussed that open the approach to near eye displays.
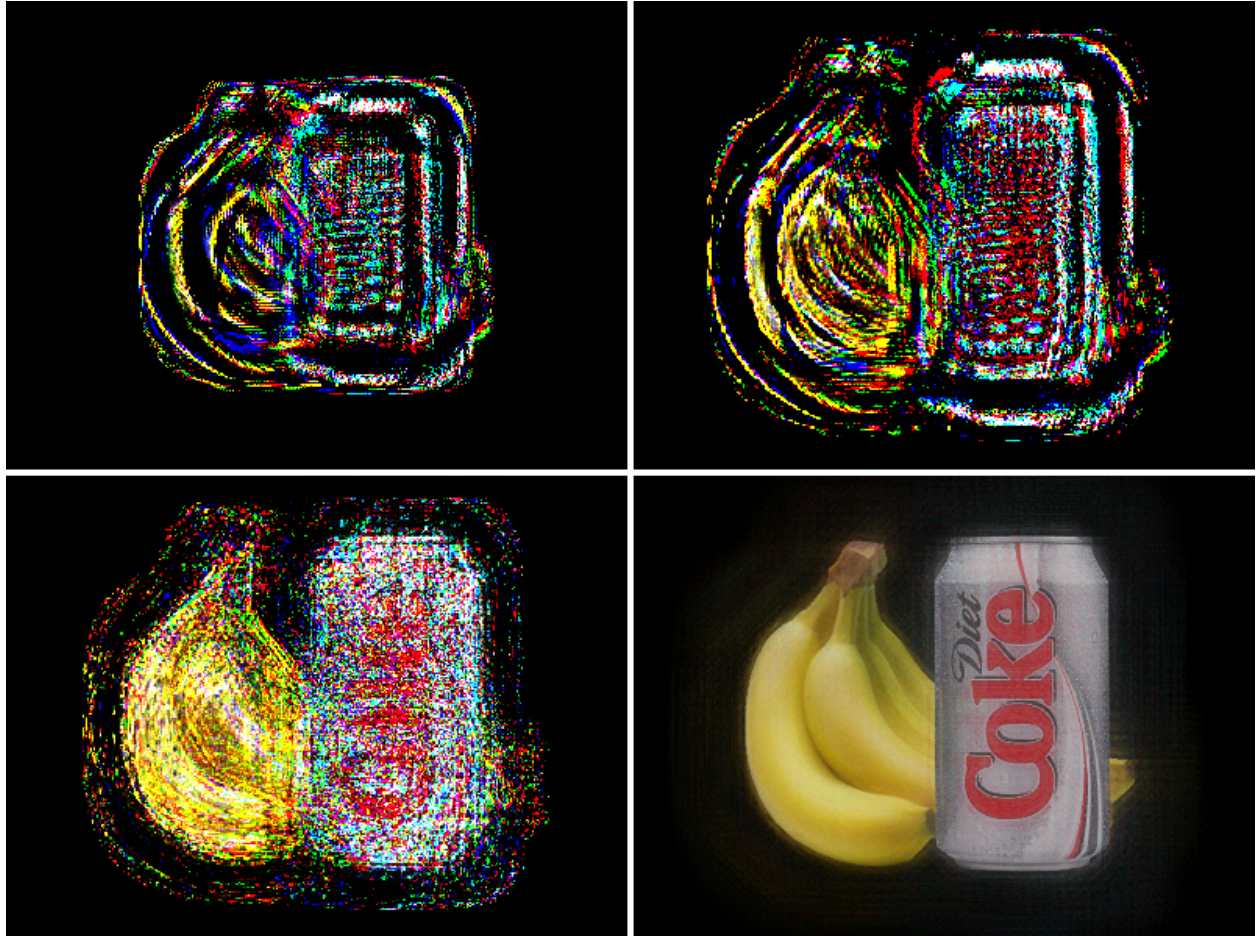
Figure 3.3: Example multilayer optimization. *Top row:* one of a set of eight time-multiplexed patterns to display on front and rear modulation layers. *Bottom left:* Simulated reconstructed view from one constrained point on the pupil, summed over the time-multiplexed patterns. *Bottom right:* Simulated accumulation of the 121 reconstructed view zones over the pupil area: the image perceived by the user if the eye is focused at the intended focal depth of the image. Section 3.5.5 describes how noise is canceled across views. Optimization was performed with a brightness scaling factor $\beta = 0.1$ (see Section 3.5.4), a focal distance of 15 cm, and the geometry of the prototype display (see Section 3.6).

3.5.5.1 Ray and Light Field Constraints

Unlike existing multilayer designs designed for desktop use, the primary objective of the proposed display is to produce a distant and focused augmented image using display layers placed closer than the typical eye accommodation distance. To meet this objective, the display should ideally reproduce the same set of rays that would be emitted from an augmented object as if it were physically present at its apparent location in the scene (see Figure 3.1). A benefit of creating such a display is support for imagery at multiple focal depths; it is believed that the ability to create multiple focal depths will reduce the eye fatigue caused by the accommodation-convergence conflict (Shibata et al., 2011), and will increase the sense of presence of augmented objects, as they will have consistency between depth cues that more closely resembles natural objects. To produce imagery that appears at focal depths other than the display layers, the eye cannot be treated as a single point; ray variation over the pupil must be considered.

Reproducing a light field with sufficient angular resolution to allow ray variation over the pupil is generally difficult for multilayer (and other) displays. Focus 3D (Maimone et al., 2013) demonstrates that a set of modulation layers alone provides insufficient angular resolution for ray variation over the pupil for a typical desktop display, requiring the addition of a high angular resolution backlight. However, achieving a high angular resolution in the present work is less challenging due to the short eye to display distance. In particular, the theoretical maximum angular resolution (in terms of the number of possible distinct rays per unit distance at the eye) for a two layer display is

$$r = \frac{d_s}{d_p d_e} \tag{3.5}$$

where d_p is the pixel pitch, d_e is the display to eye distance, and d_s is the layer separation distance, as shown in Figure 3.4. Since angular resolution is proportional to layer separation d_s, display performance will generally improve as the device thickness increases. For a near eye display, the minimum theoretical view spacing becomes a small multiple d_e/d_s of the pixel pitch, which is typically two orders of magnitude smaller than the size of a human pupil when a high density microdisplay is used. (Diffraction, however, will reduce the effective angular resolution as described in Section 3.5.6.3.) Due to this high angular resolution, optimization should be performed with a densely placed set of constraints over the pupil area, up to the maximum angular resolution. Since the regions in between constraints are undefined and may produce an unacceptable image, it is crucial that the constraints are sufficiently dense.

26
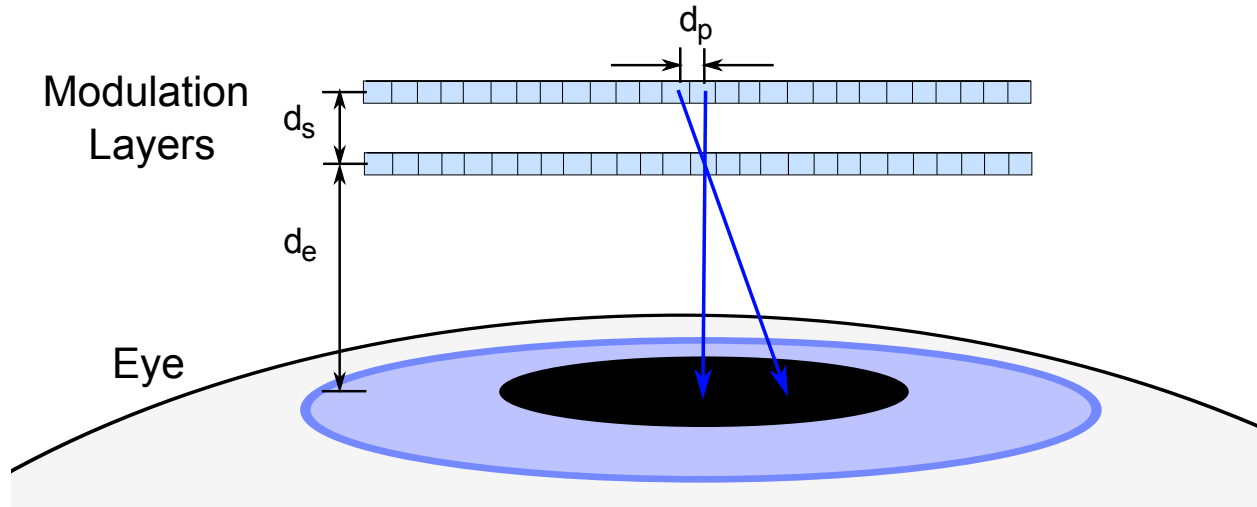
Figure 3.4: Display geometry to determine angular resolution for two layers. The theoretical maximum angular resolution r at the eye is computed as $\frac{d_s}{d_p d_e}$.

However, as Focus 3D (Maimone et al., 2013) also observes, the actual performance of a multilayer display using compressive optimization may be significantly less that the theoretical limits defined by the display's geometry, limiting sharpness and depth of field. Empirical testing confirmed this result; when compressing general light fields compression performance was found to be exceptionally poor; extreme noise in reconstructed views (see Figure 3.3, bottom left) overwhelmed angular variation between closely spaced views, causing images formed on the retina to be very blurry with no depth of field.

A first key insight to improve image quality for near-eye display configurations was to convert input light fields to those without local angular variation, i.e. to "planarize" the light field into a set of diffuse planes at a distinct set of focal depths (see Figure 3.5). With this restriction, the reconstructed views over different points on the pupil are very noisy as before; however, each represents a noisy version of the *same* image. Therefore, when all views over the pupil are summed (i.e. light is collected by the lens of the eye), much of the noise *cancels*, significantly improving image quality.

This process is demonstrated quantitatively in Figure 3.6. The plot shows that individually constrained viewpoints over the pupil (gray dots), are very noisy, typically showing a variation of approximately $\pm 10\%$ or more of the entire intensity range when compared to the intended signal (green dashed line). However, when the views are summed over the pupil area (solid red line), much of the noise cancels and the result is much closer to the intended signal.
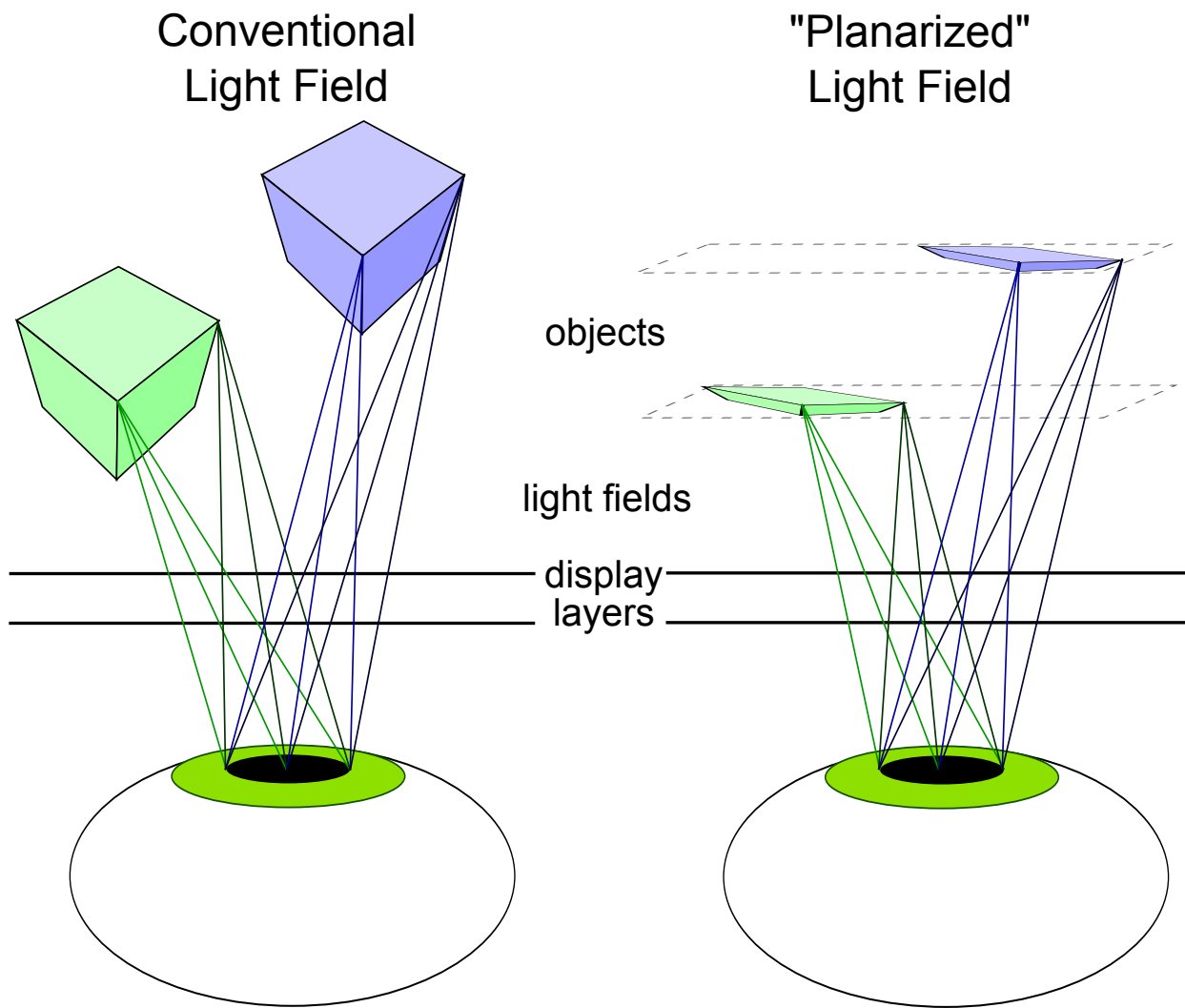
Figure 3.5: Multilayer planarized light field. *Left:* Conventional light field. Each object point emits rays from its true depth. *Right:* Planarized Light Field. Each object point emits rays from its projected position on one of a distinct set of planes.
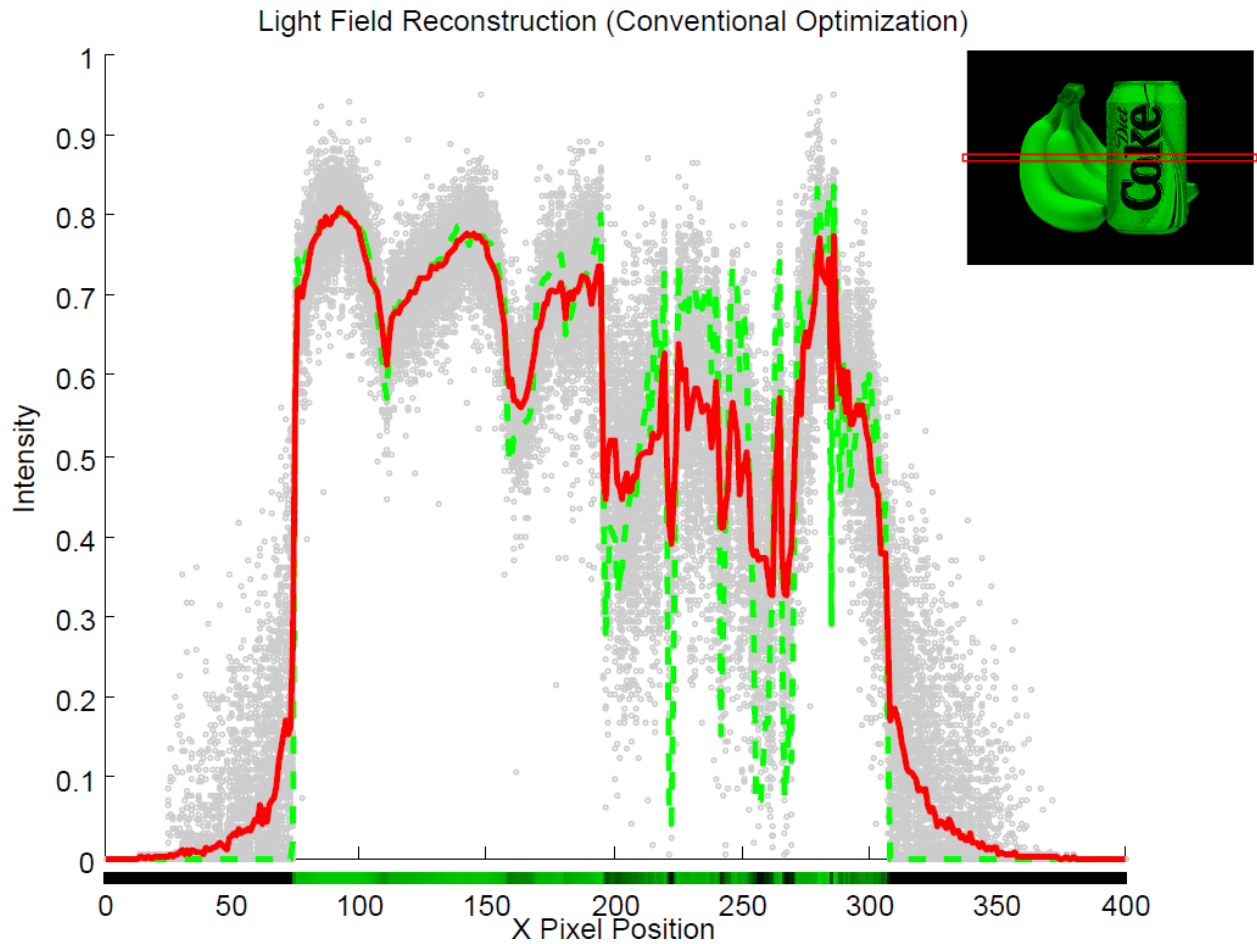
Figure 3.6: Noise cancellation of the conventional optimization of Section 3.5.4 using a "planarized" light field. The plot provides a 2D slice of a 4D reconstructed light field: the data from a single scanline (shown below x-axis and in top right inset image) over all of the constrained view zones on the pupil. The x-axis corresponds to the pixel position across the scanline and the y-axis represents the pixel intensity value. At each x-value, the set of gray dots provides the intensity values over the set of all views on the pupil. The data from all view zones should ideally match as the light field represents a diffuse plane. Note gray values are very noisy compared to the intended signal (green dashed line). However, when the views are collected and averaged over the pupil (solid red line), the reconstructed signal more closely matches the intended signal.

Of course, eliminating local angular variation in the input light field comes at a cost: objects grouped in the same plane will have the same focal depth, and thus the focal depth cues (accommodation and retinal blur) will be constant *within* each grouping. However, as shown in Figure 3.7, it is possible to preserve these focal cues *between* groups; that is, to create multiple planar regions over the display that each appear at their own focal depth.

3.5.5.2 Retinal Optimization

Although restricting light fields to diffuse planes for noise cancellation significantly improves image quality, the result is still substandard (see Figure 3.9, bottom left). The second key insight to improve image fidelity for near-eye display configurations was to optimize the *perceived* image rather than simply attempt to reconstruct the original *light field*. To form an image, rays enter the eye, are refracted by the lens, and strike the retina. For a given eye focal state, the individual intensities of the rays falling on a spot on the retina are not important, but rather their perceived sum; however, preserving individual ray intensities places additional constraints on the optimization. By constraining groups of rays only by their sum of intensities, it is expected that there will be more freedom to meet other constraints.

Performing this *retinal optimization* as described will, however, require knowledge of the focal state of the eye; it is necessary to know the state of the eye's lens to determine where rays will fall on the retina. Rather than attempt to measure this focal state, it is instead assumed that image quality is most important when an object is in focus. Therefore, optimization may be performed as if the eye is simultaneously focused on each object in the scene, improving in focus performance at the expense of out-of-focus blur quality.

To perform retinal optimization, the reconstructed light field matrix $\mathbf{FG}$, computed during each iteration of the rules specified in Equation 3.4, should be replaced with matrix $\mathbf{R}$, in which the reconstructed intensity of each ray is replaced with the average intensity of the set of rays falling on the same retinal spot. Matrix $\mathbf{R}$ is computed according to:

Algorithm 1 Compute Retinal Reconstruction $\mathbf{R}$

for each $(x, y) \in \boldsymbol{E}$ **do**
    $\boldsymbol{S} = RetinaRays(x, y)$
    $t = \frac{1}{|S|} \sum_{(i,j) \in \boldsymbol{S}} (\mathbf{FG})[i, j]$
    for each $(i, j) \in \boldsymbol{S}$ **do**
        $\mathbf{R}[i, j] = t$
    end for
end for

Figure 3.7: Multiple focal position simulation. Synthetic objects at different focal depths can be displayed simultaneously on the display. *Top row:* Light field refocused with synthetic camera at close (10 cm) focus. *Bottom row:* Light field refocused with synthetic camera at far (500 cm) focus. *Left column:* Original light field, provided as input to the optimization process. *Right column:* Simulated reconstructed views using multilayer optimization. Optimization was performed with two layers, eight time-multiplexed frames, brightness scaling factor $\beta = 0.1$ (see Section 3.5.4) and the geometry of the prototype display (see Section 3.6).

where $\boldsymbol{E}$ is the set of unique spots on the retina, and $RetinaRays(x, y)$ returns the set of rays that fall on retinal spot (x, y). Note that conventional optimization is performed before retinal optimization to provide a reasonable initial solution. For a human-viewable display, also note that retinal intensity calculations should also account for the Stiles-Crawford effect (Westheimer, 2008), the perceived dimming of light as it enters nearer the edges of the pupil.

Figure 3.8 provides a quantitative insight into retinal optimization. Note that the individual constrained zones over the retina (gray dots) are extremely noisy – often spanning the entire intensity range; however, when averaged over the retina (solid red line), they more closely match the target signal (green dashed line) then when conventional optimization is performed (see Figure 3.6). Figure 3.9 provides a qualitative comparison of conventional and retinal optimization. Note that when the eye is focused at the depth of the imagery, the retinal optimized image (bottom center) appears much sharper than the conventionally optimized image (bottom left). When the eye is focused elsewhere, the conventional optimization provides a smoother, more accurate retinal blur (top left), but the retinal optimization still provides a reasonable out-of-focus cue (top center). The author believes that a much sharper focused image at the expense of a moderate degradation of focal blur is a very acceptable trade-off. However, additional analysis is needed to determine how much the focal blur can be degraded while still stimulating an accommodative response.

3.5.5.3   Perceptual Optimization

Perceptual, rather that least squares, optimization for multilayer displays has been suggested (Lanman et al., 2010; Wetzstein et al., 2012) and implemented (Masia et al., 2012) through the use of complex non-linear objective functions. A simple alternative perceptual optimization scheme is proposed and used here that relies on the conventional optimization formulation (see Section 3.5.4) through the use of the existing ray weighting mechanism. In past work, weight matrix $\mathbf{W}$ (see Equation 3.4) was *binary*-valued and used to select which of the set of all possible rays emitted by the display should be constrained (i.e. those within the field of view of the display). A *real*-valued matrix $\mathbf{W}$ is instead proposed, in which a zero value indicates an unconstrained ray, and a non-zero value in the range $(0..1]$ indicates the perceptual importance of the ray.

The weight of each ray could be mapped to a variety of factors: a perceptual metric (e.g. contrast with adjacent rays), object importance, or the distance from the center of the field of view. A simple illustrative example of weighting is provided in Figure 3.10, in which edges were deemed most perceptually valuable. In the perceptually weighted case, object edges (detected with a Sobel filter) were assigned a high weight
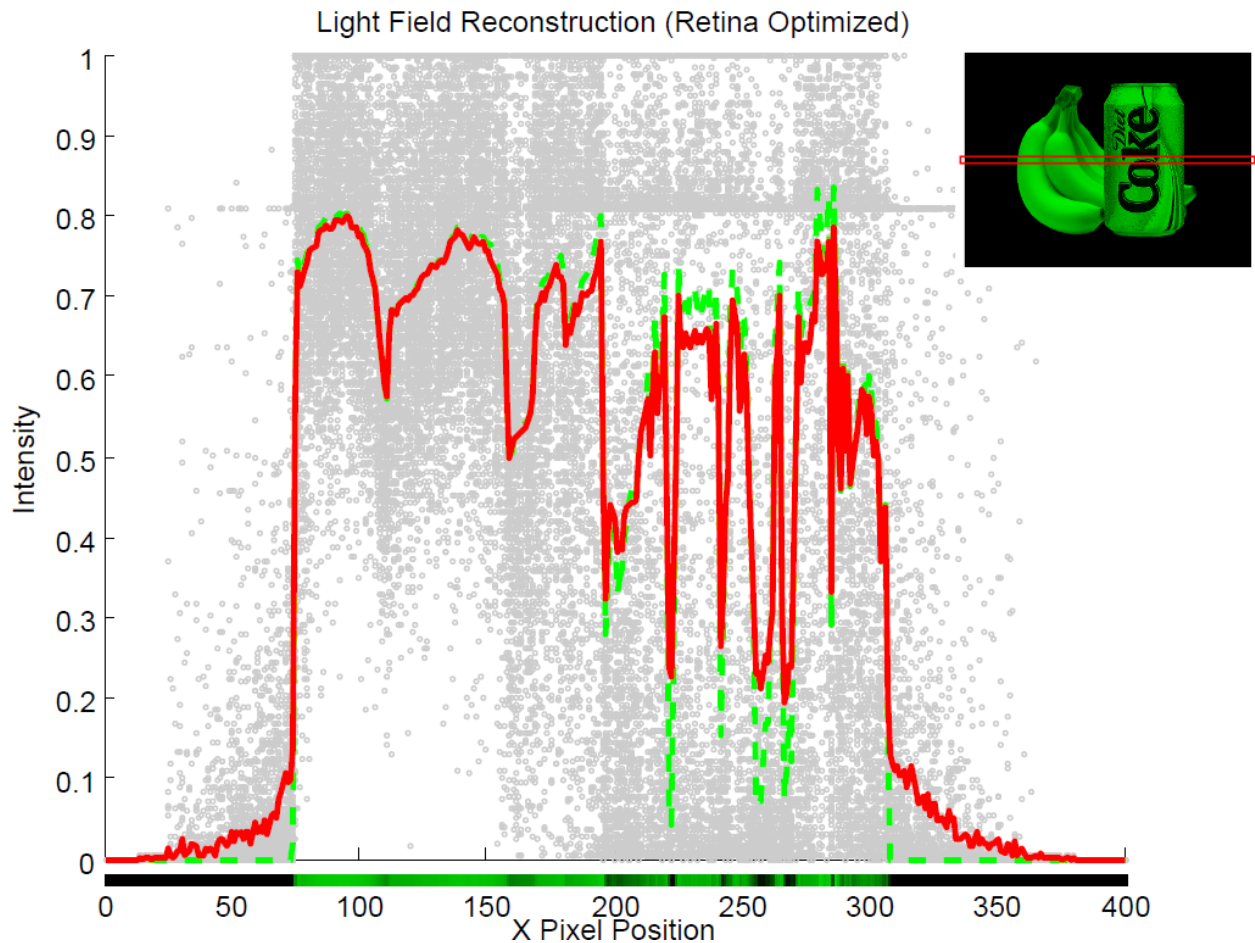
Figure 3.8: Improved noise cancellation with retinal optimization. Plot provides a 2D slice of a 4D reconstructed "planar" light field (shown below x-axis and in top right inset image) described in Figure 3.6, except that retinal optimization was performed as described in Section 3.5.5.2. Note that individual reconstructed views (gray dots) have extreme noise when compared to intended the signal (green dashed lines); however, the average of views over the pupil (red solid line) more closely matches the intended signal than with conventional optimization (Figure 3.6).
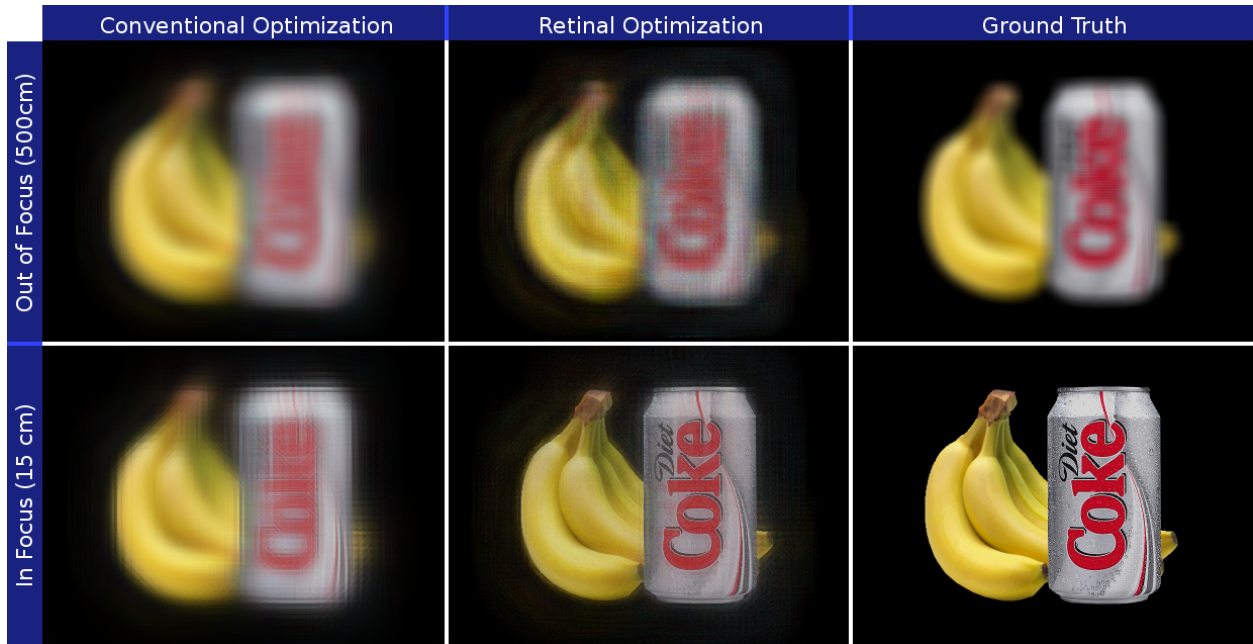
Figure 3.9: Qualitative comparison of conventional and retinal optimization. *Left column:* Simulated views reconstructed with conventional optimization. *Center column:* Simulated views reconstructed with retinal optimization. *Right column:* Original light field, provided as input to optimization. *Top row:* Light field synthetically refocused at far distance (500 cm). *Bottom row:* Light field refocused at close distance (15 cm). Note improved sharpness of retinal optimized image (bottom center) as compared to conventionally optimized image (bottom left) for the in focus case. Optimization was performed with two layers, eight time-multiplexed frames, brightness scaling factor $\beta = 0.1$ (see Section 3.5.4) and the geometry of the prototype display (see Section 3.6).

Figure 3.10: Perceptual Optimization. *Left:* Simulated light field reconstruction using uniform ray weighting. *Center:* Simulated reconstruction using ray weighting based on contrast with adjacent rays. *Right:* Original light field, used as input to the optimization process. Inset images show magnified region of topmost flower. Note improved contrast of green stem and leaves when using edge prioritized weights (center) as compared to use of uniform weights (left). Optimization was performed with two layers, eight time-multiplexed frames, brightness scaling factor $\beta = 0.1$ (see Section 3.5.4), a focal distance of 15 cm and the geometry of the prototype display (see Section 3.6).

(~ 0.75) while all other rays were assigned a low weight (~ 0.25). Note that the perceptually weighted case (center) appears to have higher contrast than the uniformly weighted case (left), at the expense of more noise in low contrast areas.

3.5.5.4 Occlusion Masks

As described previously, the modulation layers can be purposed both to form augmented images and provide an occluded view of the real environment. Occlusion masks are also specified as light fields (or more precisely *light field occluders*) and are optimized into sets of time-multiplexed layer patterns so that they appear in focus and at the correct focal depth.

Occlusion masks are typically used to block real light rays behind augmented objects so that the augmented objects appear opaque. In this scenario, the occlusion mask light field is a copy of the augmented image light field where object rays are assigned an intensity of 0 (opaque) and all other rays are assigned an intensity of 1 (transparent). Object rays can also be assigned non-zero values for partial transparency. Using this scheme, a light field occluder can be optimized using the same optimization process described in Sections 3.5.4 and 3.5.5. An example occlusion mask is presented in Figure 3.17E.

3.5.6 Practical Implementation Details

In this section, additional optical considerations relevant to creating a practical, human viewable multilayer near-eye display are discussed.

3.5.6.1 Eyebox

As described in Section 2.1.3, a key parameter of a near-eye display is the eyebox, the region over which the eyes can be positioned to see a complete image. In the proposed design, the expected location of the eyes with respect to the display is defined in software and can be adjusted for each user; however, in the present work it is assumed that there is no relative motion between the eyes and the display (i.e. the eyebox matches the eye pupil size). In a practical display, the eyebox should be larger to account for eye movement. One possibility for increasing the eyebox size is to enlarge the constrained region during optimization, although some image degradation is expected as there are more constraints to satisfy. It is also expected that less noise cancellation will occur (see Section 3.5.5) as the complete set of rays that contribute to the noise cancellation may not enter the eye. Another possibility is to track the eyes; a promising approach is to place a small camera on the edge of the backlight waveguide that can see the eyes through total internal reflection, as proposed by Travis et al. (2013).

3.5.6.2 Spatial Light Modulator Performance

The proposed design places demanding constraints on spatial light modulator performance that are not met by current hardware. As shown in the simulations (see Section 3.7.1), generally eight or more time-multiplexed frames are needed for adequate performance, requiring 480 Hz modulators for smooth motion at 60 Hz. However, current LCD panels are limited to 240 Hz rates or less. Additionally, if the occluded real image and augmented real image phases are time-multiplexed, high speed modulation is required to ensure that these phases are reasonable synchronized. For example, for a moderate head movement speed of $50°$/sec and modest spatial resolution of 10 pixels/degree, a 500 Hz modulator is needed to ensure that the occlusion mask is approximately aligned to the augmented content within 1 pixel. The augmented image and occlusion masks are typically a multi-frame sequence, further exacerbating this demanding requirement. High light efficiency is also an important consideration for a see-through design. Unfortunately, much light is lost through the polarizers of stacked LCDs. Finally, to provide high spatial resolution over a wide field of view, very high

density modulators are needed in moderate sizes (e.g. the size of eyeglasses lenses); however, current very high density displays are limited to small microdisplays.

Upcoming display technologies may alleviate these limitations. A recent spatial light modulator technology, digital microshutters (Hagood et al., 2007), provides a claimed $10\times$ light efficiency advantage over LCDs, binary switching rates in the kHz range, and has been demonstrated in a mobile form factor. Samsung also now manufactures LCDs with high transmittance for transparent applications[3]. Eyeglass lens sized, high density LCD displays are also beginning to become available; Japan Display Inc. has demonstrated a 2.3 inch 651 dpi panel[4]. Microdisplays are also now available with extremely high densities; Silicon Microdisplays sells a 3,000 dpi microdisplay available in commodity products[5].

3.5.6.3 Diffraction

Light that passes through an aperture expands angularly (diffracts) to a degree inversely related to the aperture size. Diffraction is burdensome in the proposed design as the viewer looks through a spatial light modulator that consists of an array of very small pixel apertures, causing degradation of both the real and augmented views, as described below.

During augmented image formation, diffraction limits angular resolution. For a rectangular pixel aperture, the first diffraction minimum occurs at the angle position

$$\theta_d \approx \frac{\lambda}{w}, \tag{3.6}$$

where λ is the wavelength of the light and w is the diameter of the pixel aperture. For 550 nm light and the configuration of the prototype display (a 770 dpi spatial light modulator placed 18 mm from the eyes), the diameter of the central diffraction order becomes 0.6 mm at the user's eyes. This spot size is significantly larger than that predicted by the display's theoretical angular resolution as described in Section 3.5.5.1, but still is considerably smaller than the size of a human pupil, allowing angular variation over the eye.

During real image formation, light from the real environment will diffract to varying degrees by wavelength when passing through the pixel apertures. When light from a point in the scene is diffracted at the

---

[3]http://www.samsung.com/in/business/business-products/smart-signage/specialized-display/LH22NLBVLVC/XL

[4]http://www.j-display.com/english/news/2012/20120604.html

[5]http://www.siliconmicrodisplay.com/st1080.html

layers, it will no longer remain a point when focused by the lens of the eye, causing ghosting and rainbowing effects. See Section 4.5.5.6 for more analysis of diffraction through pixel apertures, and see Sections 4.5.5.2 and 4.7.4 for suggestions to mitigate diffraction effects.

3.6 Implementation

3.6.1 Hardware

As shown in Figures 3.11 and 3.12, a prototype display was constructed from two small stacked spatial light modulators (SLMs) and a backlight (per eye), and a 3D printed eyeglasses frame. Monochrome SLMs were used in conjunction with a color sequential RGB backlight, allowing for larger physical pixels at the same effective resolution to reduce diffraction. LCD spatial light modulators were obtained from an Epson Powerlite 5000 3LCD projector[6] and have a 26.4 mm $\times$ 19.8 mm active area, an 800$\times$600 resolution and an 85 Hz refresh rate. To prevent a set of crossed polarizers from blocking light through the display, the integrated rear polarizer was removed from the rear LCD, and polarizers were installed at the front and rear of the unit so that there was a alternating set of linear polarizers surrounding each LCD. The LCDs were stacked 8 mm apart center-to-center with spacers in between. To construct the backlight, five RGB LEDs (for field sequential color) were installed on an edge lit waveguide that was extracted from the backlight of a larger LCD panel and cut to the necessary size. The LCD and backlight assemblies were installed in a custom eyeglass frame, designed in Google SketchUp and printed on a Rostock MAX 3D printer (see Figure 3.12). The LCDs were driven by the original projector hardware (housed externally), and the backlight was connected to an external power supply. An Intel Core i7 PC with an NVIDIA GeForce GTX 580 GPU running Linux was used both to generate the display patterns and drive the prototype display over a VGA link.

Note that for convenience a shutter was not installed on the rear of the unit to switch between the augmented and real image formation phases (see Section 3.5.3); instead, an opaque sheet was placed behind the display when photographing the display in the augmented image phase (see Section 3.7.2). A liquid crystal shutter from a pair of LCD shutter glasses would be a likely candidate to be installed in a future device, and would add approximately 1 mm thickness and some additional light attenuation.

---

[6]These panels are obsolete; it is recommended that more modern panels with better performance are used – see Section 4.6.1.2.
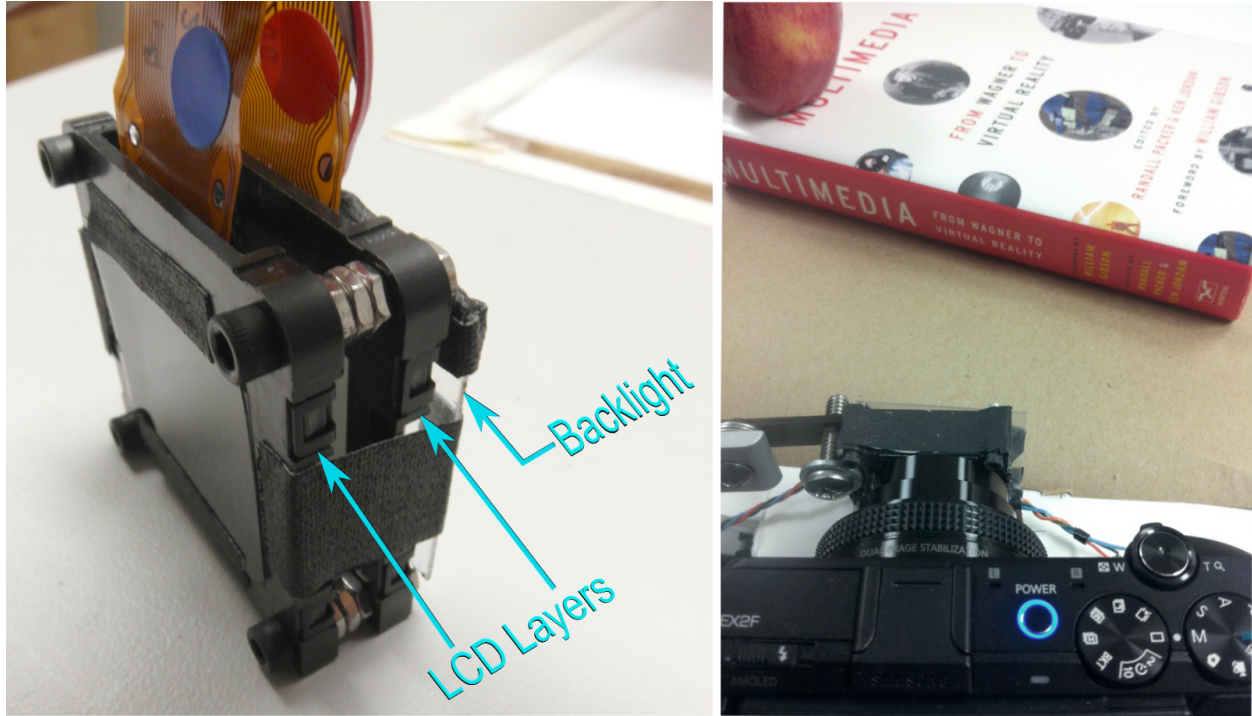
Figure 3.11: Prototype hardware and testing configuration. *Left:* Prototype display unit for one eye, consisting of two stacked LCD microdisplays and an RGB backlight. *Right:* Testing configuration for experiments described in Section 3.7.2. The camera setup was configured to approximate a human wearing eyeglasses: the camera lens was placed nearly in contact with the display so that the display to camera center of projection distance was 18 mm, and a 3.7 mm aperture was used.

The total thickness of the prototype device is approximately 14 mm: 9 mm for the LCDs and spacing, 1 mm for the backlight, and 4 mm of LCD housing that could be removed in a future device. With an eye center of projection to display distance of 18 mm (used for testing in Section 3.7.2), the display provides a $65°$ diagonal field of view, limited by the selected LCD panel size.

3.6.2 Software

All light fields used as input for layer optimization for simulations or prototype display consisted of one or more diffuse planes located at a distance of 10 cm to 500 cm from the eye, as indicated in each figure. A light field resolution of $800{\times}600$ was used for results shown on the prototype display and a resolution of $400{\times}300$ was used for the simulations. The optimization was constrained at $11{\times}11$ equally spaced zones over a 4 mm pupil. Unless otherwise indicated, all results use eight time-multiplexed frames, a brightness scaling factor of $\beta = 0.1$ (see Section 3.5.4) and the geometry of the prototype display described above. All results also use
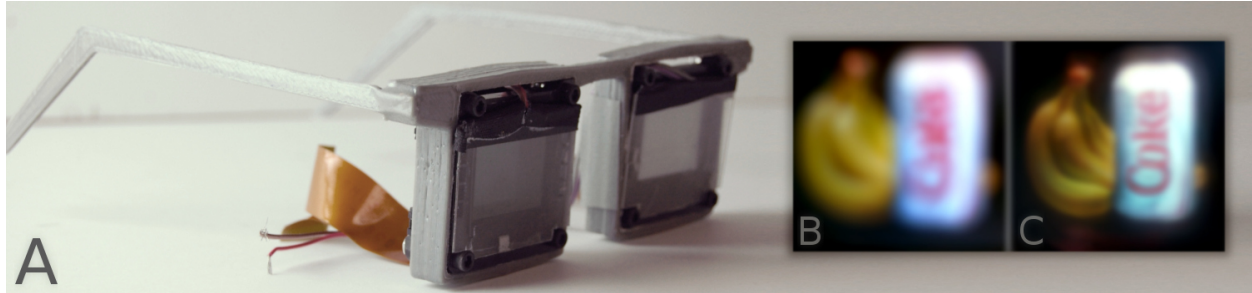
Figure 3.12: Multilayer prototype display and sample result. A) Prototype display (driving electronics are external). B,C) Photos taken of the prototype display with a camera that approximated a human viewer wearing eyeglasses. Photo B) was taken with an image naively displayed on one of the layers, while photo C) was taken using the multilayer optimization technique described in this paper.

an eye to display distance of 18 mm, which corresponds to the distance between the display and the camera's center of projection when the display was tested (see Section 3.7.2).

Optimization was performed using the GPU-accelerated nonnegative tensor factorization routines of Tensor Displays (Wetzstein et al., 2012), modified to accommodate near-eye displays as described in Section 3.5.5. Using the parameters described above, simulation results were performed at a rate of approximately 100 iterations per minute; acceptable results were reached at approximately 150 iterations and near convergence at approximately 500 iterations. Recent work in adaptive display optimization has shown the possibility of increasing optimization performance significantly (Heide et al., 2013).

3.7 Results

3.7.1 Simulated Results

Configuration The input light field of the simulations is a diffuse plane located at a distance of 15 cm from the eye, which corresponds to the augmented overlay tested on the prototype display (see Section 3.7.2). The simulations were performed using the geometry of the prototype display (see Section 3.6), with any additional layers beyond two being equally spaced within the same total layer separation distance (8 mm). Simulations were performed across two key display parameters: the number of display layers and the number of time multiplexed frames.

Simulated images were generated by reconstructing the light field from the 11×11 constrained viewing positions while summing over the set of time-multiplexed frames. The reconstructed views were averaged to produce the perceived image when the eye is focused at the apparent depth of the imagery (15 cm).

Qualitative Results Figure 3.13 shows a qualitative assessment of performance for the various optimization steps described in Section 3.5.5. When only basic optimization is performed (top right), the image is very noisy and barely recognizable. When planarization is performed and the image is summed over the retina (bottom left), much of the noise cancels but high frequency details are still lost. When retinal optimization is also performed (bottom right), much of the high frequency details are recovered, but contrast is lower than the target image.

Figure 3.14 shows a qualitative assessment of performance over various display configurations. Note that a recognizable image can be formed even in a very modest configuration (two display layers, two time-multiplexed frames), although perceived sharpness and contrast is low and there is a glow extending outside the image. As the number of layers and frames increase, the perceived sharpness and contrast improve, and the glowing recedes. In a more advanced configuration (3 layers, 12 frames), image quality is further improved; however, there is still a noticeable loss of contrast as compared to the ground truth image.

Figure 3.15 also provides a qualitative comparison of simulated occlusion masks with and without optimization, showing a significant increase in sharpness when optimization is enabled. However, the optimized mask still retains some softness at the edges and has significantly reduced contrast.

Quantitative Results Figure 3.16 shows a quantitative measure of performance for the formed retinal image, expressed in peak signal-to-noise ratio (PSNR), over various display configurations. The PSNR calculations exclude the black background area (see Figure 3.14) to prevent overinflating the PSNR in unconstrained regions and overpenalizing the PNSR in the glowing areas that extend beyond the image (since the true background when displayed is not black, but rather the real scene behind the display). The plot shows that the proposed approach is clearly lossy; however, with sufficient layers and frames performance on the order of lossy video compression ($\geq$30 dB) can be achieved. Note that even in the highest bandwidth configuration plotted (4 modulation layers, 20 time-multiplexed frames), the display is compressive: the $11 \times 11 = 121$ constrained viewpoints over the pupil are compressed into $4 \times 20 = 80$ times the bandwidth of a single viewpoint.
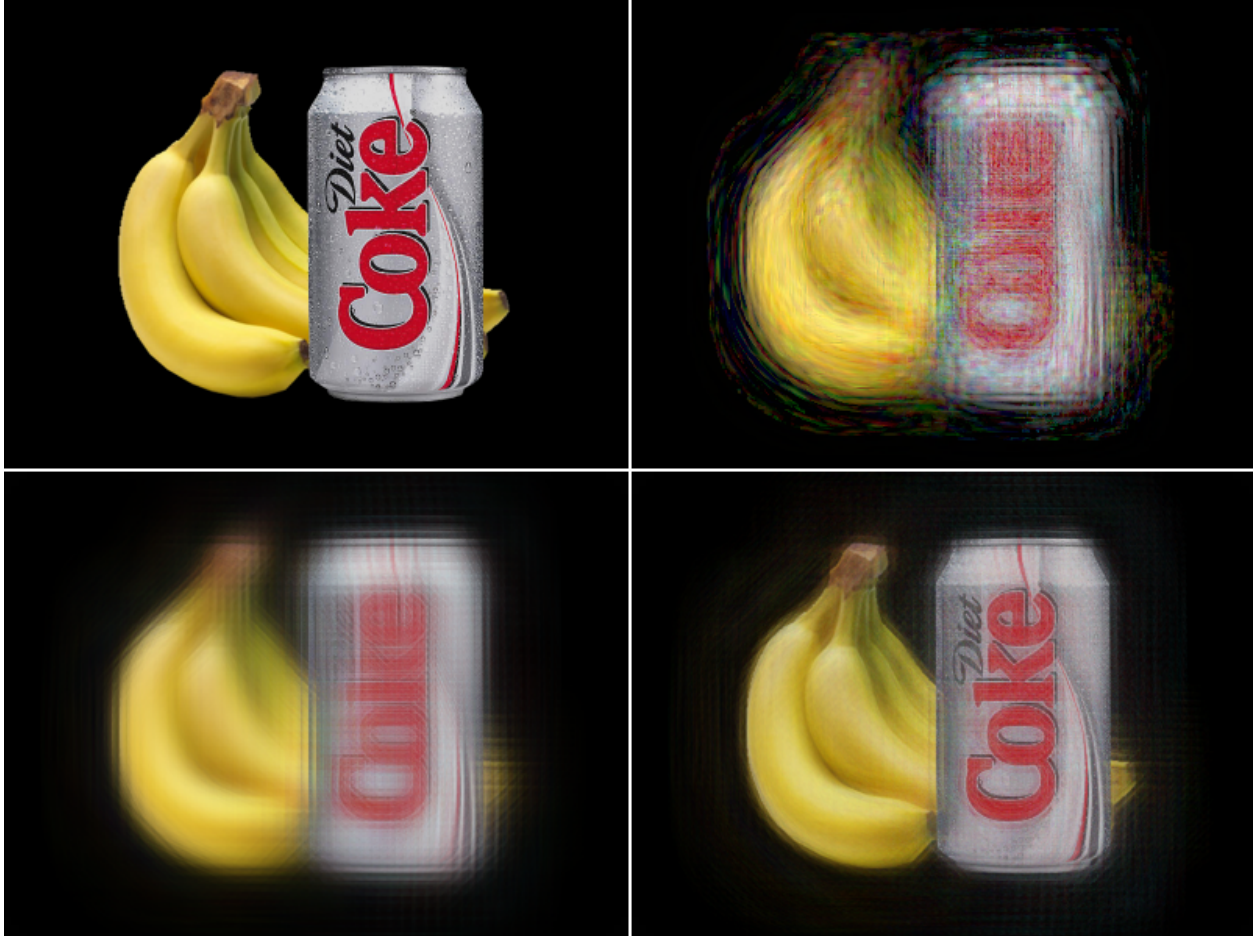
Figure 3.13: Multilayer qualitative performance for various optimization steps. *Top Left:* Target image for optimization. *Top Right:* Simulated reconstructed image from one viewpoint on pupil, representing the approximate perceived image when no additional optimization steps are performed. *Bottom Left:* Simulated summed image over retina when light field is planarized. *Bottom Right:* Simulated summed image over retina when light field is planarized and retinal optimization is performed. The optimized simulations use two layers, eight time-multiplexed frames, and a brightness scaling factor $\beta = 0.1$ (see Section 3.5.4).
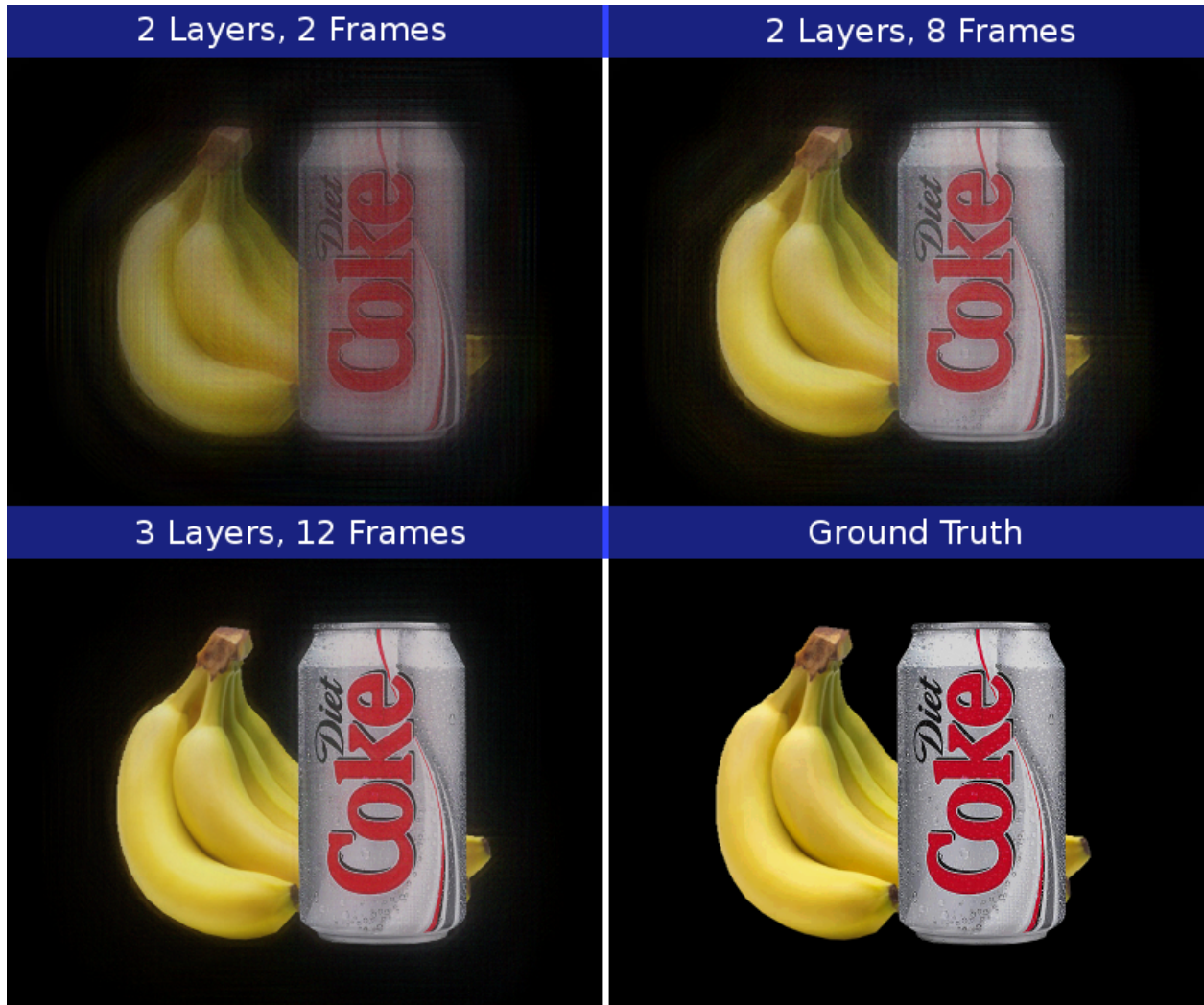
Figure 3.14: Qualitative performance measure across parameter space. Each image is a simulated view reconstruction with a varying number of display layers and time-multiplexed frames, as indicated above each image. All optimizations used brightness scaling factor $\beta = 0.1$ (see Section 3.5.4), a focal distance of 15 cm, and the geometry of the prototype display (see Section 3.6), with any additional display layers above two equally spaced to fit within the original layer separation distance.
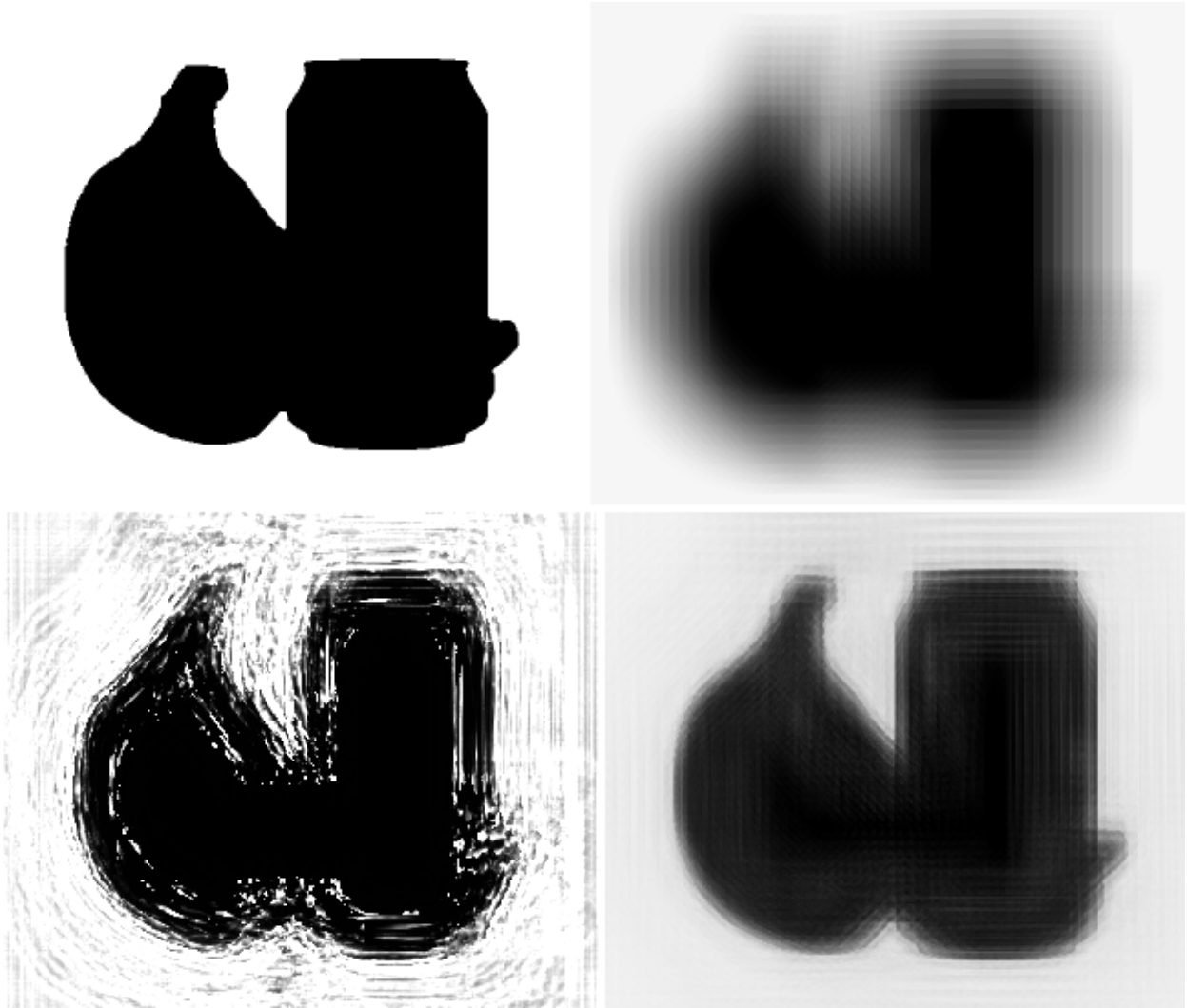
Figure 3.15: Qualitative performance measure for world occlusion mask. *Top Left:* Target occlusion mask (where white pixels indicate "clear"). *Top Right:* Simulation of occlusion mask displayed without optimization (i.e. naively displayed on panel). *Bottom Left:* Simulation of optimized occlusion mask from one constrained viewpoint over pupil. *Bottom Right:* Simulation of optimized mask, summed over the 121 constrained viewpoints over the pupil. The optimized simulation used two layers, eight time-multiplexed frames, and a brightness scaling factor $\beta = 0.1$ (see Section 3.5.4). The simulations used a focal distance of 15 cm and the geometry of the prototype display (see Section 3.6).

Image Quality (PSNR) for Various Display Configurations
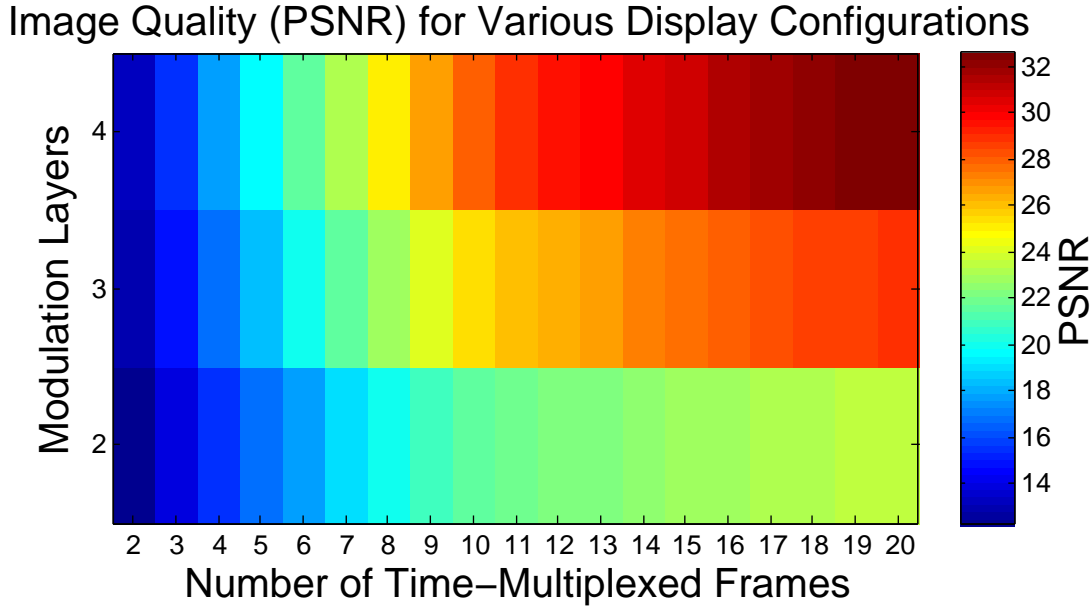
Figure 3.16: Quantitative performance measure across parameter space. Plot shows reconstruction accuracy for the retinal image, expressed as peak signal-to-noise ratio (PSNR), compared to a reconstruction of the original light field used as input to the optimization process (Figure 3.14, lower right). All optimizations used brightness scaling factor $\beta = 0.1$ (see Section 3.5.4) and the geometry of the prototype display (see Section 3.6), with any additional display layers over two equally spaced to fit within the original layer separation distance.

For the selected configuration shown in Figure 3.13 (two layers, eight time-multiplexed frames), the PSNR (as calculated above) is 16.6 dB for the basic optimization case (top right), 17.6 dB for the planarized case summed over the retina, and 20.0 dB for the retinal optimized case. These results show that the near-eye multilayer optimization techniques described in Section 3.5.5 can yield a significant advantage in numerical accuracy.

For the simulated occlusion masks (Figure 3.15), the PSNR over all pixels in the image is 12.8 dB for the unoptimized image (top right) and 16.2 dB for the optimized image (bottom right) when compared to the reference mask (top left). Note that the PSNR in the optimized case remains low because of the reduced contrast, but in the author's opinion it is perceptually much sharper than the unoptimized image.

3.7.2 Prototype Display

Configuration To test the prototype (described in Section 3.6), a camera was placed behind one of the backlit LCD units and photographs were taken through the display (see Figure 3.11, right). To determine the relative

position and orientation of the camera and the display, calibration patterns were displayed on the layers and extrinsic parameters were computed using OpenCV. The camera center of projection to display distance was determined to be 18 mm when the lens was placed nearly in contact with the display, approximating the typical distance between eyes and eyeglasses. The camera was set to its maximum aperture of 3.7 mm, within the range of human pupil diameters. To produce color images with the monochrome LCDs, three color sequential exposures were taken – each with the backlight set to one of the RGB primaries and the corresponding channel displayed on the LCDs.

Several accommodations were made for the limitations of the LCD panels. To simulate a faster display that would allow more time-multiplexed frames within the flicker fusion threshold, the camera was set to the least sensitive setting (ISO 80) and a long exposure of $1/4$ sec was used. Additionally, since the older panels have a slow response time, each pattern was displayed for two display frames to give the liquid crystals more time to stabilize. Another shortcoming of the LCDs is their poor viewing angle – a problem when viewed at extremely close range – which caused pixels to appear increasingly dark toward the bottom left corners of the LCDs. To compensate, a higher brightness scaling factor of $\beta = 0.4$ was used, which shifted displayed colors toward white at the expense of compression performance. It is expected that these problems will be ameliorated with the use of modern projector LCD panels, which have faster refresh rates, lower response times, and wider viewing angles through the use of vertically aligned (VA) or in-plane switching (IPS) liquid crystals. The panels used in a subsequent display design (see Section 4.6.1.2) showed improvement in switching times and viewing angles.

Augmented Reality Scene Results A first test measures the ability of the prototype to display an optical see-through augmented reality scene. An augmented image (see Figure 3.14, bottom right) was optimized to appear at a distance of 15 cm from the camera, in front of a real apple and book placed on a desk (see Figure 3.11, right). The camera was set for focus at 15 cm, which corresponds roughly to the depth of the spine of the book. The scene was shot in two combined sets of RGB exposures – one of the augmented image and one of the occluded real scene – to simulate the time-multiplexed operation described in Section 3.5.3. To keep the combined result within $0 - 255$ intensity range, the exposures E_1, E_2 were combined with the following formula:

$$I = 255 - \frac{(255 - E_1)(255 - E_2)}{255} \tag{3.7}$$
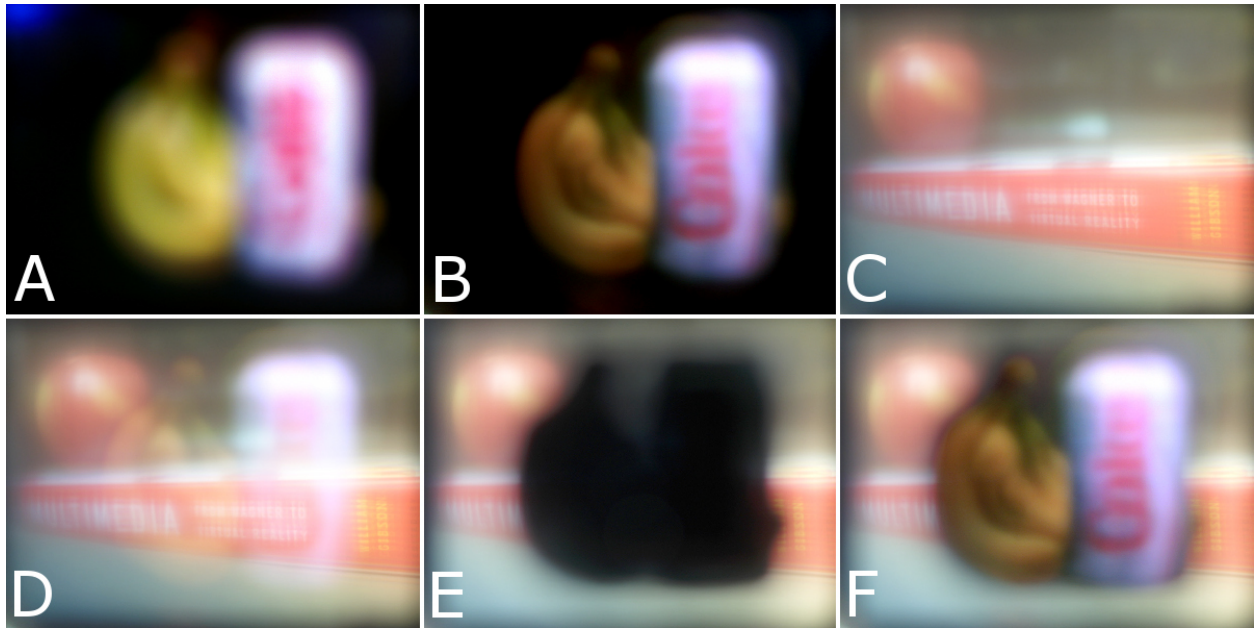
Figure 3.17: Augmented reality scene on prototype display. *A)* Photograph of prototype showing image naively displayed on farthest layer. *B)* Photo of prototype displaying a series of optimized time-multiplexed patterns. *C)* Photo of real environment taken through the display with backlight off. *D)* photographs (B) and (C) combined to simulate an augmented display without occlusion support. *E)* Photo of real environment with an optimized occlusion mask displayed on the device. *F)* photos (B) and (E) combined to simulate an augmented scene with occlusion support.

The results are shown in Figure 3.17, which demonstrates the ability of the prototype display to meet three key objectives. First, the display is able to create a (moderately) focused image using layers placed closer than the camera focus setting. Note that the optimized image (B) is much clearer than an image naively displayed on the layer farthest from the camera (A) when the camera is focused into the scene. Second, the display exhibits occlusion of real objects (F); the augmented banana and soda can appear solid and in front of the real apple and book. Finally, the synthesis of the augmented scene (F) shows the ability to combine an occluded real scene and augmented image using a shared set of display layers.

Despite these promising results, the prototype also shows several shortcomings. The augmented image (B), although recognizable, is not nearly as sharp as the simulated result (see Figure 3.14, top right). The generated occlusion mask (E) has blurry edges, causing the augmented image to appear surrounded by a black halo (F). Finally, the view through the display (C) is somewhat blurry due to diffraction through the LCDs and to some diffusion in the backlight. Prototype results may be improved by utilizing LCD panels with wider viewing angles and replacing the backlight with a completely clear OLED lighting panel.

Multiple Focus Results A second test measures the ability of the prototype to display objects at multiple focal depths simultaneously. The left half of the display shows a pattern optimized for a 10 cm focal depth, while simultaneously the right half shows the same pattern optimized for a 500 cm focal depth. Two sets of RGB exposures were taken which varied only by the camera's focus setting, which was adjusted between 10 cm and 500 cm.

The results, shown in Figure 3.18, demonstrate the ability of the prototype to display images at varying focal depths simultaneously: each pattern is in moderate focus when the camera is focused at the intended focal depth, and out of focus when the camera is focused elsewhere. These results show some promise; however, in the future it would be useful to demonstrate this ability with overlapping objects and more challenging high spatial frequency patterns.

3.7.3   Assessment

A compact, see-through AR display design supporting a wide field of view and focal depth cues was proposed and demonstrated in some preliminary results. However, the quality of the simulated results is significantly better than those achieved on the prototype display. This discrepancy could be reduced through more accurate modeling of the display in the optimization and simulation software. Since the number of constrained viewing zones (11×11, limited by GPU memory) is less than the theoretical maximum angular resolution, optimization may benefit from an even denser set of view constraints. Recent work in display optimization (Heide et al., 2013) provides a time and memory efficient optimization method for very dense sets of view constraints. Simulation and optimization may also benefit from a better model of the spatial light modulators that considers off-axis viewing performance, diffraction, and the varying display intensities across layers. Note that diffraction through the panels was more significant here than through more modern panels used in a subsequent display design (see Section 4.6.1.2). Finally, the hardware prototype could also benefit from improved layer calibration; camera-based layer calibration was challenging due to the fine pixel pitch of the utilized spatial light modulators and the close proximity of the camera to the display (yielding slightly unfocused images at the nearest focal setting of the lens).
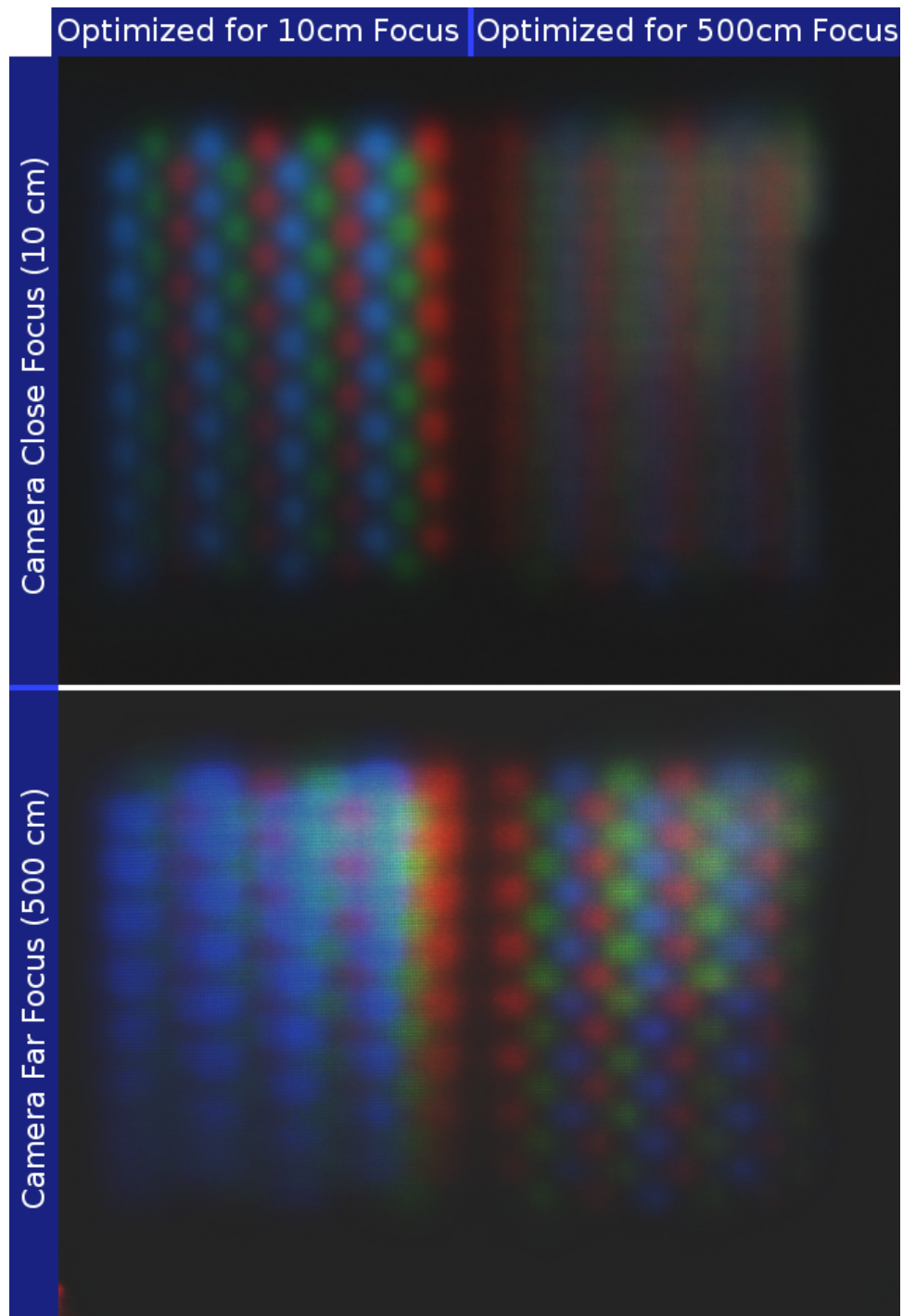
Figure 3.18: Multifocal scene on prototype display. An optimized set of patterns was generated to simultaneously display a tiled RGB pattern at two focal depths, 10 cm (*left* half of photographs) and 500 (cm) (*right* half of photographs). *Top:* Photograph with camera focus set to 10 cm. *Bottom:* Photograph with camera focus set to 500 cm.

CHAPTER 4

Additive Approach: Pinlight Displays

4.1 Introduction

As described in Section 3.1, most near eye displays form the perceived image optically as a set of *point* sources, each of which represent a "pixel" in the image. The point sources are typically emitted over a wide angle from a small display panel (e.g., an LCD microdisplay) and collected by the eye. The wide emission angle is useful to allow the eye to see the displayed image even with eye movement. If the image is to appear in focus, the rays emitted from each point source must appear to intersect at the focal plane of the eye so that the lens of the eye can focus them to a small spot on the retina. The rays can be adjusted to intersect at a more distant location than the display panel if a suitable refractive, diffractive, or reflective element is placed in front (see Figure 4.1, top). However, such hardware limits the field of view or adds bulk to the near-eye display. Additionally, if a transparent display is desired, such elements which obscure or distort the see-through view must be placed outside of view and optically recombined or corrected with additional elements, further adding bulk and limiting field of view.

The Pinlight Displays approach (Maimone et al., 2014b) provides a method of image formation that is different than other known optical see-through near eye displays. In the Pinlight method, each point in the image is formed not as a point source but as a highly directional *"ray"*[1] source; rather than attempt to reproduce all rays emitted from a virtual point, only a single "ray" is reproduced. The ray sources are generated from a sparse set of point light sources (called *pinlights*) that are modulated by a single spatial light modulator (see Figure 4.1, bottom). With this method it is much simpler to produce a coherent focused image: it is not possible for a single "ray" source to go out of focus as it will always fall on a small spot on the retina regardless of how it is refracted by the eye's lens. Thus no additional optics are needed to collect

---

[1]In this context, a "ray" is meant as a narrow beam of light that has a width or angular extent on the order of the size or angular extent of one pixel in the display.
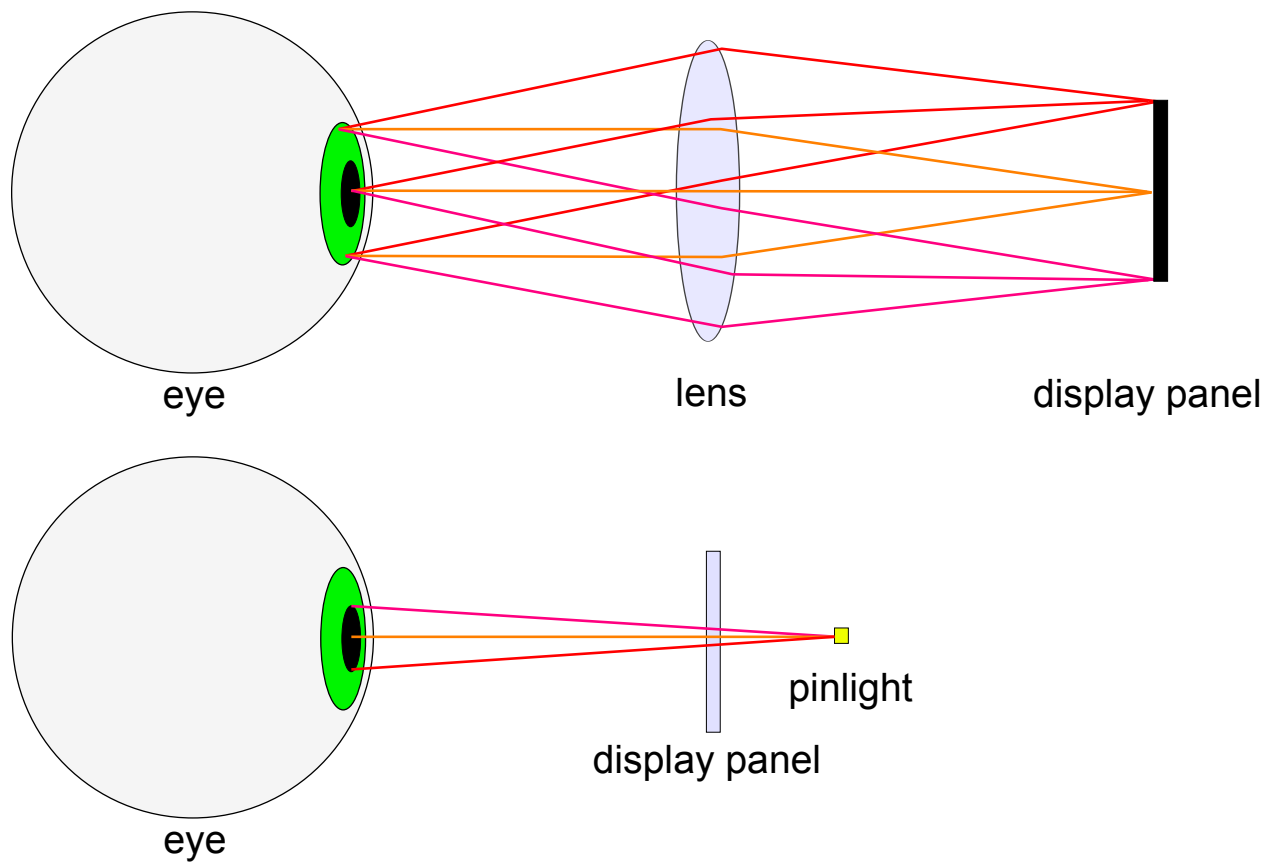
Figure 4.1: Pinlight method compared to conventional optical design. *Top:* In a typical optical design, bundles of rays emitted from each pixel on a display panel are brought into focus with an optical element. (Here a lens collimates the bundles to create an image at optical infinity.) *Bottom:* In the pinlight display design, a bundle of rays from a point light source is modulated by a transparent display panel so that each display pixel is represented by a single "ray".

and focus the light emitted from the display, eliminating much of bulkiness and field of view limitations of the display. In essence, the pinlight display hardware acts as a series of miniature projectors aimed directly into the eye. Additionally, since all of the components of the pinlight display hardware are transparent[2], a see-through display can also be achieved with a single shared optical path, eliminating the majority of field of view limits associated with optical recombination hardware. In contrast to the previously discussed multilayer *subtractive* approach (see Chapter 3) in which one starts with all possible rays and removes the unnecessary ones, the pinlight approach is *additive*: a near minimum number of rays needed to form the desired image are generated.

4.2   Contributions

Pinlights Displays is a novel approach to see-through near-eye displays. Specific contributions include:

- The use of point light sources, modulated with an LCD and placed *near the eye*, that act as *imperceptible, see-through* projectors

- The use of such projectors in a *tiled* configuration encoded with a *virtual aperture* to expand the FOV of the display

- The use of such tiled projection arrays in an alternative configuration to provide a *light field* over the eye, as a *see-through* alternative to existing near-eye light field displays

- Example *hardware designs* for creating transparent tiled projection arrays, described and evaluated in two prototype devices

4.3   Benefits and Limitations

The proposed design offers several benefits over existing see-through near-eye displays. A wide FOV is supported with no theoretical upper bound (or approaching $180°$ if the display components are planar), and a prototype display achieving a $110°$ diagonal FOV in each eye is demonstrated. The design also supports a lightweight, eyeglasses-like form factor and uses simple, low cost hardware components.

---

[2]In practice, the hardware components are *partially* transparent: the spatial light modulator and pinlight array may have small regions that are opaque or attenuate light, causing some light loss in the display and diffraction effects as described later in the chapter.

The proposed design also exhibits several limitations. Precise pupil tracking, while not strictly required, strongly benefits the possible image quality and resolution. Diffraction limits image quality and provides a challenge to scaling beyond modest image resolutions. The brightness of the environment is also dimmed through the display when LCD spatial light modulators are used (which are practical and low cost), providing the effect of wearing sunglasses. The see-through capability is also reduced if the augmented image content fills a small area of the display; however, this effect can be mitigated with a more complex design. Some lightweight computation ($\approx$1 ms on a high-end GPU of 2014) must be performed to generate an image for the display, and accurate geometric and radiometric calibration of the display is needed to eliminate tiled periodic structures from appearing in the perceived image. See Section 4.7.4 for additional discussion of current limitations and possible solutions.

4.4   Related Techniques

The pinlight approach is related to work in defocused projection systems. In the Bokode (Mohan et al., 2009) system, Mohan et al. use a combination of a backlit, microprinted transparency and a microlens to create a compact 2D barcode visible to a distant defocused camera. Hiura et al. (2010) extend this approach to curved arrays of Bokodes that operate like compound eyes. In contrast to the pinlight approach, this approach relies on refractive lenses and is intended to be distantly located from an imager.

Pamplona et al. (2010, 2011) demonstrate near-eye applications of these principles to interactively assess the refractive errors in human eyes. A near-eye light field display is created by covering a microdisplay with either a microlens array or a pinhole grid to present test patterns containing accommodation cues. More recently, Lanman and Luebke (2013) show that such displays can be optimized for general applications to make compact virtual reality (VR) eyeglasses. While sharing conceptual similarities to the pinlight design, neither approach directly facilitates optical see-through applications as refractive microlens arrays distort the see-through view.

There are also similarities with light field displays that exploit *parallax illumination*. In a related work, Son et al. (2007) substitute a 2D point light source (PLS) array for the microlenses and pinholes routinely found within light field displays. There are also similarities in the pinlight approach to projector-based displays. Jurik et al. (2011) describe a light field display in which each projector in a dense array acts as a single display "pixel" with a high degree of angular variation. However, unlike pinlight displays, PLS arrays

and such projection arrays are designed to function within the accommodation range of the observer (e.g. as a 3D TV). The pinlight design instead explores making these arrays imperceptible and suitably configured for defocused near-eye use; instead of the LED array of Son et al. (2007) or the pico projectors of Jurik et al. (2011), an edge-lit etched acrylic sheet is used.

4.5 Display Design and Analysis

4.5.1 Components

The proposed design relies on the following principal hardware components:

1. A high-speed, high density, *transmissive* spatial light modulator (SLM) that controls the intensity of passing light through attenuation, e.g. a transparent LCD panel

2. A array of point light sources (with spacing on the scale of the human pupil diameter, $\approx 1 - 3$ mm apart) that appears transparent when defocused, e.g. an edge-lit acrylic waveguide with small etched scattering features where the point light sources appear

The components are sandwiched together with a small spacing in between as illustrated in Figure 4.2.

4.5.2 Physical Attributes

The proposed design supports a compact form factor and wide field of view that can approach ordinary eyeglasses. The thickness of the device is determined primarily depending on the required spacing between the pinlight array and SLM (typically 3-15 mm). The field of view of the device is limited only by the size of the spatial light modulator layers, the distance between the eye and layers (typically $\sim$20 mm), and the supported viewing angles of the SLM and pinlight array. A $110°$ diagonal field of view per eye was achieved on a prototype display. The relationships between these factors are quantified in Section 4.5.4.

4.5.3 Single Pinlight Projector

The core approach is directly encoding light from a *defocused* point source placed near the eye (outside of the eye's accommodation range), called a *pinlight*, with a transmissive SLM placed between the eye and the pinlight. This system acts as a miniature projector, which is called a *pinlight projector*, that projects
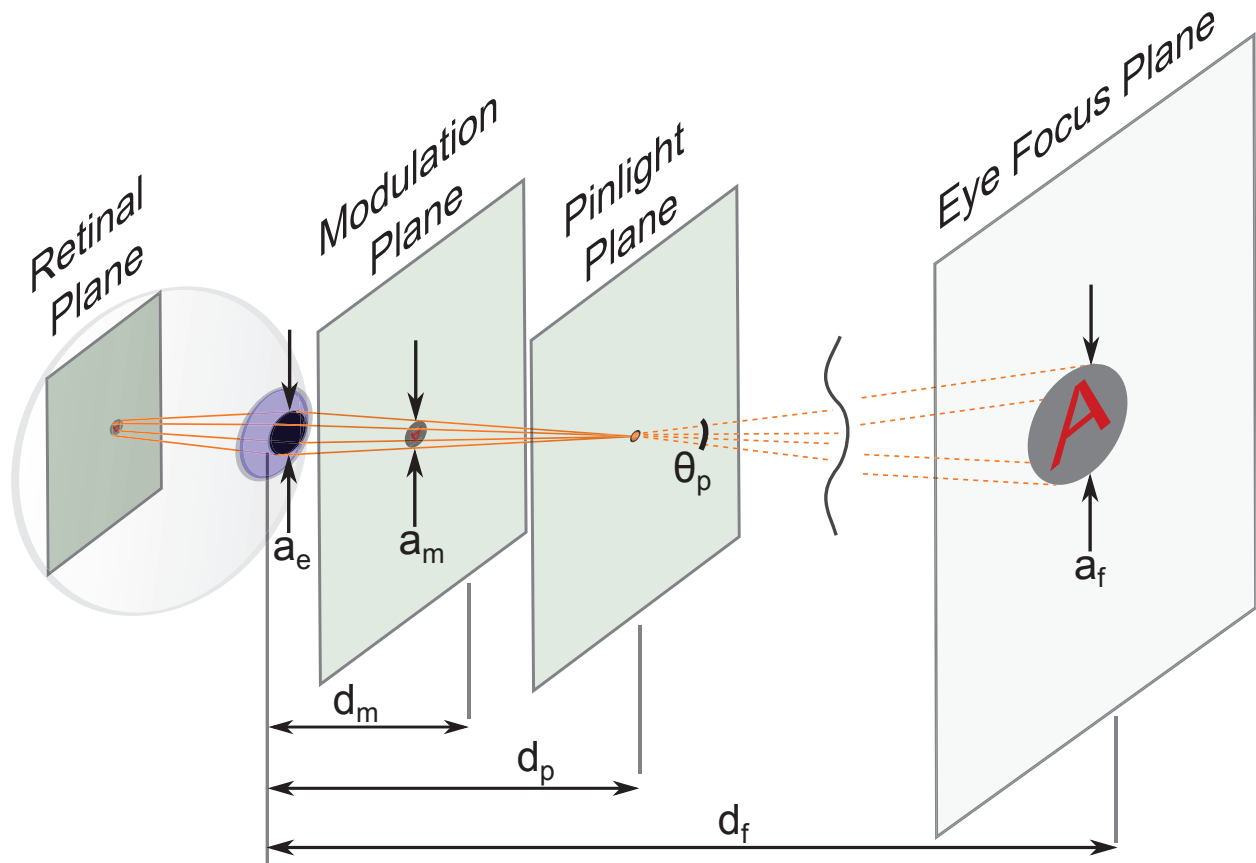
Figure 4.2: Pinlight Projection. A defocused point light source, or *pinlight*, placed near the eye is coded with a spatial light modulator to create a narrow field of view image that appears in focus without the use of refractive or diffractive optics.

light directly onto the retina, as illustrated in Figure 4.2. Assuming for simplicity that the pinlight and the pixels on the SLM are true mathematical points, each location on the SLM receives light from a single direction (a "ray") that is modulated and then enters the eye. The set of all rays is refracted by the lens of the eye; however, since the rays originate from a single point their angular ordering is preserved, creating a "sharp" (i.e. apparently in-focus) copy of the modulated image on the retina. Since the point source is nearer than the viewer's minimum accommodation distance, the image is not flipped by the eye, and thus the modulated image must be inverted along both axes by the display. The image formed on the retina is "sharp" regardless of the focal state of the eye (much as with a camera with a pinhole aperture); the eye's focal state changes only the degree of refraction and therefore the scaling of the image formed. Note that the illumination is simply a coded point source; therefore it also captures the point spread function of the eye, which has some non-uniformity in shape and intensity. Also note that the image produced by a pinlight projector is generally round, due to the shape of a human pupil, which adds complexity to optical design as described in Section 4.5.4. See Section 4.5.5.1 for details concerning the creation of real point sources with non-zero extent and Section 4.5.5.5 for handling changes in eye state. See-through ability is achieved through the use of the previously described transparent components; further details are found in Section 4.5.5.

4.5.3.1 Projection Geometry

From Figure 4.2, it can been seen that the necessary diameter of modulation on the SLM a_m and the diameter of the image on the focus plane a_f for a *single* pinlight projector can be computed as:

$$a_m = a_e \left(1 - \frac{d_m}{d_p} \right), \; a_f = a_e \left(\frac{d_f}{d_p} - 1 \right),$$ (4.1)

given eye aperture diameter a_e, pinlight plane distance d_p, modulation plane distance d_m, and eye focus distance d_f. Likewise, the field of view θ_p through the single pinlight can be computed as:

$$\theta_p = 2 \tan^{-1} \left(\frac{a_e}{2 d_p} \right).$$ (4.2)

The pinlight plane distance d_p and modulation plane distance d_m are the variables under the control of the display designer. From Equation 4.2, observe then that the FOV through a single pinlight projector θ_p is increased by decreasing pinlight plane distance d_p. Given a selection of d_p, modulation plane distance d_m

can be chosen to select the desired modulation scale on the SLM. For a typical eye pupil diameter $a_e = 3$ mm, the pinlight plane set at a practical distance of $d_p = 25$ mm yields a FOV through the pinlight projector θ_p of only $6.9°$. Obtaining a wide field of view of $100°$ would require an impractically close pinlight plane distance of $d_p = 1.26$ mm. However, the use of multiple pinlight projectors is explored to increase the FOV as described in the following section.

4.5.4 Tiled Pinlight Projectors

A single pinlight projector does not alone provide a useful field of view for most applications. However, multiple pinlight projectors may be *tiled* to significantly increase the field of view. In this section, various tiling geometries are discussed.

4.5.4.1 Ideal Tiling Geometry

When tiling pinlight projectors, it is assumed that the pinlight plane now contains an *array* of point light sources, all of which are modulated by a single SLM. Further, it is assumed that the pinlights emit light over a wide angle so that each is visible to the eye (but without concern with light rays that do not enter the eye). In this configuration, the pinlights can be tiled to an arbitrarily wide field of view (or approaching $180°$ if the pinlight array and SLM are restricted to planes[3]) where the total display area is approximately proportional to the number of pinlight projectors used. The field of view is also subject to practical limitations on the emission angle of the pinlights and the size and supported viewing angles of the SLM, but these components can be found with support for very wide angles.

The tiling must satisfy two primary conditions: the eye focus plane must contain a *continuous* tiled image among all the projectors (so that no gaps appear in the image), and the modulated areas on the SLM must be *disjoint* among the projectors (so that no SLM pixel is simultaneously mapped to multiple rays). Resolution should also be maximized by using as much area on the SLM as possible. Considering first an *ideal* one-dimensional case (illustrated in Figure 4.3): given a desired eye aperture a_e, eye focus distance d_f, and selected pinlight plane distance d_p, the optimal pinlight spacing s and modulation plane distance d_m are computed as:

---

[3]Human anatomy and form factor constraints may limit the field of view in practice. For example, with an eye relief of 20 mm and interpupillary distance of 64 mm, the maximum field of view on the nasal side is $\approx 58°$ horizontally before the left and right eye display elements reach each other.
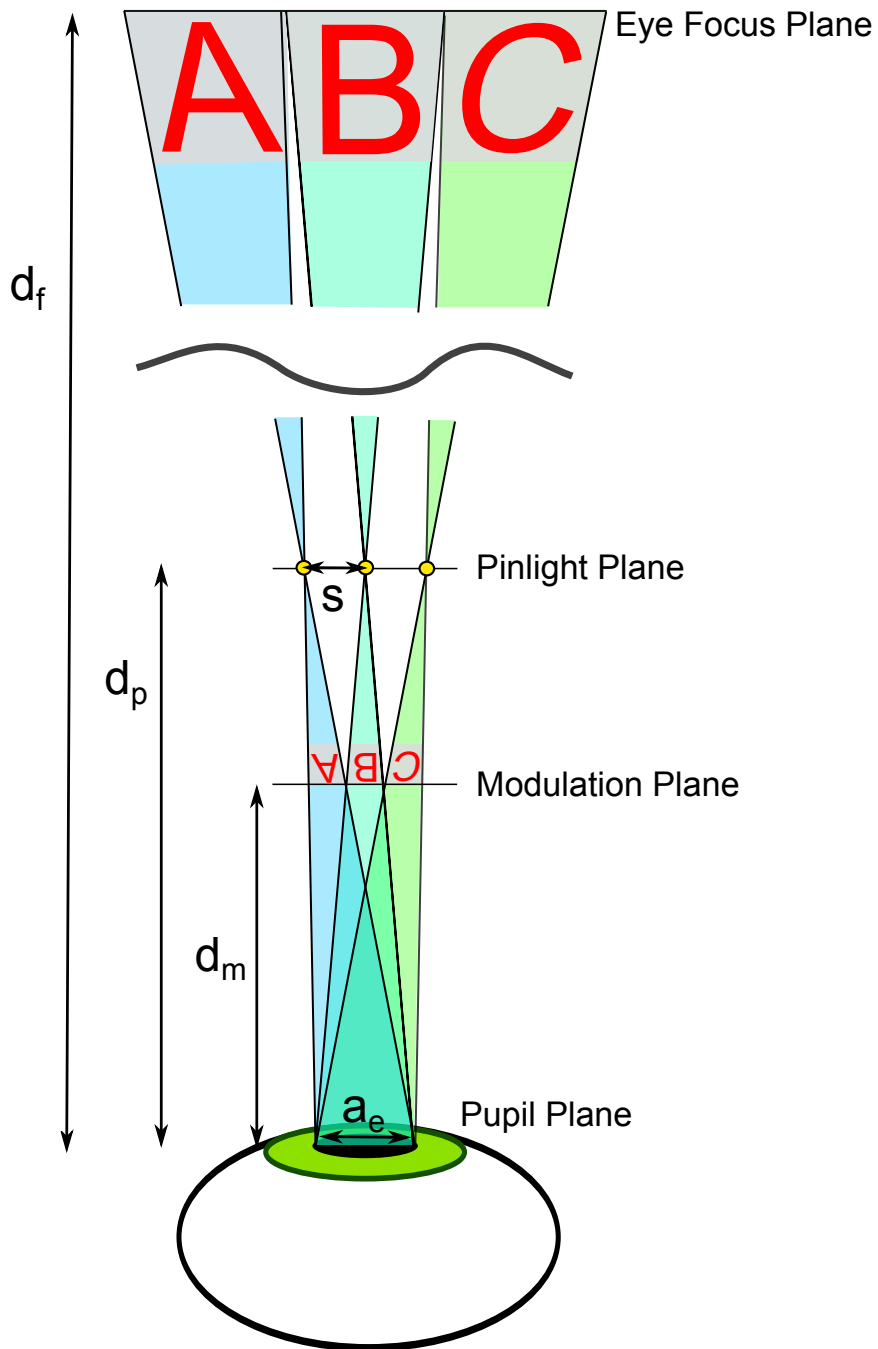
Figure 4.3: Ideal tiled pinlight projector geometry. The display is configured so that the pinlight projectors abut on the eye focus plane and on the modulation plane, creating a continuous image and using the SLM efficiently. (Note that the pinlights emit light over a wide angle, but only light entering the pupil is drawn).

$$s = a_e \left(1 - \frac{d_p}{d_f} \right) \quad \text{and} \quad d_m = \frac{a_e d_p}{a_e + s}. \tag{4.3}$$

This spacing ensures that the projected cone of light that enters the eye from neighboring pinlight projectors abut at the modulation plane and focus plane, providing a continuous image at the full resolution of the SLM. Note that this geometric configuration places pinlights sparsely (with spacing on the order of the eye pupil size), so that the structure of an *array* of small pinlights will remain imperceptible when defocused. Also note that the display becomes thinner as it is moved nearer the eye (i.e., decreasing d_p decreases $d_p - d_m$).

The effective resolution r in terms of the preserved fraction of modulation plane resolution (by area) can be computed as:

$$r = \left(\frac{d_f (d_p - d_m)}{d_m (d_f - d_p)} \right)^2 \tag{4.4}$$

This equation compares the size of a pixel projected onto the focus plane to the size of the focal plane, by area. Note that this equation assumes that the pinlight geometry is valid: the eye focus plane must contain a continuous tiled image, and the modulated areas on the SLM must be disjoint among projectors. In the ideal configuration, ratio r equals 1: the entire modulation plane has been used to form an image on the focus plane. The horizontal field of view f_h from a point on the eye can also be computed as:

$$f_h = 2 \tan^{-1} \left(\frac{c}{2d_m} \right) \tag{4.5}$$

where c is the width of the modulation plane.

4.5.4.2 Challenges for Practical Tiling

Tiling has the potential to create a wide field of view display. However, in the *ideal* one-dimensional case described above several factors have not been considered which must be addressed to create a practical *human-wearable* display:

1. The ideal case can only be directly extended to 2D if the image areas created by the pinlight projectors can be tiled; however, a circular image area is created when projecting into the round human pupil, which cannot be tiled without the inclusion of gaps or overlapping areas.

2. The ideal model assumes that the eye position is fixed relative to the display, an invalid assumption for real-world use. If the eye moves the image will circularly shift *within* the tiled sub-images corresponding to each pinlight projector, resulting in a corrupted image.

3. A device using the ideal model is only modestly thin. For example, a device designed for eye aperture $a_e = 3$ mm focused at $d_f = \infty$ with the pinlight plane placed at $d_p = 30$ mm yields a device $d_m = 15$ mm away from the eye that is $d_p - d_m = 15$ mm thick. While still thin for a wide field of view display, this is significantly thicker than ordinary eyeglasses, although note that most of the thickness consists of empty space so that the device may remain lightweight.

4. The ideal model is affected by changes in pupil size and focal state, which are expected to change over time.

In the remainder of Section 4.5.4, items $1 - 3$ are addressed under two alternative configurations. Item 4 is addressed in Section 4.5.5.5.

4.5.4.3 Tracked Virtual Aperture Configuration

Eye Tracking In one alternative tiled configuration, eye movement relative to the display is allowed under the assumption that the position of the eye on the pupil plane[4] is known (i.e. tracked) and that the image on the modulation plane is adjusted to account for the new eye position. As the eye moves, the view cones corresponding to each pinlight projector (illustrated in Figure 4.3) shift and intersect with a new portion of the intended image on the eye focus plane, which is then flipped and scaled appropriately to be encoded on the corresponding region of the modulation plane (see Section 4.5.5.4 for details). In this section, theoretical error-free tracking without latency in sensing or display is assumed. Physical eyetracking is discussed later in Section 4.7.4.

Virtual Eye Apertures Although eye tracking allows compensation for eye movements, it does not resolve the issue of how to tile the circular images formed by individual pinlight projectors due to the round aperture of the eye. In particular, if the circular projectors are tiled without overlap, gaps will remain in the image, and if the circles overlap to fill the plane, the overlapping areas will have greater intensity than the non-overlapping

---

[4]In practice, the pupil generally does not move on a plane but over a curved surface that is related to the center of rotation of the eye. However, a plane was found to be suitable approximation when testing the display with human viewers as described in Section 4.7.3.
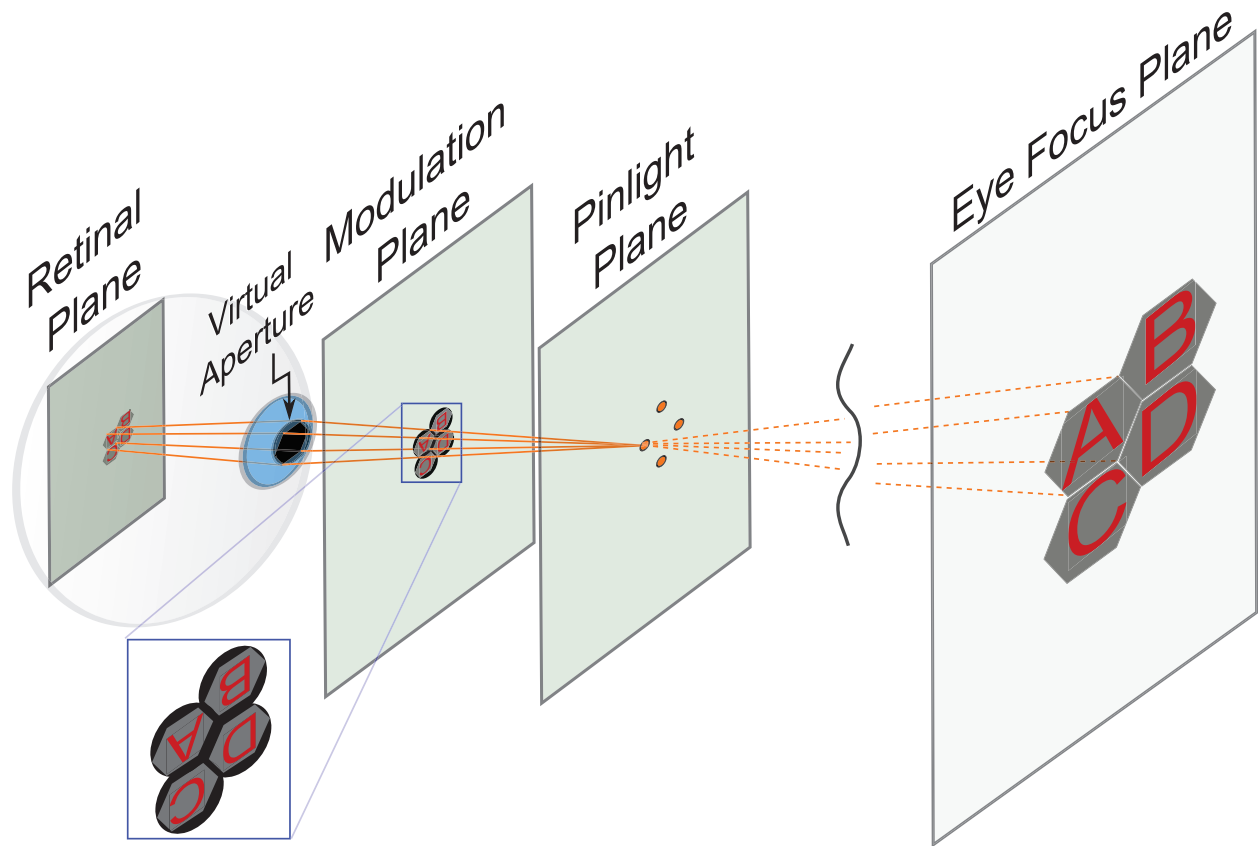
Figure 4.4: Tiled pinlights in tracked virtual aperture configuration. An aperture mask is encoded over the desired image on the modulation plane to create a virtual hexagonal aperture over the eye that can be tiled, eliminating uneven brightness in the perceived image. The black areas on the image represent the aperture mask and result in a small amount of lost display resolution. The image on the modulation plane is recomputed based on the tracked eye position to allow eye movement relative to the display. (Note that the pinlights are intended to emit light over a wide angle, but only light that enters the pupil is drawn).
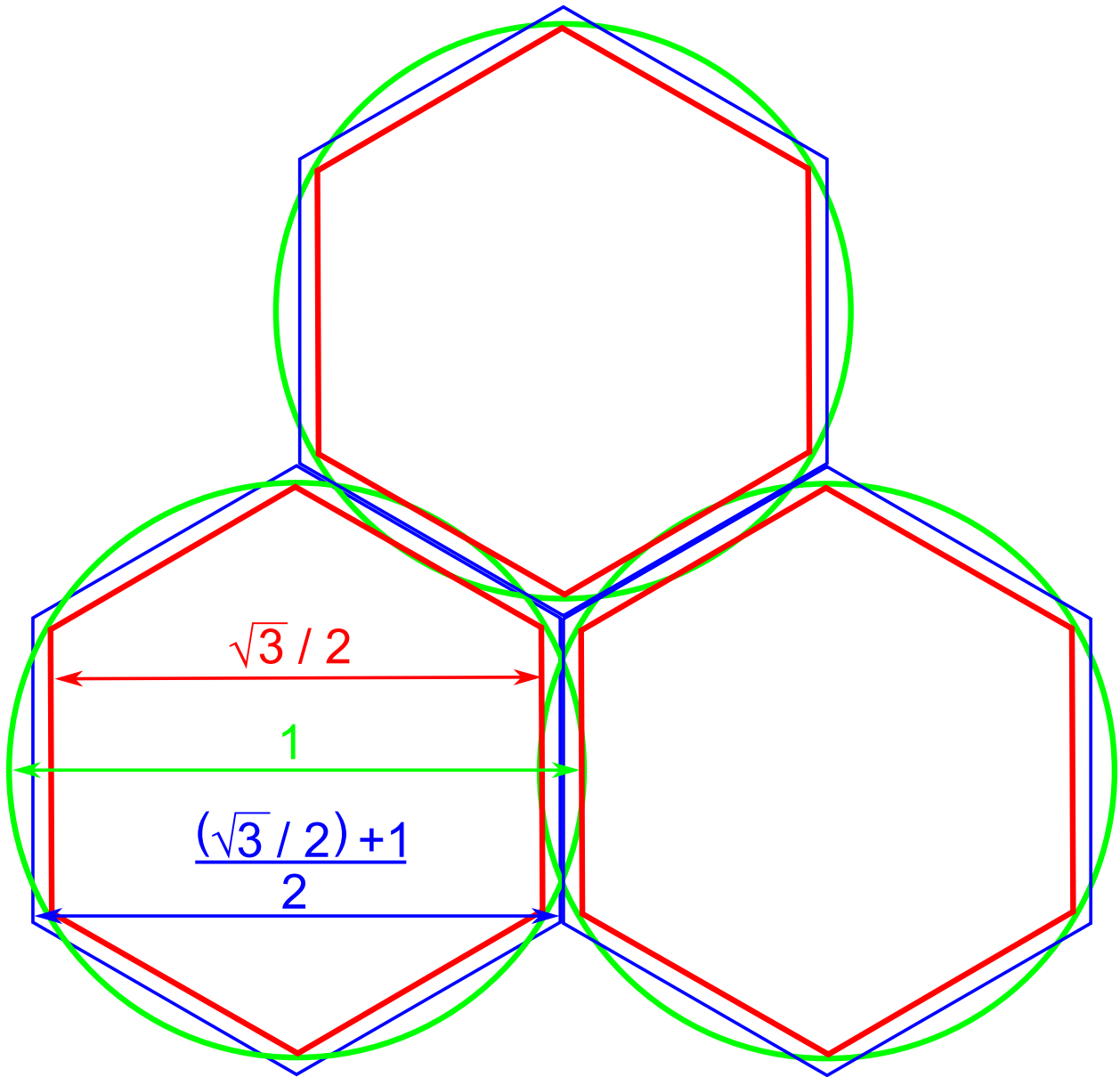
Figure 4.5: Pinlight hexagonal tiling geometry for the tracked configuration on the SLM. The green circles show the projected location of the minimally overlapping pupil circles on the SLM. The red hexagons show the utilized display area of the SLM; areas outside these regions are set to black to eliminate toning in the perceived image from the overlapping pupil circles, resulting in a small resolution loss. The larger blue hexagons exactly tile the viewing plane and are used as a reference to compute the effective utilized area on the SLM of $\approx 86\%$. The dimensions show the relative scale of the shapes in relation to a unit diameter pupil circle.

areas (see Figure 4.8E). A solution to create an evenly-toned image is to configure the pinlight projectors so that they *minimally* overlap to fill the focus plane (i.e. maximizing resolution) and to encode a *virtual aperture* over the modulation plane so that light from the overlapping regions in the modulation plane does not reach the eye. This process is illustrated in Figure 4.4. The virtual aperture has the same effect as if the viewer's pupil were a shape that could be tiled (e.g. a rectangle or hexagon), or the user were wearing a contact lens masked with such a shape.

Geometry The ideal display geometry, given by Equation 4.3, is updated to support a *hexagonal* virtual eye aperture as follows.

$$s_{t_h} = \frac{\sqrt{3}}{2} a_e \left(1 - \frac{d_p}{d_f} \right), \; s_{t_v} = \frac{\sqrt{3}}{2} s_{t_h} \tag{4.6}$$

$$d_{m_t} = \frac{\frac{1}{2} \left(a_e + \frac{\sqrt{3}}{2} a_e \right) d_p}{\frac{1}{2} \left(a_e + \frac{\sqrt{3}}{2} a_e \right) + s_{t_h}} \tag{4.7}$$

This geometry assumes that the virtual aperture is encoded as a hexagon with vertical left and right sides and a pointed top, as seen in Figure 4.21B. Note that the horizontal pinlight spacing s_{t_h} and vertical spacing s_{t_v} are asymmetric and that odd pinlight projector rows should be offset by $\frac{s_{t_h}}{2}$ due to the staggered hexagonal packing of the plane. Also note the similarity to the geometry in the ideal case (Equation 4.3), except that pinlight spacing s has been decreased to allow the circular image areas of the pinlight projectors to overlap to cover the focus plane, and modulation plane distance d_m has been adjusted to allow space for the virtual aperture mask. In particular, a regular hexagon is inscribed into the circular area on the modulation plane that represents each pinlight projector, and d_{m_t} is set so that the projectors are positioned as closely as possible without intersecting the inscribed hexagons of neighboring projectors. This process causes a resolution loss as some of the modulation plane is now dedicated to providing an aperture mask rather than contributing to the virtual image. The resolution loss can be computed according to Equation 4.4. Equivalently, as shown in Figure 4.5, the effective resolution (by area) in a tracked virtual aperture configuration can also be computed as the ratio of the area of the inscribed hexagon in a unit diameter circle $\left(\frac{3\sqrt{3}}{8} \right)$ with the area of the larger hexagon that would exactly tile the plane without the virtual aperture mask $\left(\frac{7\sqrt{3}+12}{32} \right)$, yielding effective resolution ratio r_t:

$$r_t = \frac{12\sqrt{3}}{12 + 7\sqrt{3}} \approx 86\%. \tag{4.8}$$
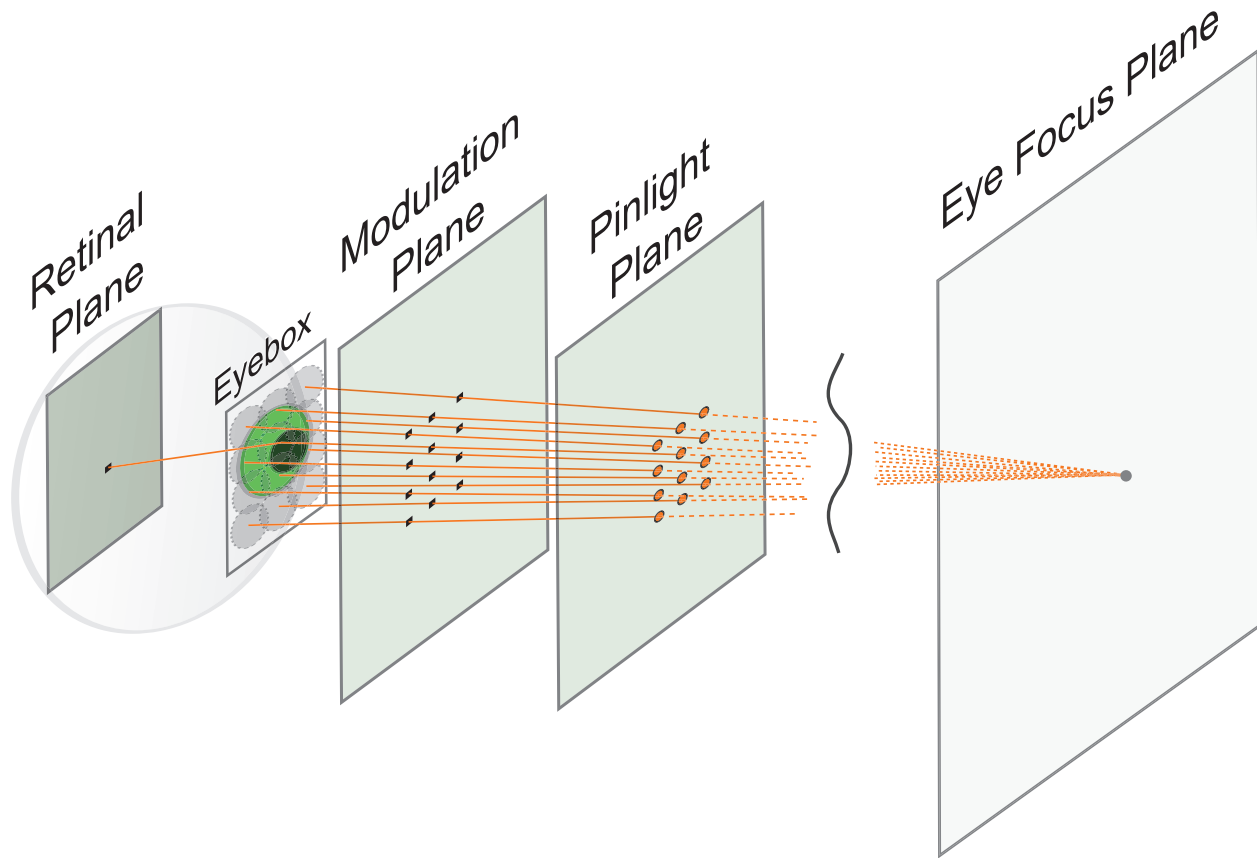
Figure 4.6: Untracked light field configuration. For each point on the eye focus plane, a set of rays are created through a tiled set of pinlight projectors that are distributed about an eyebox near the eye, allowing eye movement. The modulated pixels corresponding to the same point in the scene are distributed over the image.

Note that spatial resolution loss is very modest compared to existing near-eye light field displays, showing the benefit of a design that allocates nearly of all the display area to forming the image perceived by the user, through tracking. Also note that moving the modulation plane closer to the pinlight plane to accommodate the virtual aperture has the positive side effect of creating a slightly thinner display; e.g., an aperture $a_e = 3$ mm focused at $d_f = \infty$ with the pinlight plane placed at $d_p = 29$ mm now yields a device $d_{m_t} = 15$ mm away from the eye that is $d_p - d_{m_t} = 14$ mm thick (i.e. 1 mm thinner than in the ideal configuration).

4.5.4.4 Untracked Light Field Configuration

Near-Eye Light Fields In another alternative tiled projector configuration, the display is designed to generate a *light field* near the eye to allow additional capabilities. The display is designed so that the view cones between pinlight projectors *overlap* away from the display, allowing angular variation in the image among
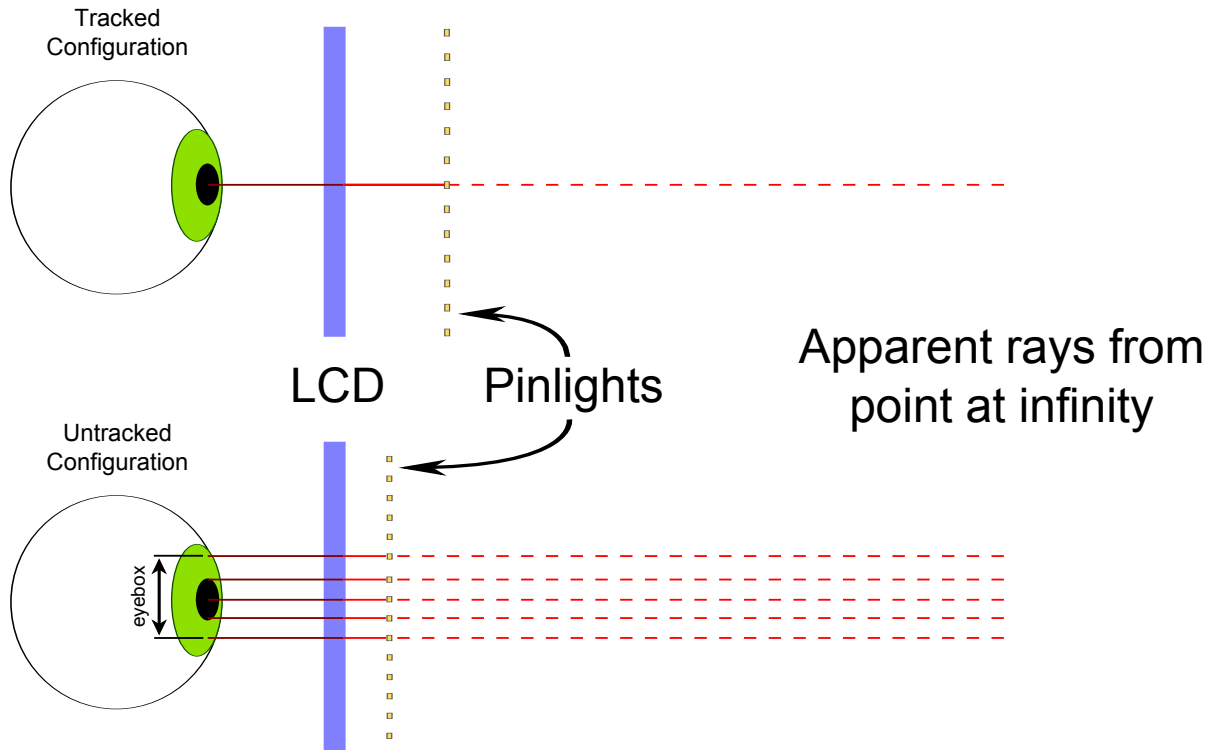
Figure 4.7: Pinlights tracked and untracked configurations comparison. *Top:* Tracked configuration. For each pixel in the augmented image (here at focal infinity), the display emits a single "ray" that is modulated in intensity according to the tracked viewing position, maximizing resolution. *Bottom:* Untracked light field configuration. For each pixel in the augmented image (here at focal infinity), the display emits multiple rays spanning an eyebox; rays are spaced closely enough so that at least one ray always enters the viewer's pupil. The geometry of the untracked configuration also results in a thinner display.

projectors. In particular, one can design the display to allow angular variation *around* the eye to create an *untracked* configuration, which is explored in this section. It is also theoretically possible to design the display to emit a light field with angular variation *over* the pupil (creating depth of field and accommodation cues), but due to the very high display panel resolution requirements and diffraction limitations, this configuration was not investigated in this work.

Untracked Configuration The tracked display configuration offers high spatial resolution, but the need for pupil tracking adds engineering complexity. An alternative display can be defined with an *eyebox*, a region over which the eye can move around while still seeing the complete intended image. Here, the display is configured to emit *multiple* light rays that appear to originate from each point on the eye focus plane, each of which is directed towards different regions of the eyebox, as illustrated in Figure 4.6. To maximize resolution,

the display geometry minimizes the number of modulated rays emitted from each point on the eye focus plane with assurance that an eye placed anywhere in the eyebox will receive one such ray; Equations 4.6 and 4.7 take the following form in this configuration:

$$s_{u_h} = \frac{\sqrt{3}}{2} a_{e_m} \left(1 - \frac{d_p}{d_f} \right), \; s_{u_v} = \frac{\sqrt{3}}{2} s_{u_h}, \; d_{m_u} = \frac{a_{e_b} d_p}{a_{e_b} + s_{u_h}} \tag{4.9}$$

Horizontal pinlight spacing s_{u_h} and vertical spacing s_{u_v} are defined with respect to a minimum eye aperture a_{e_m}; a viewer with a smaller aperture will perceive gaps in the image. Modulation plane distance d_{m_u} is defined with respect to a constrained window around the eye a_{e_b}, creating an eyebox of size $a_{e_b} - a_e$. Unlike a conventional display, each ray in the eyebox can be modulated individually in this light field mode, allowing different images to be perceived with eye movement and without tracking. For example, the display could be configured to show different images (e.g. a menu) when the user glances left or right. The display is also considerably thinner in this configuration; an eye with minimum aperture $a_{e_m} = 3$ mm and eyebox of $a_{e_b} - a_e = 7$ mm (when $a_e = a_{e_m}$) focused at $d_f = \infty$ with the pinlight plane at $d_p = 18.9$ mm yields a device $d_{m_u} = 15$ mm from the eye that is $d_p - d_{m_u} = 3.9$ mm thick. Figure 4.7 provides a comparison of the geometry of the untracked configuration to the tracked configuration.

Limitations The untracked display configuration described above offers two notable challenges, however. First, the displayed image contains unevenly brightness levels (toning) since the number of pinlight projectors that cover a region in the image vary over the display. This is a consequence of distributing rays in the eyebox with assurance that the *round* human pupil receives light from every point in the scene; however, the inability to exactly tile the eyebox plane with circles causes some overlapping regions. This effect cannot be directly corrected in a static display without knowledge of the eye position; however, approaches for mitigating the effect are discussed in Section 4.7.4. Another limitation of the untracked configuration is a significant loss in spatial resolution. In the example above, the preserved SLM resolution computed using Equation 4.4 is approximately 6%, which is comparable to other near eye light field displays (Pamplona et al., 2010, 2011; Lanman and Luebke, 2013). In future work, resolution loss may be minimized, however, by creating a hybrid of the tracked and untracked configurations: a display with a small eyebox that compensates for tracker error and latency.
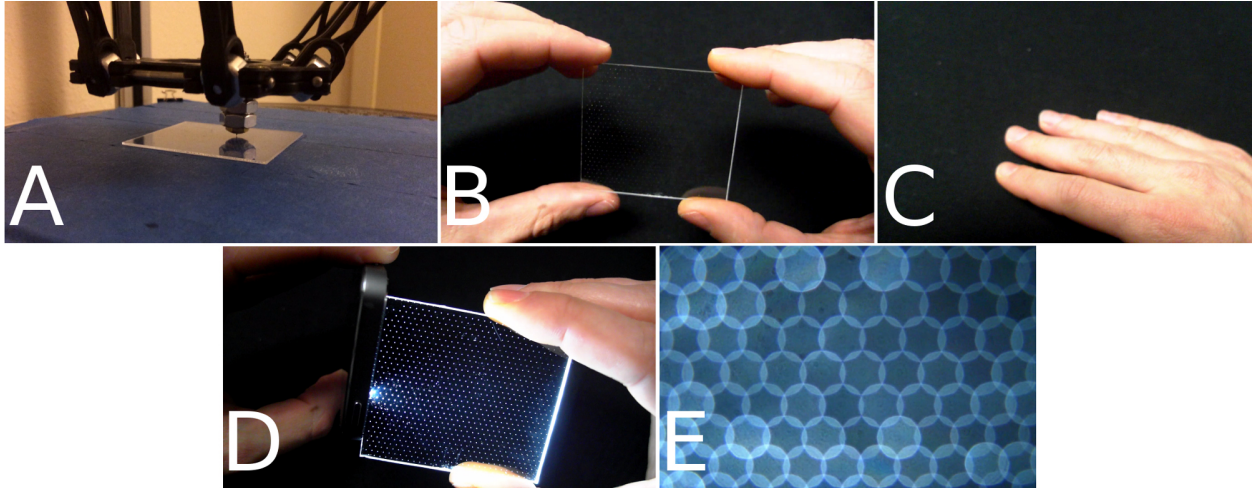
Figure 4.8: Waveguide-based pinlight array. A) Array manufactured by creating small divots in a common acrylic sheet using a needle attached to the robotic arm of a 3D printer. B) The waveguide contains a sparse array of visible divots. C) When viewed near a camera with an aperture and focus similar to a human eye, the divots are imperceptible and see-through. D) When illuminated from the side with an LED, bright spots appear in the divots of the plastic sheet. E) When defocused near the camera, the pinlights form discs that tile the image plane. Notice the brighter overlapping areas that cause resolution loss and also toning artifacts in the untracked configuration.

4.5.5   Practical Implementation Details

In this section, issues related to realizing a practical human wearable display are discussed.

4.5.5.1   Creating Point Light Sources

Requirements To create point light sources for pinlight projectors, note three primary requirements. First, the pinlight sources should be bright, i.e. the total light emitted should be on par with a normal display panel, but concentrated into the much smaller area of a sparse grid of dots. Second, the emission area of each pinlight should be very small (i.e. on the order of the size of a pixel on the SLM) as the image formed is essentially convolved with the pinlight shape. To maximize resolution, the effective aperture size of the SLM pixels (with consideration for the pixel fill factor) plus the size of the pinlight should be less than or equal to the pixel pitch of the SLM. Finally, the pinlight array should be highly transparent and any structures comprising the pinlight array should be small enough to be imperceptible when defocused.

Implementation The proposed pinlight array implementation is a common acrylic or polycarbonate sheet (acting as a waveguide) that is etched with tiny divots and edge illuminated with LEDs, as illustrated in
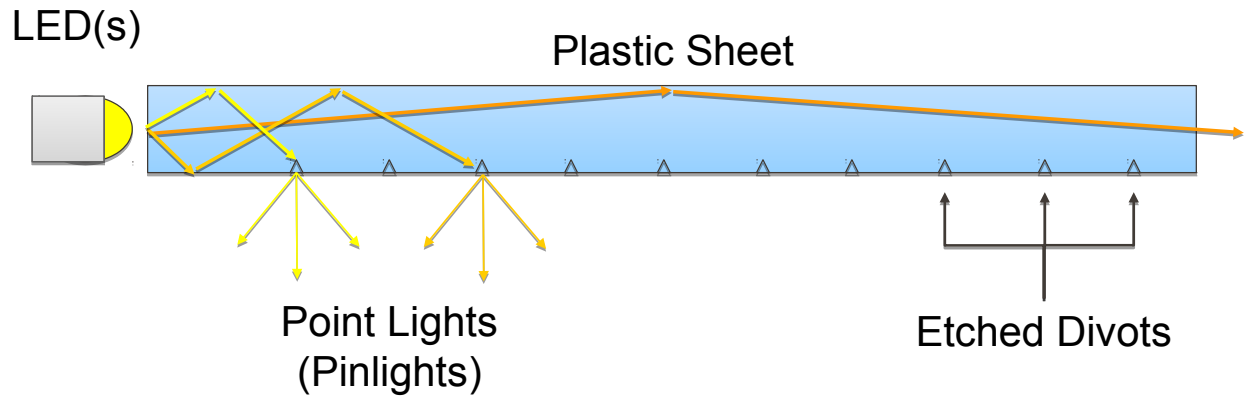
Figure 4.9: Waveguide-based pinlight array. Light enters the side of the plastic sheet and travels through the sheet by total internal reflection to the other end, except where it encounters the etched features and scatters.

Figure 4.8. Light from the LEDs is channeled through the waveguide by total internal reflection (TIR) to the opposite side except where it encounters the divots, which scatter the light and cause it to exit the waveguide, forming small bright spots (see Figure 4.9). Note that a similar configuration is commonly found in the backlights of LCD panels, but here the design is optimized for transparency and small sparse emission sites, rather than the creation of a uniform area light source.

Two methods were explored to create the divots: mechanical etching and laser drilling. The mechanical etching was performed using a needle attached to the moving robotic platform of a hobbyist 3D printer, which was programmed to create a staggered pinlight array according to Equation 4.6. This method produced cone shaped divots with a maximum diameter of ≈ 100 μm, potentially limiting resolution. Laser drilling was performed by a commercial service (Lennox Laser), which produced precisely drilled divots of a 35 μm diameter and a depth of 50 μm or 150 μm. Note that deeper divots will collect more light and appear brighter, but will also have a larger apparent size off-axis, potentially lowering resolution. Since the divots near the edge of the waveguide will appear brighter (due to adjacency to the light sources), but will also have a larger apparent footprint, it is recommended that depth is maximized at the center of the waveguide and minimized at the edges. The etched features were visible to the naked eye with both fabrication methods, but were imperceptible defocused when held near the eye. Note that this implementation is light inefficient since much of the light escapes through the waveguide edges; however a bright array can be achieved with a modest number of LEDs (six per eye in the prototype displays). It is possible that more precise structures could be etched into the plastic that collect more light and direct more of the light output toward the eyebox.

Note other possibilities for creating an array of transparent light sources: transparent emissive displays (e.g. transparent OLEDs), LED or laser chips mounted on a transparent conductive substrate (e.g. ITO coated glass), fiber optics, holograms, and quantum dots.

4.5.5.2 Modulating Light Sources

A spatial light modulator intended for a pinlight projector display must work in a transmissive mode and should have high transmittance, a wide viewing angle, and should minimally diffract light. An LCD microdisplay intended for use in a projector was selected for its relatively high transmissivity and pixel density. To minimize the diffraction and light loss through the display caused by color filters, a monochrome panel was selected and operated in conjunction with a color sequential pinlight array. Further implementation details are provided in Section 4.6.1, and additional factors (e.g. contrast) are discussed in Section 4.7.4.

4.5.5.3 Optimizing See-Through Capability

To achieve a see-through capability, it is assumed that the SLM and pinlight array are effectively transparent when defocused near the eye. Note, however, two complicating factors. First, light from the environment may permeate the SLM in addition to illumination from the pinlights, causing a soft defocused glow around the synthetic image (see Figure 4.21D). Second, light from the environment only reaches the eye through the defocused mask on the SLM, causing a soft, uneven coloring of the environment and allowing little light to reach the eye in areas where there are no synthetic objects (see Figure 4.21C). To mitigate these issues, one may rapidly alternate between displaying an augmented image with the pinlights on (see Figure 4.21D), and displaying a occlusion mask of the augmented image with the pinlights off (which appears defocused, see Figure 4.22E). This allows light from the environment to reach parts of the display where no augmented imagery is shown and reduces the apparent soft glow around the augmented images (see Figure 4.21E). It should be noted, however, that an acceptable result may be achieved even if this technique is not used if the augmented image fills much of the user's field of view.

4.5.5.4 Creating Modulation Masks

To create the modulation masks to display on the SLM, the virtual scene is projected through the pinlights onto the modulation layer with respect to the eye as illustrated in Figure 4.10. This process was implemented in software by rendering the virtual scene with a camera that uses an off-axis projection frustum defined
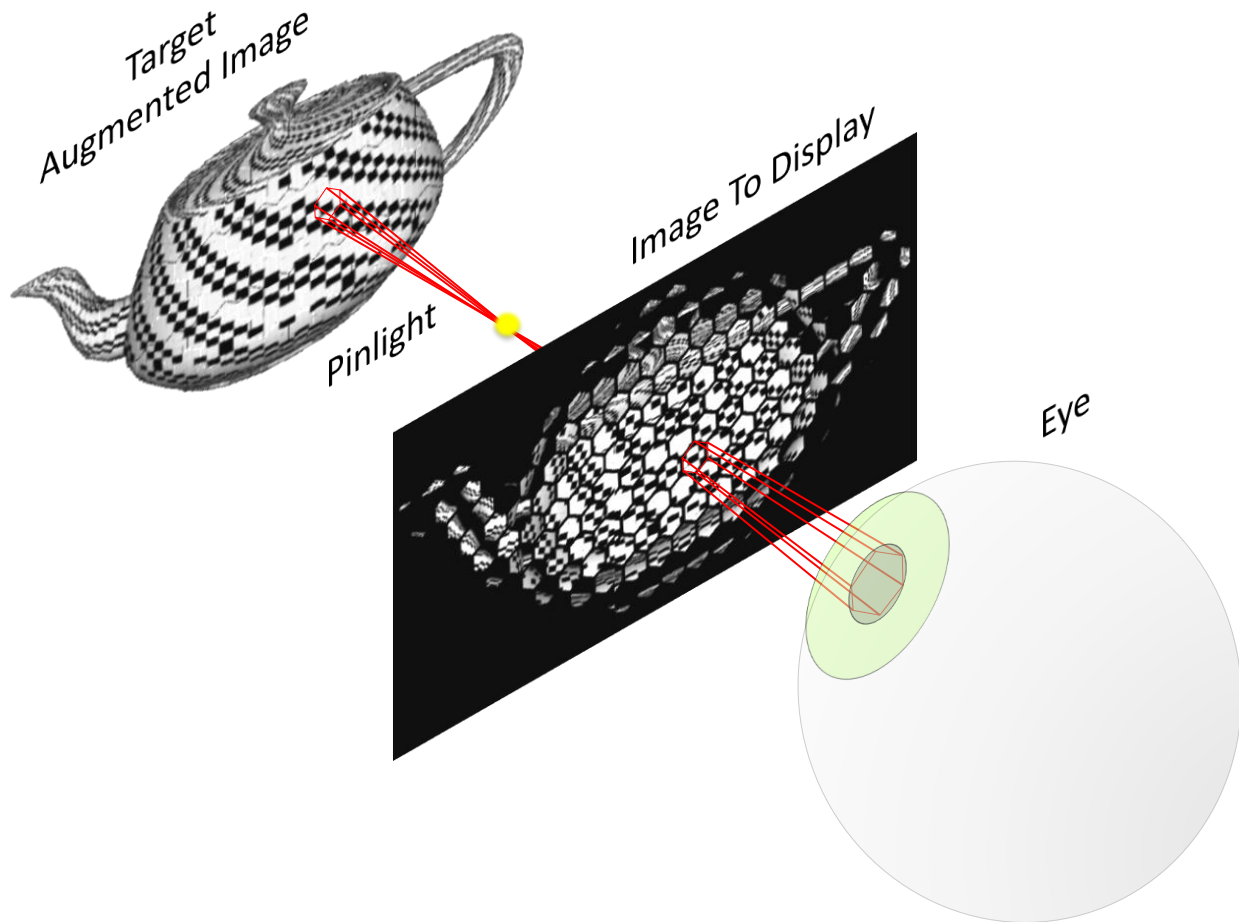
Figure 4.10: Pinlight Modulation Masks. To generate the imagery to display on the SLM, the scene is first rendered and then projected onto the modulation layer through each pinlight with respect to the eye.

by the eye and each pinlight; this process is repeated for each pinlight. If the scene is a simple 2D plane in each eye at the eye focus distance (rather than an arbitrary light field), this process is performed more efficiently by rendering the scene once and finding the intersection of the camera frustum and the focus plane and transferring a flipped and appropriately scaled copy of the rendered image region onto the correct region of the modulation plane. For images representing occlusion masks, an unmodified copy of the image on the modulation plane is simply drawn, which appears defocused when viewed. Example modulation and occlusion masks are shown in Figure 4.22.

4.5.5.5 Changes in Eye State

Although accommodating the display for eye movement was discussed in Section 4.5.4, the eye may change in other ways that may affect the display; notably the pupil size or focal state may change.

Handling Change in Pupil Size In the tracked configuration (Section 4.5.4.3), it was assumed that the eye aperture size was fixed in Equation 4.6. One can account for a dynamic aperture size using one of four methods. First, the display geometry can be configured for the maximum allowable aperture size and a *virtual aperture* (see Section 4.5.4.3) can be created with the minimum allowable size. This allows only a smaller central area of the pupil to receive light as the pupil size changes, but will cause a resolution loss proportional to the ratio of the largest and smallest apertures. Second, the display could be configured with a small eyebox sufficient to allow the pupil size to expand beyond a minimum size. This approach also results in a similar loss in resolution but the addition of a small eyebox also helps compensate for tracker error and latency. Third, the display could be outfitted with a *dynamic* pinlight array (e.g. a transparent OLED display) that can adjust the spacing of the pinlights according to tracked pupil sizes. Finally, the amount of light reaching the eyes could be controlled in an active loop by the SLM and/or pinlight array in an attempt to control pupil size to a predetermined size. The display could also simply be set sufficiently bright to force the pupils to contract to near their minimum size. However, note that other factors than lighting may affect pupil size. In the untracked configuration (Section 4.5.4.4), variable aperture sizes are already considered in Equation 4.9.

Eye Focus In a tiled pinlight projector display, changes from the expected eye focus do not cause the image to become appreciably "blurred", but rather change the scaling of the tiled sub-images among the various projectors so that they expand or contract, causing small tile gaps or overlaps to occur in some cases. This is expected to appear less natural to the user than normal focal blur. However, the change in scaling is small unless the viewer is focused at very close range; e.g. the change in scaling from a distance of 1 m to infinity is $\leq 3\%$ in a typical display (see Equation 4.6). Gaps caused by nearer than expected focus can be avoided by configuring the pinlight projectors so that they slightly overlap.

4.5.5.6 Diffraction

Light passing through an aperture diffracts (expands angularly) with a magnitude inversely related to the aperture size. This effect is troublesome in the proposed design as the viewer sees the augmented image and real environment through a series of small apertures on the SLM, degrading both views.

One can approximate the diffraction-limited resolution of a pinlight display by assuming that the image is viewed at infinity (i.e. Fraunhofer diffraction) and that two adjacent points are no longer resolvable when

the diffraction maximum of one point coincides with the first minimum of the other (i.e. Rayleigh criterion). The angular limits of the display are then computed as:

$$\theta_d = \frac{\lambda}{w}, \;\; \theta_g = \frac{w}{2(d_p - d_m)} \tag{4.10}$$

where θ_d is the angle of the first diffraction minimum, λ is the wavelength of the light, w is the SLM aperture size, θ_g is half the angular extent of a pixel defined by the geometry of the display, and $d_p - d_m$ is the distance between the pinlight plane and modulation plane. The maximum angular resolution is then calculated by equating these angles:

$$w_{opt} = \sqrt{2(d_p - d_m)\lambda}, \;\; r = \frac{1}{\left(\frac{180}{\pi}\right)\left(\frac{\lambda}{w_{opt}}\right)} \tag{4.11}$$

where w_{opt} is the optimal aperture size (i.e. when θ_d and θ_g are equal) and r is the corresponding angular resolution in pixels/degree.

Diffraction through the display is illustrated in Figure 4.11 where modulation layer pitch is selected according to Equation 4.11. Light originates from the environment or a pinlight and is diffracted by the pixel apertures at the modulation plane, where the first diffraction orders form at the angle θ_d. When the light reaches the eye's lens (of focal length f), it has spread so the size between the zeroth and first orders (i.e. half the spot size on the lens) reaches:

$$x_1 = d_m \tan \theta_d. \tag{4.12}$$

The light is then refracted by the lens of the eye, assumed to be a thin lens focused at infinity. By the thin lens formula, the diffracted orders will converge at distance of x_2 beyond the lens, conjugate to the modulation plane, where:

$$x_2 = \frac{1}{\frac{1}{f} - \frac{1}{d_m}}. \tag{4.13}$$

Since the eye is focused beyond the modulation plane (at infinity), the diffracted orders will not be brought into focus, but instead will converge at a distance of $x_2 - f$ beyond the retina. The spot size of the diffracted light (up to the first order) can then be calculated as:
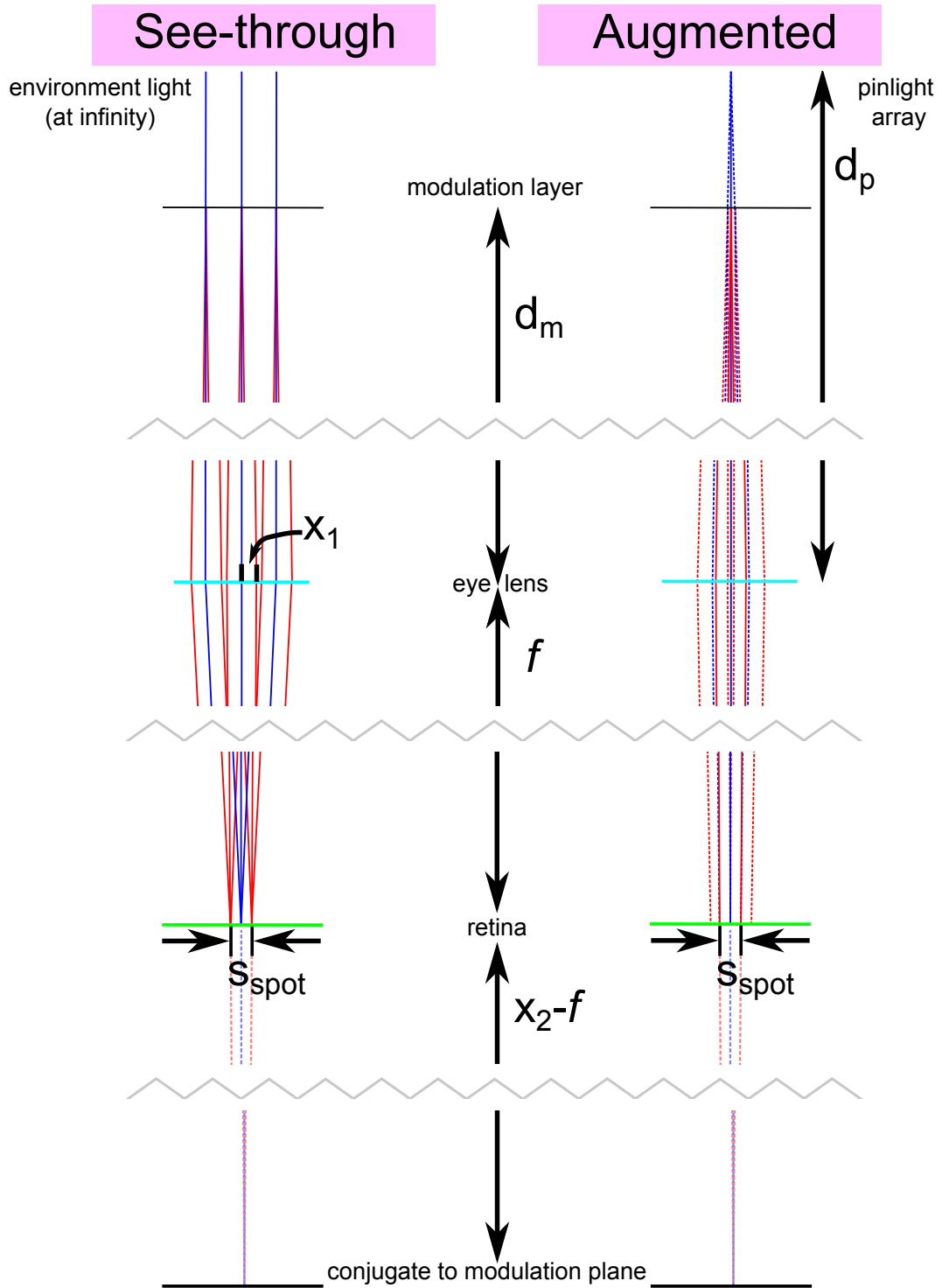
Figure 4.11: Pinlight geometric model of diffraction. Light from the environment or a pinlight travels through the modulation layer and diffracts and is refracted by the lens of the eye to form an image on the retina. Blue lines show zeroth diffraction order and red lines show first orders. *Left:* Light from environment (at infinity) propagating through the display into the eye. *Right:* Light from pinlight (forming augmented image) propagating to eye. Dotted lines before retina show light traveling from adjacent pixels.
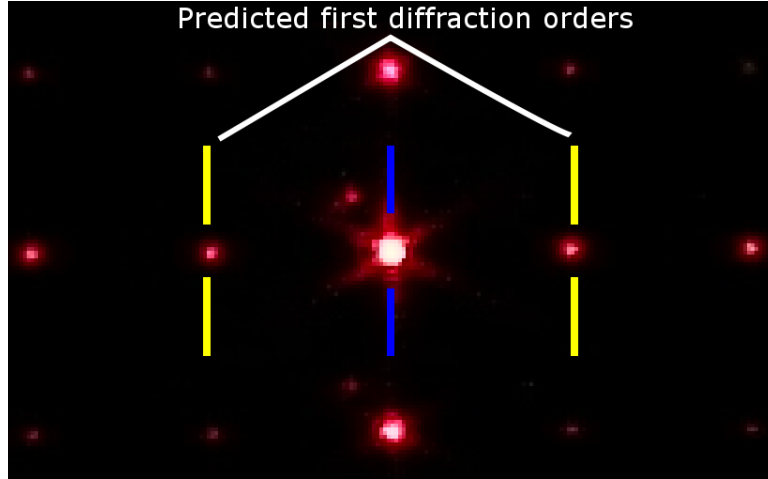
Figure 4.12: Pinlight see-through diffraction orders. Photograph taken of distant laser light source through a LCD panel with all pixels set to most transparent state and the camera focused at the laser source. The zeroth order is indicated with a blue line. The predicted horizontal location of the first orders are indicated with yellow lines.

$$s_{spot} = \frac{2x_1(x_2 - f)}{x_2} = 2f \tan \theta_d. \tag{4.14}$$

This geometric construction was verified experimentally in Figure 4.12. The image shows a photograph of a distant laser point source taken through an LCD panel with the camera focused on the laser source. For laser light with wavelength $\lambda = 635$ nm, LCD pixel pitch $w = 0.036$ mm, modulation plane distance $d_m = 25$ mm, and focal length $f = 5.2$ mm, the distance to the first diffraction order (i.e. $\frac{1}{2} s_{spot}$) was calculated as 0.0917 mm by Equation 4.14. The pixel pitch of the camera sensor is 0.00186 mm so that the first diffraction orders should be ≈ 49 camera pixels from the zeroth order. The predicted horizontal position of the first orders (relative to the manually identified zeroth order, indicated in blue) are illustrated with yellow bars in Figure 4.12 and match the apparent positions of the first orders within approximately one camera pixel.

Note although spot size s_{spot} appears the same for both the see-through (left of Figure 4.11) and augmented views (right of Figure 4.11), there are important consequences of the distance of light sources from the modulation layer. In the see-through view, the rays from a point at infinity are parallel, so the corresponding diffraction orders among rays emerging from the point remain parallel. Therefore, when these orders are imaged onto the retina with a lens focused at infinity, each converges to a small spot. In other words, diffraction of the see-through view consists of multiple focused and "sharp" copies of the scene, each

superimposed and shifted by $\frac{1}{2}s_{spot}$. As shown in Figure 4.13, this allows the see-through view to maintain high resolution, but ghosting effects appear in high contrast areas. However, if a high fill factor LCD panel is used (bottom left of Figure 4.13), these see-through diffraction effects can be subtle for typical indoor scenery. However, bright, wide spectrum objects (e.g. a lamp) will still exhibit strong rainbowing effects (bottom right of Figure 4.13) because the amount of separation between orders is wavelength dependent. Note that see-through diffraction is not a unique problem to the proposed approach and must be mitigated in other see-through designs (e.g. waveguide designs using holographic out-couplings), providing motivation for industry to mitigate this problem in the future.

For the augmented view, the right Figure 4.11 shows that the light from the first diffraction orders is not focused to a sharp point on the retina, but is rather spread over the size s_{spot}: the consequence of the light source being out of focus with respect to the lens. As formulated in Equations 4.10 and 4.11, the first diffraction order for the central pixel (red solid line) coincides with the zeroth order for the two adjacent pixels (blue dashed lines) when the adjacent pixels are angularly spaced by $\theta_g(=\theta_d)$, yielding the estimated resolution limit r of Equation 4.11.

However, in practice, the resolution limit is more difficult to estimate in the augmented image case because the LCD panel does not act as a simple aperture or grating but rather displays a spatially varying image. The light passing though the grating also has some degree of coherence as it comes from a light source with moderately narrow spectral width (e.g. an LED with a 20 nm width) and from the small emission area of a pinlight; therefore, some interference effects are expected as well. The wavefront from a pinlight that has propagated through the LCD layer can be calculated by treating it as series of point sources (according to the Huygens-Fresnel principle) whose phase vary according to the distance to the pinlight source and amplitude varies according to the modulated image. This wavefront can then be propagated to the eye and through the lens onto the retina. Such simulations are useful for predicting performance limits of spatial light modulators of varying density and under ideal conditions (e.g. LCDs with perfect contrast).

Some example wavefront simulations used this method are shown in Figure 4.14 for a display matching the specifications of the second prototype display (see Section 4.7.3): a LCD with pixel pitch $w = 0.036$ mm and fill factor of 85% is placed at a distance $d_m = 25$ mm from the eye and illuminated with a pinlight with a wavelength of $\lambda = 470$ nm at a distance $d_p = 31$ mm. Note that the simulations make some simplifying assumptions: the light is a true monochromatic point light source, the LCD is a thin ideal attenuation mask, and the wavefront is approximated discretely at wavelength scale. However, by comparing the simulations
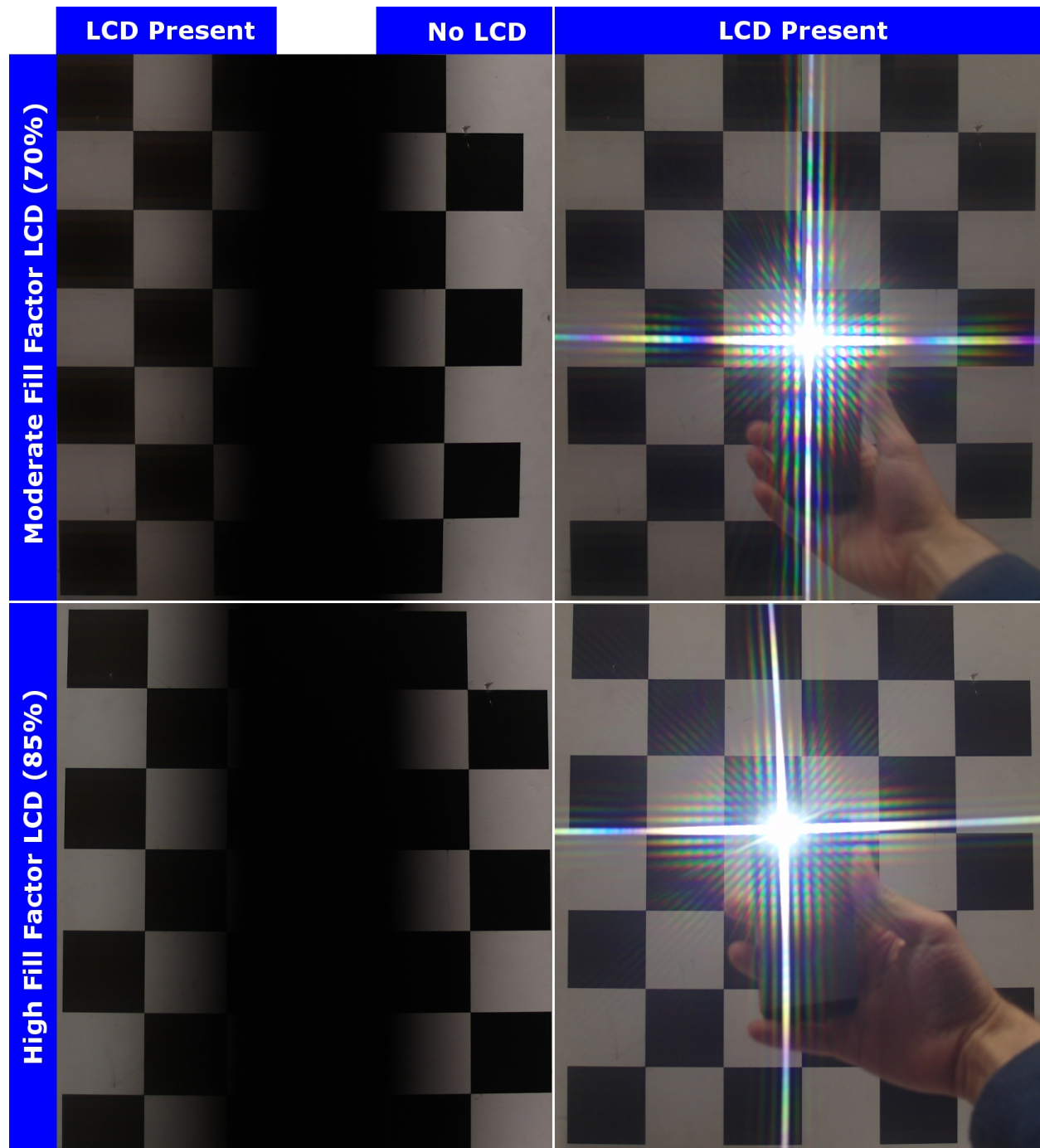
Figure 4.13: Pinlight see-through views. *Left column:* Photographs taken of scene with transparent LCD panel (with all pixels set to most transparent state) covering left half of image; right half of image is uncovered. *Right column:* Photographs of scene with bright light source with transparent LCD covering entire image. *Top row:* Photographs taken with moderate fill factor (70%) LCD panels. *Bottom row:* Photographs taken with high fill factor (85%) LCD panels. Both panels have a pixel pitch of $w = 0.036mm$ and are the same type of panels used in the prototype displays. Panels do not have polarizers installed, which will globally attenuate the LCD covered views.
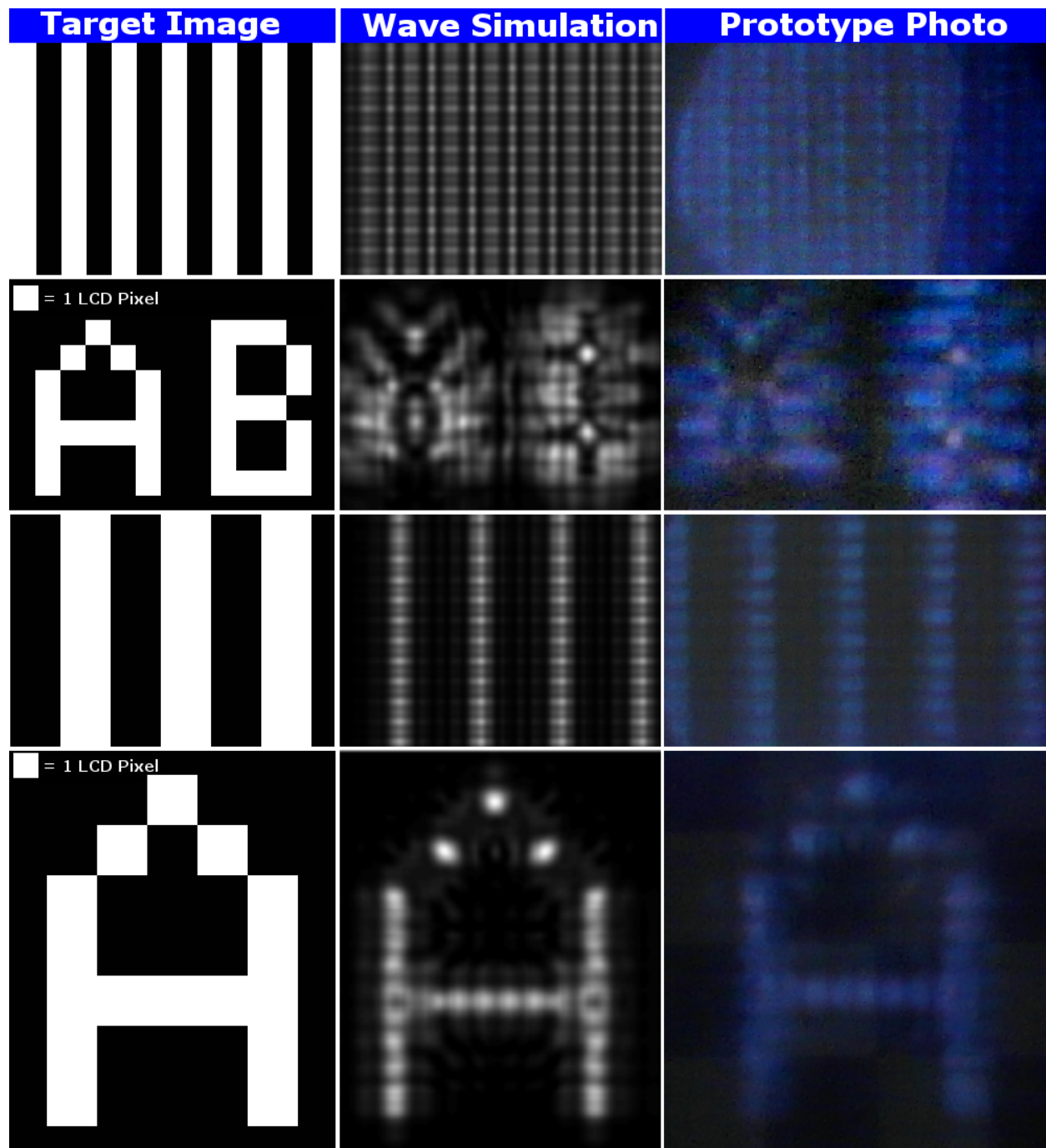
Figure 4.14: Pinlight simulated and prototype diffraction. *Left column:* Desired image to display on LCD panel. *Center column:* Wave optics simulation of image appearance using configuration of prototype display. *Right column:* Photograph of image on prototype display. *Row 1:* Images of one pixel width line pairs. *Row 2:* Image of one pixel line width letters. *Row 3:* Images of two pixel width line pairs. *Row 4:* Image of two pixel line width letters.
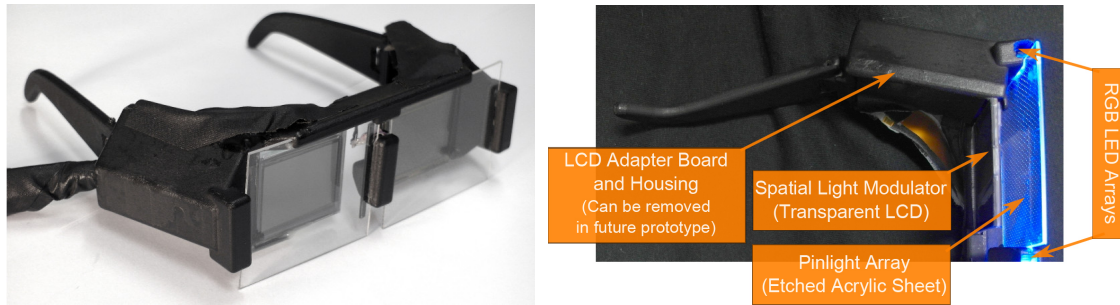
Figure 4.15: First prototype display. Left: Photograph of display. Right: Identification of display components.

(center column) to photographs of a prototype display (right column), one can see that the approximations are reasonable, and the performance of the prototype approaches the theoretical limits, albeit with reduced contrast. The bottom two rows of Figure 4.14 show that two pixel line pairs and text with a two pixel line width can readily be resolved with fair contrast. Note that the width of the two pixel lines (0.0072 mm) is very close to the ideal pitch predicted by Equation 4.11 (0.0075 mm), indicating that the Rayleigh criterion may be too conservative a measure of resolvability. As an alternative, the Sparrow resolution limit is reached when there is no drop in intensity between two points, yielding an upper resolution limit of approximately twice that of the Rayleigh criterion. In this case, features approximately the size of one pixel in the selected configuration should be just resolved. As shown in rows 1 and 2 of Figure 4.14, this appears to be the case: there is almost no contrast between one pixel line pairs, and the one pixel line width text is hardly recognizable.

In summary, among the display geometries discussed in this paper and the wavelengths of visible light, the display design is capable of angular resolutions in the range of $2-5$ pixels/degree by Equation 4.11, or up to approximately 10 pixels/degree using a more liberal definition of resolvability. These resolutions are very modest; however they are sufficient for a range of possible augmentations as shown in experimental results of Sections 4.7.2 and 4.7.3. The possibility of further increasing resolution is discussed in Section 4.7.4.

4.6 Implementation

4.6.1 Hardware

4.6.1.1 First Prototype

To test the design experimentally, a first prototype device was constructed that operated in the "tracked" configuration (see Section 4.5.4.3), but with the use of a camera in a known location rather than a human eye in order to evaluate image quality under ideal circumstances. The prototype device consisted of two main optical components: LCDs (Sony LCX017, 36.9 x 27.6 mm active area, 1024x768 pixels, 705 dpi, 60 Hz, monochrome) and an acrylic pinlight array that was constructed as described in Section 4.5.5.1 using the mechanical etching method. The acrylic pinlight array was 1 mm thick, etched with a 1.8 mm horizontal pinlight pitch, and illuminated with RGB LED color sequential illumination. The modulation plane and pinlight plane were spaced at a distance of $d_p - d_m = 13.5$ mm, creating an optical assembly with a total thickness of 15.5 mm including component thickness and 10.5 mm of empty space. The components were mounted in a 3D printed plastic eyeglasses frame designed in OpenSCAD and printed by Shapeways. See photos in Figure 4.15. The bulky components on the far left and right of the glasses house an LCD adapter board that can be relocated or removed in a future device.

The display was tested using a camera that approximated a human eye (PointGrey Flea2 camera body with a Fujinon F1.2, 2.3 mm aperture lens). The camera was mounted behind the display (see Figure 4.16) at a distance that approximated a human wearing eyeglasses (i.e. a camera center of projection to modulation plane distance of $d_m = 16$ mm). The field of view of the display through the camera is $\approx 110°$ diagonally in each eye (see Figure 4.17), limited by the camera's field of view and a cropping of the top of the camera image to remove portions of the LCD with poor viewing angles. The FOV was measured by placing a protractor at the camera's center of projection as shown in Figure 4.16; this measurement method was inspired by similar measurements taken in William Steptoe's AR Rift System (Steptoe, 2014).

An Arduino microcontroller board and transistor circuit were used to drive the RGB LEDs in a color sequential mode that was synchronized to the monochrome display. Since the utilized LCD panels have a very slow response time (greater than the 16.67 ms duration of a 60 Hz frame), each color sub-frame was doubled and illuminated by the LEDs for a short period at the end of the second sub-frame to reduce switching blur. A occlusion mask sub-frame was also included to improve the see-through ability (see Section 4.5.5.3). With

79
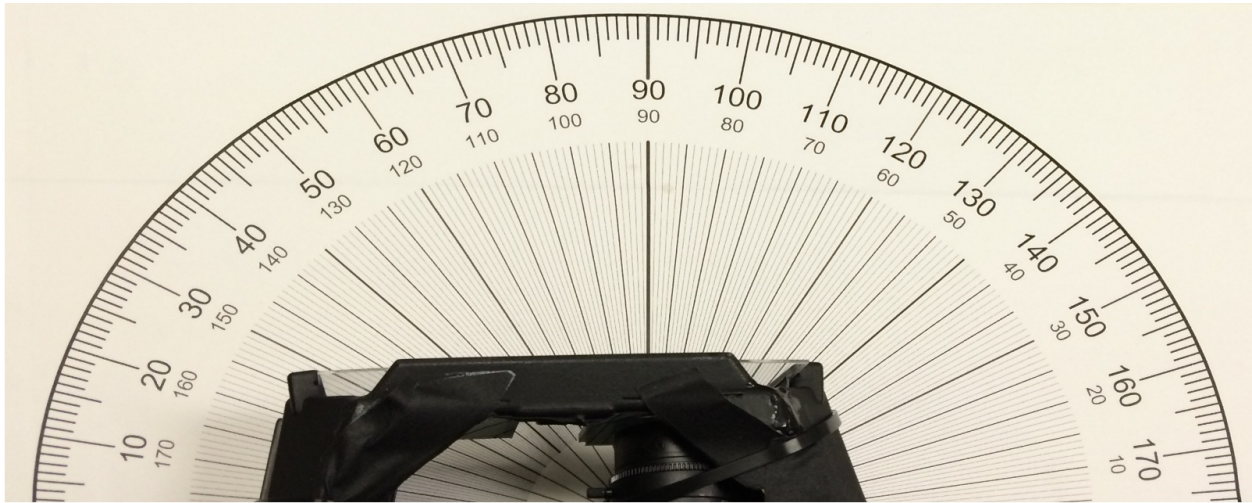
Figure 4.16: Display field of view measurement. A camera is mounted behind the first display prototype that approximates a human viewer wearing eyeglasses. A protractor is placed with the origin at the camera's center of projection to measure the field of view.
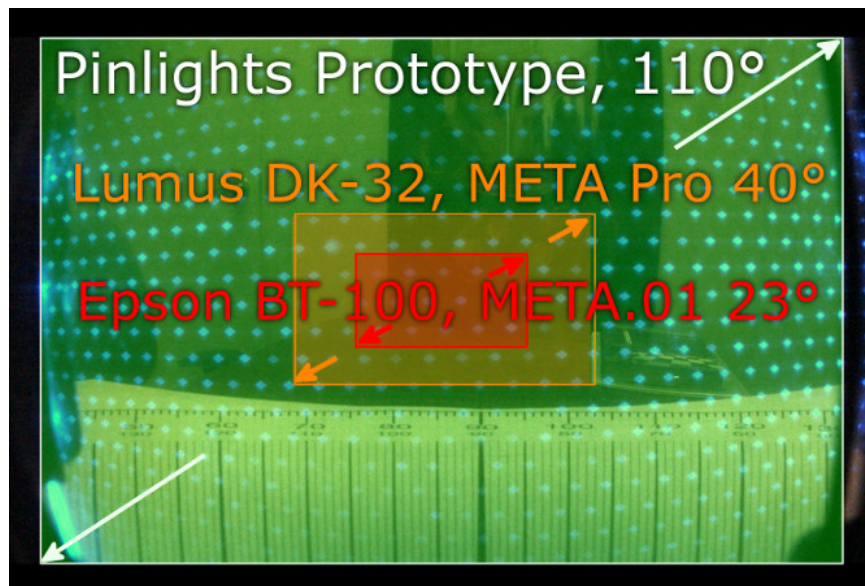
Figure 4.17: Display field of view comparison. An image taken through the camera, with the field of view of the first prototype display compared to state of the art commercial eyeglasses-like augmented reality displays.

the doubled subframes, the effective framerate was 7.5 Hz. A bright monochrome color mode was also tested with three duplicated color subframes that operated at 12 Hz; the LEDs were turned on the entire third subframe after the LCDs had fully switched during the first two subframes. All photographs of the display were taken from real-time video capture at the rates listed above with the exposure time set to the entire frame time (to integrate over all sub-frames) and synchronized to the camera. The LCD panels were controlled by an externally housed driver board connected by a DVI link to an Intel Core i7 PC with an NVIDIA GTX 580 GPU running Linux.

4.6.1.2 Second Prototype

To further test the design, a second prototype design was developed that was intended to test how well the design worked for human views. This prototype was designed in the untracked configuration, providing a constrained window around the eye of $a_{e_b} = 12$ mm, which provides an eyebox of $a_{e_b} - a_e = 9$ mm for a viewer pupil size of $a_e = 3$ mm. Two instances of the second prototype design were constructed: one with the same Sony LCX017 LCD panels as used in the first prototype design and one upgraded to Sony LCX086 panels. The newer LCX086 panels had significantly higher contrast than the LCX017 panels, but the specifications are otherwise the same. The resolution loss resulting from the large untracked eyebox reduced the effective horizontal spatial resolution to approximately 250 pixels per eye. The panels support a 60 Hz refresh rate and were operated in a monochrome mode to avoid the flicker that would result from color sequential operation. An occlusion frame was also not used to further avoid flicker. The pinlight array consisted of a 0.5 mm thick polycarbonate film commercially laser etched with 35 μm holes at a depth of 150 μm. The pinlight arrays were each illuminated by six RGB LEDs that were illuminated simultaneously at 100% duty cycle to create a white monochrome display. The modulation plane and pinlight plane were spaced at a distance of $d_p - d_m = 6$ mm, creating an optical assembly with a total thickness of 11 mm including 5 mm of component thickness, empty space, and plastic housing that could be reduced or eliminated in a future device. The eye center of projection to modulation plane distance was set to $d_m = 25$ mm, resulting in a combined stereo field of view of approximately $\approx 100°$ diagonally. The components were mounted in a slimmer redesigned 3D printed glasses frame (see Figure 4.18). The second prototype using the newer Sony LCX086 panels were driven using a Sanyo PLC-XF71 projector though a custom manufactured 2 m long flat flex cable. The projector was run without its lamps and most of its fans by using an Arduino microcontroller to simulate the status signals of these disabled components. The projector was connected to an Intel Core i7

81
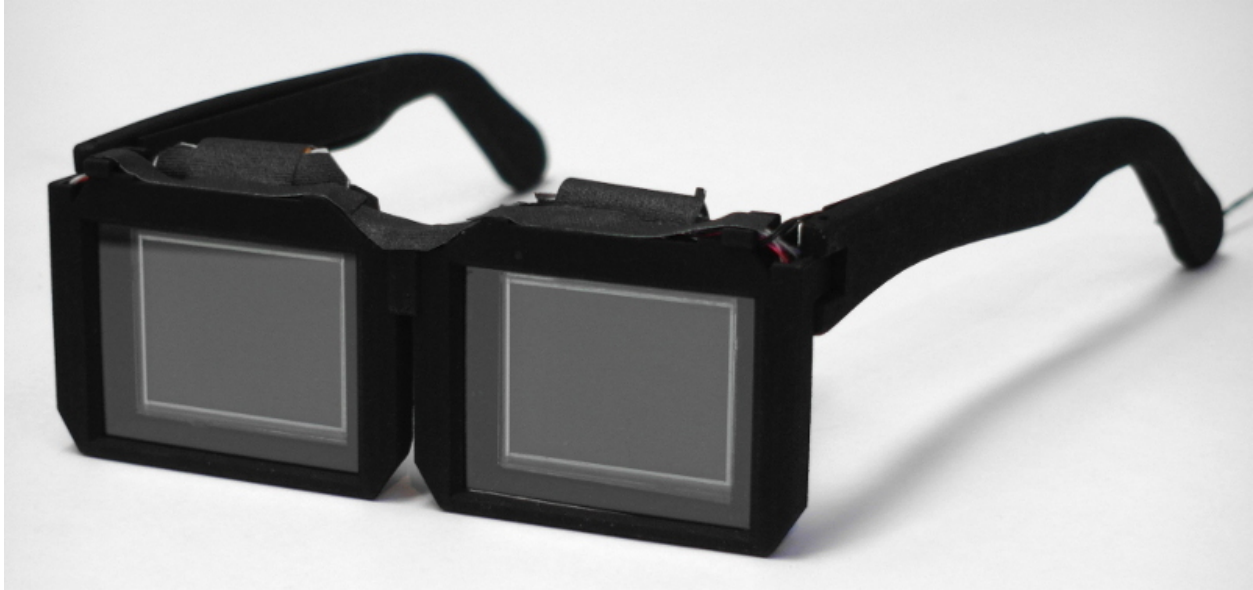
Figure 4.18: Second prototype display. (Driving electronics are external and not shown.)

PC with an NVIDIA GTX 580 GPU running Linux over a DVI link. The glasses were mounted in a rigid metal frame for protection during user testing (see Figure 4.25). The frame allowed the glasses to rotate about the Y-axis and tracked the rotation position along this axis using a shaft encoder, allowing spatially registered content to be shown in the glasses. A wireless gamepad was used with the glasses to allow the user to calibrate the system and control interactive demos.

4.6.2 Software

Real-time stereo tiled imagery for the prototype displays was generated with an OpenGL/GL Shader Language based program using the fast image transfer method described in Section 4.5.5.4. For the first prototype, the program also decomposed the image into color channels for color sequential operation and adjusted the image to compensate for poor LCD contrast at wide angles. For the second prototype, the software drew stereo tiled geometry at approximately 700 Hz for interactive applications (including the time needed to render a simple 3D scene), allowing the display to easily operate at the 60 Hz refresh rate of the LCD panels.

For the second prototype, a small set of demo and calibration routines were also developed. For calibration, the user was instructed to shift an image in each eye using a joystick until tiling artifacts were minimized, which centered the eyebox on each eye. The user was then instructed to align a rectangle in each eye to a physical rectangle in the room to align the two eyes and also align the glasses coordinate system to

the room. Once calibrated, the user could view a floating user interface, view various floating 3D models, or arrange a set of furniture items that appear spatially registered in the room.

Simulated images appearing in this section were generated by drawing the modulation plane image and a grid of pinlight dots at their apparent positions with respect to a virtual camera placed at the eye position. Images were summed over a densely sampled set of positions over a round camera aperture that matched the pupil diameter of the intended viewer.

4.7 Results

4.7.1 Simulated Results

To evaluate the theoretical performance of the proposed display design, various tiled pinlight projector configurations were simulated as described in Section 4.6.2 with a modulation plane that matched the specifications of the prototype display LCD panel (see Section 4.6.1).

The following configurations were tested: (1) a tracked virtual aperture configuration of $d_p = 29$ mm and $d_m = 15$ mm, (2) an untracked, small eyebox configuration of $d_p = 22.8$ mm, $d_m = 15$ mm, and $a_{e_b} = 5$ mm, and (3) a untracked larger eyebox configuration of $d_p = 18.6$ mm, $d_m = 15$ mm, and $a_{e_b} = 11$ mm. All configurations were simulated with an eye pupil size of $a_e = 3$ mm focused at a distance of $d_f = \infty$. Note that the tracked configuration is similar to that used by the first prototype display. The untracked, small eyebox configuration provides a miniature eyebox of $a_{e_b} - a_e = 2$ mm that could be used to compensate for tracker error, latency, pupil size variation. The untracked, larger eyebox configuration provides an eyebox that is close in size to the second prototype display.

Results are shown in Figure 4.19. Note that resolutions follow predictions of Sections 4.5.4.3 and 4.5.4.4 and that the untracked configurations exhibit the predicted uneven image toning. The tracked virtual aperture simulation may be compared to the result achieved on the first prototype display in Figures 4.23C and 4.23D. Note that these results do not simulate diffraction, which may cause additional image degradation.

Figure 4.20 also shows a view of the tracked configuration when the eye is not in the expected position and no eyebox is used. Notice that even a small shift out of the expected eye position causes significant display artifacts, providing motivation to use a small eyebox to compensate for tracker error. Also note that there is a repeating set of correct eye positions that correspond to a shift to the next pinlight projector; in the

Figure 4.19: Display simulations. Outlined inset images show magnified regions. Top Left: Reference image, used as target image for simulation. Top Right: Tracked configuration simulation. Bottom Left: Untracked configuration simulation (small eyebox). Bottom Right: Untracked configuration (larger eyebox).

figure, notice that display artifacts increase with error in eye position, but then subside when the eye nears the adjacent correct view position.

4.7.2 First Prototype Display

To evaluate real-world optical performance, the tiled pinlight projector design was initially tested using the first hardware prototype with a camera placed behind the display. The hardware prototype was configured in the tracked virtual aperture configuration (see Section 4.5.4.3) with the assumption that the eye position was known using a camera placed in a fixed position. See Section 4.6.1 for full specifications.

Figure 4.21 shows the steps of image formation on the first prototype. The image begins as overlapping defocused discs (Figure 4.21A) from the pinlight array which are converted to a plane of abutting hexagonal
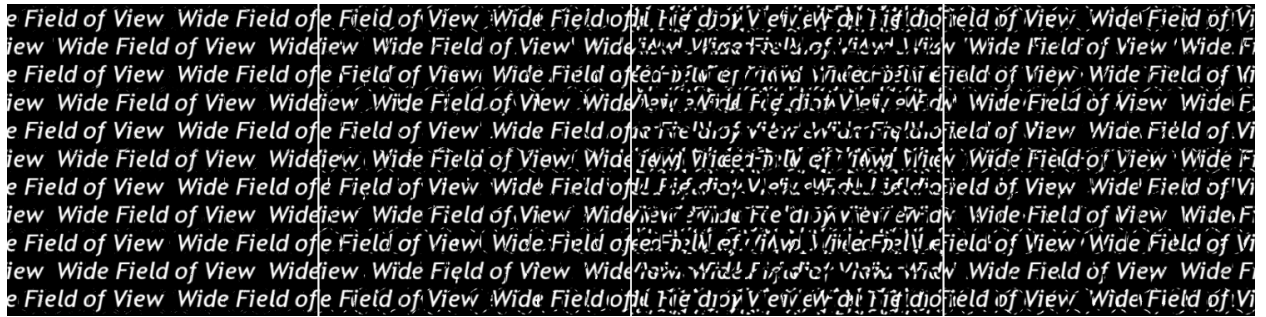
Figure 4.20: Tracker error. Images show simulated view of display in a tracked configuration when the eye is not in the expected position and the display is not configured with an eyebox. *Left to right:* (1) Eye is in expected position. (2) Eye is shifted laterally 0.25 mm. (3) Eye is shifted laterally 1.0 mm. (4) Eye is shifted laterally 2.5 mm and begins to return to a repeating correct eye position.

tiles through the encoding of a virtual aperture on the modulation plane (Figure 4.21B). An augmented image is then encoded in the modulation plane, which appears out of focus when viewed with unstructured illumination from the scene (Figure 4.21C). When illuminated, the strongly directional light from the pinlight array causes a focused augmented image to appear, although the dark regions of the modulation mask cause the background to appear dark, except for a glowing region around the virtual image (Figure 4.21D). When an occlusion mask sub-frame is included with the pinlight array off, the see-through ability is improved (Figure 4.21E).

Figure 4.22 shows the inputs to the first prototype display and the captured outputs when the display is operated in a color sequential mode with an occlusion mask sub-frame. During each color sub-frame, an encoded image is sent to the display's LCD panel (Figure 4.22A), which appears as Figure 4.22D when viewed through the display. During each occlusion mask sub-frame, the backlight is disabled and an occlusion image is sent to the LCD (Figure 4.22B), which appears defocused through the display (Figure 4.22E). The final perceived image is the integration of the color and occlusion mask sub-frames (Figure 4.22F), which can be compared to the theoretical performance of the display through simulation (Figure 4.22C).

Figure 4.23 shows the first prototype display generating imagery for a variety of AR scenarios: a gesture based interaction (Figure 4.23A), visualization of a detailed model (Figure 4.23B), and display of overlaid text (Figure 4.23C). Note that Figures 4.23B and 4.23C were taken in the high brightness monochrome mode and that Figure 4.23C shows the display utilizing its entire field of view of approximately $110°$ diagonally (per eye). Figure 4.23D shows a magnified region of Figure 4.23C which can be compared to the similar simulated case in Figure 4.19 (top right).
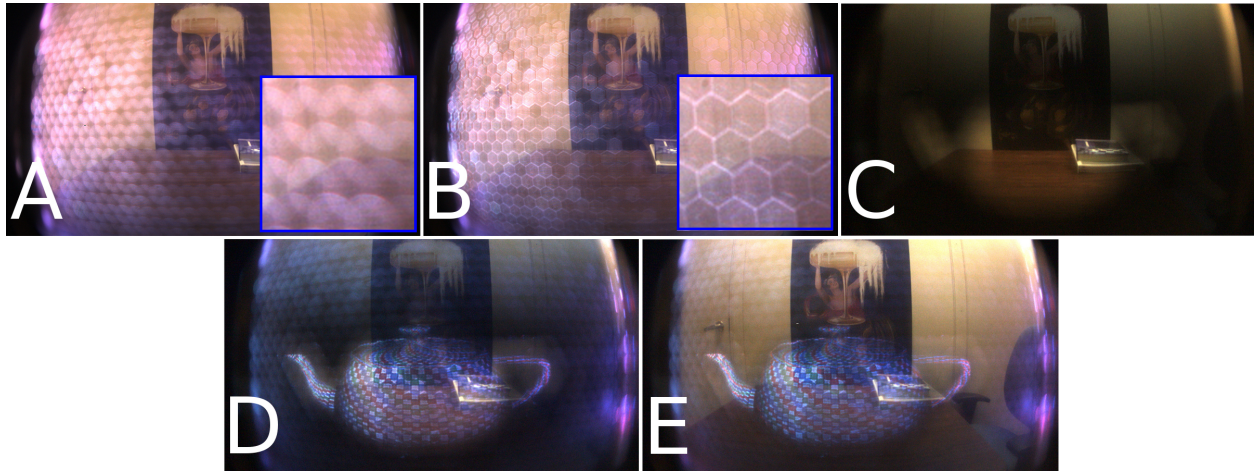
Figure 4.21: Image formation process on the first prototype device. Blue outlined inset images show magnified regions. A) Defocused pinlights form overlapping discs on image plane. B) A hexagonal virtual aperture is encoded in a modulation pattern. C) A modulation pattern is displayed without the pinlights illuminated. D) Pinlights are illuminated, causing the augmented image to appear. E) The image is displayed with a periodically displayed occlusion mask, improving see-through ability.
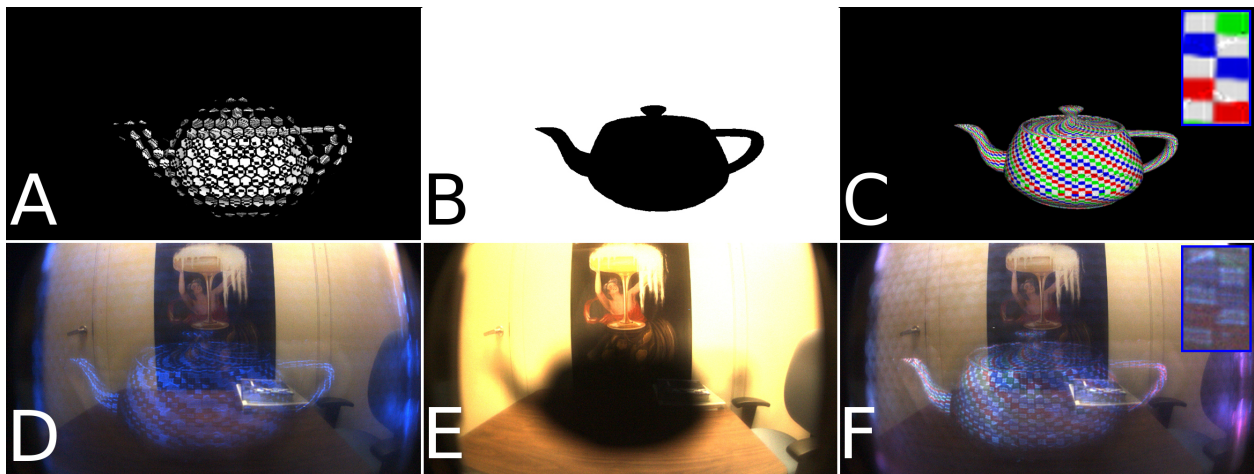
Figure 4.22: Sample display inputs and results. Blue outlined inset images show magnified regions. A) Modulation pattern for augmented image sent to display for blue color channel. B) Modulation pattern sent to display as occlusion mask. C) Simulated reconstruction of augmented image. D) Photograph of display while blue color channel displayed. E) Photograph of display while occlusion mask is shown. F) Photograph of actual augmented image and background, taken over complete color and occlusion cycle.
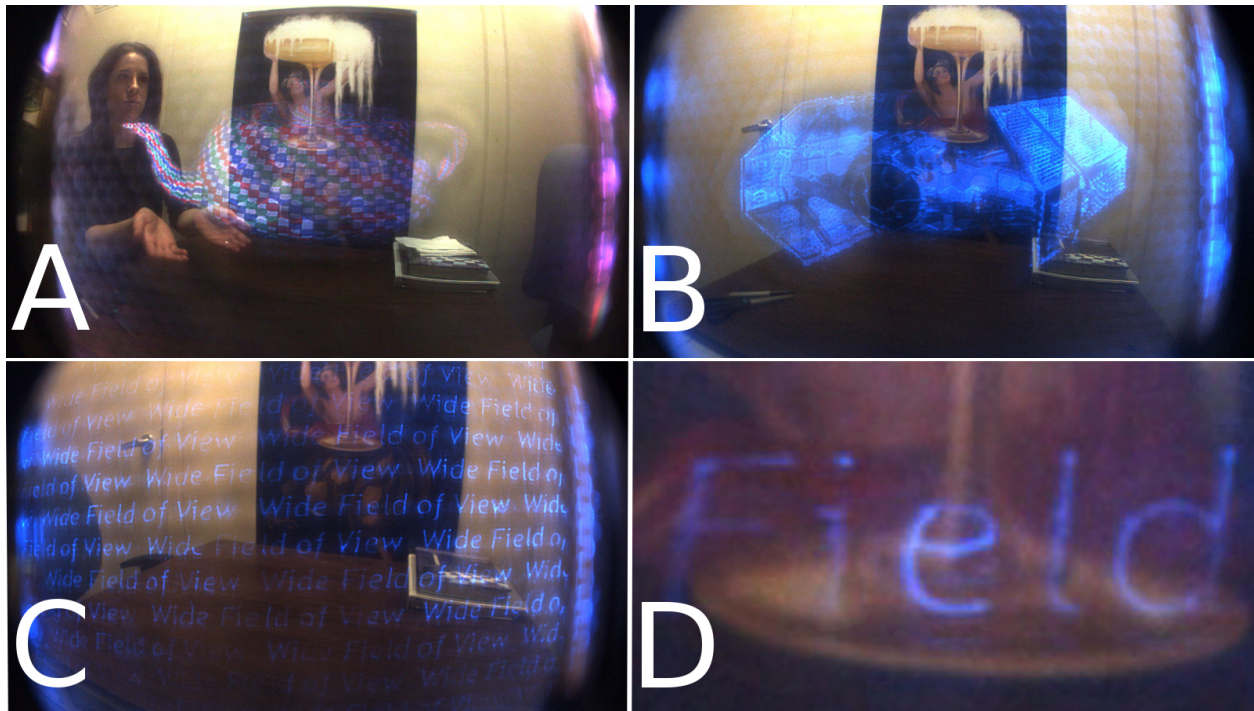
Figure 4.23: Sample results from first prototype display. A) User interacting with teapot. B) Model visualization. C) Text displayed to fill entire field of view. D) Magnified region of image C which represents a horizontal FOV of approximately $12°$. Ship model by Staffan Norling.
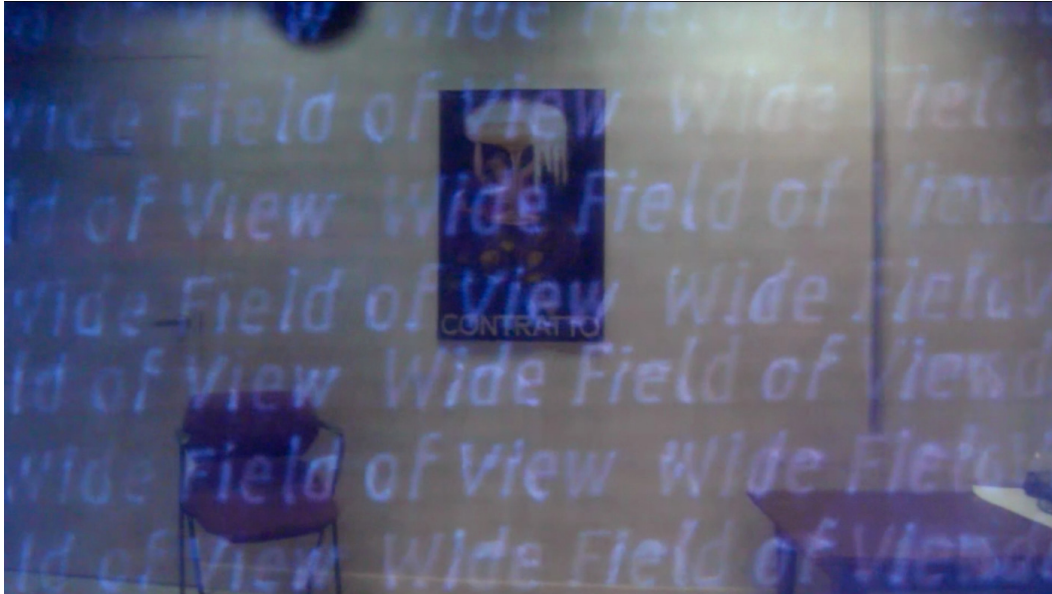
Figure 4.24: Sample results from second prototype display. Photograph was taken with a handheld camera placed in the eyebox of the display for the right eye. Note that the field of view of the display is slightly larger than could be captured with the camera.

4.7.3 Second Prototype Display and Demonstration

To show the pinlights approach is viable as a human viewable display, a second prototype was developed for demonstration purposes. The second prototype used the untracked light field configuration as described in Section 4.6.1.2. A sample result, taken with a handled camera, is shown in Figure 4.24. Note that this photograph was taken of the second prototype model using the newer Sony LCX086 panels. The resolution of this prototype is lower than that of the first prototype due to a significant portion of the display resolution being dedicated to creating a larger eyebox. Some toning artifacts (uneven brightness levels in the image) are also present, especially in large filled areas, although in the author's opinion they appear more pronounced in the photograph then when viewing the phototype display in person. Also note that this prototype produces an acceptable result (without significant glowing artifacts) without a time-multiplexed occlusion frame when the augmented imagery fills the majority of the field of view.

The second prototype display using the older Sony LCX017 panels was demonstrated at the SIGGRAPH 2014 Emerging Technologies exhibition (Vancouver, Canada, Aug. 10-14, 2014) (Maimone et al., 2014c), and the prototype using the newer Sony LCX086 panels was demonstrated at the 2015 Stanford Light Field Workshop (Stanford, CA, Feb. 12, 2015) and to various visitors to UNC. In the demos, each visitor sat down to view the prototype display mounted at a table (see Figure 4.25), and participated in the interactive

Figure 4.25: Demonstration of second prototype at SIGGRAPH 2014 Emerging Technologies.

calibration and demo routines described in Section 4.6.2. A formal study of user experiences was not performed, but some general observations were made of the several hundred vistors who participated in the interactive demos:

- **Calibration**: The calibration procedure proved tedious and unintuitive. During the first stage of the calibration procedure, users were asked to adjust a text calibration pattern (by sliding it up and down or left and right) until any tiling artifacts in the image disappeared, effectively centering the eyebox on each eye. Many users, especially those without experience in the area, were not sure of the goal of the procedure until they had degraded the image from its initial setting and then understood it should be adjusted back until optimal. This was especially a problem if the initial setting was near the optimal setting for the user. The second stage of the calibration, centering an augmented rectangle over a physical rectangle in the room, was intuitive to almost all users and was generally performed quickly. Many users had to perform the calibration procedure more than once to obtain acceptable results, taking several minutes of time – a significant portion of the total demo time. Communication could also be a

problem as many users were not fluent in the language of instruction, providing further motivation to develop future calibration procedures that are obvious without instructions.

- **User Reported Image Quality**: The vast majority of users reported they were able to see a wide field of view see-through image with acceptable quality. The most common minor complaint was image artifacts at the very edge of the display. This was likely because the eyebox was set to be just large enough for users to rotate their eyes around to see the full field of view, but did not leave sufficient tolerance for additional eye or head movement or self-calibration error. A small group of users, estimated at 10% of the visitors, reported moderate image problems (e.g. seeing image artifacts over a larger portion of the field of view or difficulty seeing the display out of one eye), but were able to complete the demonstration. A smaller group of users, estimated at 5% of the visitors, reported severe image problems (e.g. large gaps between the image tiles, or artifacts over the entire image) that prevented participation in the demo. Many of these users identified that they had specific advanced vision problems (e.g. astigmatism), or provided general comments such as "I have bad eyes".

- **User Feedback and Suggestions**: User enthusiasm in the demonstrated ranged from indifference or confusion as to the significance of a low resolution monochome display to amazement at the sight of a large object floating in front of them. The most common request was for more resolution, followed by requests for full color and tonal range. (The demos predominately used high contrast imagery, nearly black and white, due to the poor off-axis intermediate gray level reproduction on the LCD panels.)

In addition to the dynamic prototype described above, static prototypes were also presented at the demos. The static prototypes had a similar resolution and optical configuration to the dynamic prototype, but were handheld, monocular, and contained a laser photoplotted transparency film in lieu of an LCD panel (see Figure 4.26). The light sources consisted of six white LEDs powered by an AA-battery that were activated by holding down a switch on the side of the unit. Most users were able to view the static prototypes easily as the calibration procedure was much easier – the unit was simply held to one eye and moved around slightly until the viewer was in the eyebox (and thus the image appeared clear). Resolution was comparable to the dynamic prototype, but the image had a more textured, grainy appearance due to the print process. While testing, users tended to leave fingerprints on the pinlight array, which scattered light and reduced image contrast. The bare polycarbonate material used for the pinlight array was quite prone to scratching while cleaning (further

Figure 4.26: Static Prototype. *Left, Left Center:* views of prototype unit. *Right Center:* User viewing image on prototype. *Right:* Photograph taken through prototype unit.

scattering light), so it is recommended that acrylic or polycarbonate with an abrasion resistant coating is used instead.

4.7.4 Assessment

A real-time, compact, see-through AR display with a very wide field of view capable of displaying modest resolution imagery was developed and successfully demonstrated to a large group of users. However, there are a few fundamental and several practical issues to address in future work to create a display with improved image quality:

Eye Tracking As shown in the results, the display resolution would greatly benefit from eye tracking. Although some preliminary experiments were performed, a practical eye tracking system has not been integrated into any of the prototype displays. Possible approaches include placing a camera at the edge of a tapered backlight that can see the eyes through total internal reflection (Travis et al., 2013), or placing a camera highly off-axis (Świrski et al., 2012). There is also significant commercial interest in development of eye tracking solutions for near-eye displays; for example, SensoMotoric Instruments has developed a eye tracking device to incorporate in the Oculus Rift DK2 head-mounted display[5]. A display configuration with tracking and a small eyebox could be used to compensate for tracker error and latency.

Improving Contrast, Color Accuracy, and Frame Rate Display contrast depends on the contrast of the pinlight array (i.e. the ratio of the light emitted from the pinlights to that emitted or transmitted through the remaining pinlight layer area) and the contrast of the SLM (i.e. the ratio of transmitted light between the most

[5]http://www.smivision.com/en/gaze-and-eye-tracking-systems/products/eye-tracking-hmd-upgrade.html

transparent and opaque states). The prototype has moderate contrast as shown in Figure 4.24. In the case of the etched waveguide, pinlight array contrast can be improved by taking care to keep dust, fingerprints, and scratches off the array (which scatter light), to mask side light leakage, and to make the pinlights as large as possible while still preserving the maximum display resolution (see Section 4.5.5.1). The LCD panels in the prototype also suffer from poor color reproduction off axis, limiting the tonal range supported by the display. The refresh rate of the panels, while suitable for a monochrome display, is not fast enough to support a color sequential mode. LCD panels with better off axis color reproduction (e.g. IPS panels) and higher refresh rates (120 Hz+) are available, but may need to be custom manufactured to the specifications of a near-eye display.

Improving See-Through Transparency The transmittance of the selected spatial light modulators is approximately 25%; while efficient for an LCD, the see-through view is similar in transparency to medium sunglasses even when the panels are set to their most transparent state. It is expected that transparency could be improved with more light efficient SLM technologies, e.g. MEMS microshutters (Hagood et al., 2007).

Improving Color Quality The first prototype display (supporting color) exhibited "washed-out" and non-uniform color. Color quality could be improved by adjusting the individual intensities of the RGB LEDs and by performing radiometric calibration over the display.

Decreasing Image Toning The first prototype display produced a tiled effect, especially in large areas of uniform intensity, due to varying light intensity and slight misalignment between the pinlight projectors. Quality could be improved with better geometric calibration, the inclusion of radiometric calibration, and blending between neighboring projectors (Brown et al., 2005). In second prototype display (in the untracked configuration), image toning caused by overlapping projectors is a fundamental issue predicted by theory; however, the use of a dynamic pinlight array to allow the positions of the projectors to be swept or randomized over time may reduce perception of the effect.

Either configuration may also benefit from compensation of the Stiles-Crawford effect, the perceived dimming of light as it enters nearer the edges of the pupil, as the light from each pixel in a Pinlight Display generally enters only a small region of the pupil opening. The perceived intensity of light entering the pupil approximately follows the equation:

$$\log(\eta_0) - \log(\eta_r) = p \times r^2 \qquad (4.15)$$

where η_0 and η_r are the perceived intensities at the peak (near the center) and edges of the pupil respectively, r is the distance from the peak in mm, and p is a falloff constant which is typically 0.05 (Westheimer, 2008). Following this equation, a 3 mm pupil will experience an falloff to approximately three-quarters of the peak intensity at the edges, while larger pupils will experience greater falloffs.

Decreasing Aliasing Artifacts Some of the results exhibited aliasing artifacts, most evident on the ship model in Figure 4.23B. The effects could be reduced by using more careful sampling (e.g. anisotropic filtering) when projecting the scene from the eye focus plane to the modulation plane through the pinlights with respect to the eye.

Decreasing Device Thickness In the tracked configuration, the optical stack is typically thicker than ordinary glasses. However, most of the thickness consists of empty space; it is possible that this space could be reduced by placing a small refractive or diffractive element over each pinlight, causing it to appear more distant. Such elements should be small relative to the pupil size or operate on narrow wavelengths to avoid appreciable distortion of the see-through view. The use of a tracked configuration with a small eyebox, a likely candidate for a practical device, would also result in a thinner device.

Mitigating Diffraction Diffraction appears to fundamentally limit the resolution of the display to modest values below those of available SLMs and causes some degradation of the see-through view. This effect may be mitigated with LCDs designed to reduce diffraction (e.g. Chaing et al. (2005) and Benoît-Pasanau et al. (2010)). Major LCD manufactures (e.g. Samsung) are also now manufacturing LCDs designed for transparent applications, providing an industry motivation to reduce diffraction effects. A complementary approach is to attempt to predistort the image displayed on the SLM so that it more closely matches the intended image when diffracted.

CHAPTER 5

Summary and Conclusion

5.1 Summary

In this dissertation, two novel designs for see-through near-eye displays were described, each of which used computation as a fundamental part of the image formation process.

The first design used a "subtractive" process: the light from an omnidirectional area light source was filtered through a stack of time-multiplexed attenuation layers so that only the rays representing the desired light field (i.e. images within the viewer's accommodation range) remained. The patterns to display on the attenuation layers were generated using an optimization process. The attenuation patterns could also be purposed as a light field occlusion mask, allowing scenery from the environment to be selectively masked. A stack of partially transparent components placed directly in front of the eye was used to achieve a see-through capability. Compared to the state of the art, the display demonstrated a wider field of view (65° diagonally) than any known see-through near-eye display in a form factor approaching eyeglasses (with the exception of the other design proposed in this dissertation). It is also the only known see-through display in this form factor that supports an occlusion capability, and the known only display in this form factor that supports addressable focal cues over a wide ($\geq 60°$ diagonal) field of view. The primary challenge in designing the display was increasing the performance of a heavily constrained multilayer optimization; several new techniques (light field planarization, retinal optimization, and perceptual weighting) were added to state of the art multilayer optimization techniques to allow recognizable near-eye images to be formed without traditional optics and with performance generally in the range of 20 to 30 dB PSNR. Additional challenges remain, notably addressing diffraction limits and increasing the performance of the prototype hardware to match simulations.

The second design described was "additive": a near minimal set of rays needed to form an image, generated by an array of point light sources, was modulated by an SLM on the way to the eye. This hardware

could be thought of as a set of miniature projectors beaming images directly onto the retina, or as encoding images in the point spread function of the eye. A stack of partially transparent components placed directly in front of the eye was also used to achieve a see-through capability. Compared to the state of the art, the display demonstrated a much wider field of view ($110°$ diagonally) than any known see-through near-eye display in a form factor approaching eyeglasses. The primary challenge in designing the display was accounting for the round-shape, variable size, and movement of the human pupil. These challenges were handled by proposing an eye tracked configuration that encoded a tileable virtual aperture over the modulated image, and proposing an untracked configuration that generated a light field over the eye region. Addressing the diffraction limits of the display is the most fundamental remaining challenge.

5.2   Future Work

The most likely candidate for future work is to address the problem of diffraction. The wave optics model of the display (see Section 4.5.5.6) could be used to pre-distort the images displayed on the modulation layers so that they appear more correct when diffracted. The severity of the diffraction problem will increase with display panel density; however, it is expected that the ability to correct the problem will also improve as the wave nature of light is modeled on finer scales. Alternatively, a passive optical correction layer (e.g. an array of zone plates that operate over a narrow band of wavelengths) could be used to control the angular spread of light through the pixel apertures. The addition of a robust, compact eye tracking system would also help the designs achieve their best diffraction-limited resolutions.

An additional engineering effort would also help to build a more practical Pinlight Display system. The LCD panels could be replaced with faster spatial light modulators that support field sequential color and also better off-axis color reproduction. User calibration could be eliminated by using eye tracking cameras to measure the position of the eyes relative to the display and the interpupillary distance, or a new calibration procedure could be implemented that uses more intuitive calibration patterns. The display could also be fitted with a wireless video transmitter (or longer tether) that would allow the glasses to be worn in a more natural position and allow a wider range of head movements. The display could then be mated with a full six degree-of-freedom tracker to allow testing with a wider range of spatially registered applications. This enhanced system could be used to assess user experiences under a formal user study and could be used to analyze wider AR issues such as system latency. The ultimate goal of building such a system would be

to demonstrate the utility of a compact, wide field of view augmented reality display for applications that operate within the design's performance limitations (e.g. low resolution) – for example, a spatially registered object identification system.

5.3   Conclusion

The primary strength of the proposed computational approaches is that they combine thin form factors and support for very wide fields of view. The designs also use simple inexpensive hardware and support or show promise to reproduce the focal depth cues. The multilayer subtractive approach also has an occlusion capability, which is rare to find in a compact design. *To my knowledge, these attributes have not been reproduced together in any conventional optical designs.*

However, the primary weakness of the designs is modest resolution limits from diffraction, and it is unclear if these limits can be overcome without design modifications. Each design is also associated with various image artifacts, some of which may be difficult to eliminate. The designs also require eye tracking for best performance and have light to heavy computing requirements. *These limitations are rarely encountered in conventional display designs.*

Although it is unclear if either of the described see-through near-eye display designs will ultimately be suitable for commercialization, perhaps the most promising aspect of these designs is that they offer *a unique set of trade-offs as compared to conventional optical designs*. Some of these tradeoffs may be improved with additional design iterations or balanced with hybrid optical-computational designs. In this way, there is promise to obtain more portable and realistic displays that help realize the dream of a general-purpose augmented reality that unites computer graphics and human vision.

BIBLIOGRAPHY

Assaiante, C. and Amblard, B. (1992). Peripheral vision and age-related differences in dynamic balance. *Human Movement Science*, 11(5):533 – 548.

Benoît-Pasanau, C., Goudail, F., Chavel, P., Cano, J.-P., and Ballet, J. (2010). Minimization of diffraction peaks of spatial light modulators using Voronoi diagrams. *Opt. Express*, 18(14):15223–15235.

Blondel, V. D., Ho, N.-D., and van Dooren, P. (2007). Weighted nonnegative matrix factorization and face feature extraction. *Image and Vision Computing (submitted)*.

Brown, M., Majumder, A., and Yang, R. (2005). Camera-based calibration techniques for seamless multiprojector displays. *Visualization and Computer Graphics, IEEE Transactions on*, 11(2):193–206.

Cakmakci, O., Ha, Y., and Rolland, J. P. (2004). A compact optical see-through head-worn display with occlusion support. In *IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '04, pages 16–25.

Cakmakci, O. and Rolland, J. (2006). Head-worn displays: a review. *Display Technology, Journal of*, 2(3):199–216.

Cakmakci, O., Thompson, K., Vallee, P., Cote, J., and Rolland, J. P. (2010). Design of a free-form single-element head-worn display. *Proc. SPIE 7618, Emerging Liquid Crystal Technologies V*, pages 761803–761803–6.

Campbell, F. W. and Green, D. G. (1965). Optical and retinal factors affecting visual resolution. *J. Physiol. (Lond.)*, 181(3):576–593.

Chaing, H.-C., Ho, T.-Y., and Sheu, C.-R. (2005). Structure for reducing the diffraction effect in periodic electrode arrangements and liquid crystal device including the same. US Patent. US 6977705.

Cheng, D., Wang, Y., Hua, H., and Sasian, J. (2011). Design of a wide-angle, lightweight head-mounted display using free-form optics tiling. *Opt. Lett.*, 36(11):2098–2100.

Curcio, C. A., Sloan, K. R., Kalina, R. E., and Hendrickson, A. E. (1990). Human photoreceptor topography. *J. Comp. Neurol.*, 292(4):497–523.

De Smet, J., Avci, A., Beernaert, R., Cuypers, D., and De Smet, H. (2012). Design and wrinkling behavior of a contact lens with an integrated liquid crystal light modulator. *Display Technology, Journal of*, 8(5):299–305.

Dodgson, N. A. (2004). Variation and extrema of human interpupillary distance, in stereoscopic displays and virtual reality systems. In *Proc. SPIE 5291*, pages 36–46.

Gao, C., Hua, H., and Lin, Y. (2012). Ergonomic head mounted display device and optical system. US Patent App. 13/335,884.

Gotoda, H. (2010). A multilayer liquid crystal display for autostereoscopic 3d viewing. *In SPIE Stereoscopic Displays and Applications XXI*, 7524:1–8.

Gross, A. (2014). What's the problem with google glass? The New Yorker (online). Mar 4, 2014.

Guenter, B., Finch, M., Drucker, S., Tan, D., and Snyder, J. (2012). Foveated 3d graphics. *ACM Trans. Graph.*, 31(6):164:1–164:10.

Hagood, N., Barton, R., Brosnihan, T., Fijol, J., Gandhi, J., Halfman, M., Payne, R., and Steyn, J. L. (2007). 35.5l: Late-news paper: A direct-view mems display for mobile applications. *SID Symposium Digest of Technical Papers*, 38(1):1278–1281.

Heide, F., Wetzstein, G., Raskar, R., and Heidrich, W. (2013). Adaptive Image Synthesis for Compressive Displays. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 32(4):1–11.

Held, R. T., Cooper, E. A., and Banks, M. S. (2012). Blur and disparity are complementary cues to depth. *Current Biology*, 22(5):426 – 431.

Hiura, S., Mohan, A., and Raskar, R. (2010). Krill-eye: Superposition compound eye for wide-angle imaging via grin lenses. *IPSJ Transactions on Computer Vision and Applications*, pages 186–199.

Howard, I. and Rogers, B. (1996). *Binocular Vision and Stereopsis*. Oxford University Press.

Hu, X. and Hua, H. (2014). High-resolution optical see-through multi-focal-plane head-mounted display using freeform optics. *Opt. Express*, 22(11):13896–13903.

Innovega (2012). Eyewear industry presentation. http://innovega-inc.com/events.php.

Jurik, J., Jones, A., Bolas, M., and Debevec, P. (2011). Prototyping a light field display involving direct observation of a video projector array. In *IEEE International Workshop on Projector-Camera Systems (PROCAMS)*, Colorado Springs, CO.

Kiyokawa, K., Billinghurst, M., Campbell, B., and Woods, E. (2003). An occlusion-capable optical see-through head mount display for supporting co-located collaboration. In *IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '03, pages 133–.

Kress, B. and Starner, T. (2013). A review of head-mounted displays (HMD) technologies and applications for consumer electronics. In *Proc. SPIE*, volume 8720.

Lanman, D., Hirsch, M., Kim, Y., and Raskar, R. (2010). Content-adaptive parallax barriers: optimizing dual-layer 3d displays using low-rank light field factorization. *ACM Trans. Graph.*, 29(6):163:1–163:10.

Lanman, D. and Luebke, D. (2013). Near-eye light field displays. *ACM Trans. Graph.*, 32(6):220:1–220:10.

Lanman, D., Wetzstein, G., Hirsch, M., Heidrich, W., and Raskar, R. (2011). Polarization fields: Dynamic light field display using multi-layer LCDs. *ACM Trans. Graph.*, 30(6).

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by nonnegative matrix factorization. *Nature*, 401:788–791.

Levola, T. (2006). Diffractive optics for virtual displays. In *Journal of the Society for Information Display*.

Lingley, A. R., Ali, M., Liao, Y., Mirjalili, R., Klonner, M., Sopanen, M., Suihkonen, S., Shen, T., Otis, B. P., Lipsanen, H., and Parviz, B. A. (2011). A single-pixel wireless contact lens display. *Journal of Micromechanics and Microengineering*, 21(12):125014.

Liu, S., Hua, H., and Cheng, D. (2010). A novel prototype for an optical see-through head-mounted display with addressable focus cues. *Visualization and Computer Graphics, IEEE Transactions on*, 16(3):381 –393.

Maimone, A., Chen, R., Fuchs, H., Raskar, R., and Wetzstein, G. (2014a). Wide field of view compressive light field display using a multilayer architecture and tracked viewers. In *SID Display Week 2014*, San Diego, CA, USA.

Maimone, A. and Fuchs, H. (2013). Computational augmented reality eyeglasses. In *Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on*, pages 29–38.

Maimone, A., Lanman, D., Rathinavel, K., Keller, K., Luebke, D., and Fuchs, H. (2014b). Pinlight displays: Wide field of view augmented reality eyeglasses using defocused point light sources. *ACM Trans. Graph. (SIGGRAPH 2014)*, 33(4):89:1–89:11.

Maimone, A., Lanman, D., Rathinavel, K., Keller, K., Luebke, D., and Fuchs, H. (2014c). Pinlight displays: Wide field of view augmented reality eyeglasses using defocused point light sources. In *ACM SIGGRAPH 2014 Emerging Technologies*, SIGGRAPH '14, pages 20:1–20:1, New York, NY, USA. ACM.

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

Masia, B., Wetzstein, G., Aliaga, C., Raskar, R., and Gutierrez, D. (2012). Perceptually-optimized content remapping for automultiscopic displays. In *SIGGRAPH 2012 Posters*, SIGGRAPH '12, page 63:1.

Mohan, A., Woo, G., Hiura, S., Smithwick, Q., and Raskar, R. (2009). Bokode: Imperceptible visual tags for camera based interaction from a distance. In *ACM SIGGRAPH 2009 Papers*, SIGGRAPH '09, pages 98:1–98:8, New York, NY, USA. ACM.

Pamplona, V. F., Mohan, A., Oliveira, M. M., and Raskar, R. (2010). Netra: Interactive display for estimating refractive errors and focal range. In *ACM SIGGRAPH 2010 Papers*, SIGGRAPH '10, pages 77:1–77:8, New York, NY, USA. ACM.

Pamplona, V. F., Passos, E. B., Zizka, J., Oliveira, M. M., Lawson, E., Clua, E., and Raskar, R. (2011). Catra: Interactive measuring and modeling of cataracts. In *ACM SIGGRAPH 2011 Papers*, SIGGRAPH '11, pages 47:1–47:8, New York, NY, USA. ACM.

Rolland, J. P. and Fuchs, H. (2000). Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoper. Virtual Environ.*, 9(3):287–309.

Santos, P., Gierlinger, T., Machui, O., and Stork, A. (2008). The daylight blocking optical stereo see-through hmd. In *Workshop on Immersive projection technologies/Emerging display technologiges*, IPT/EDT '08, pages 4:1–4:4.

Shibata, T., Kim, J., Hoffman, D. M., and Banks, M. S. (2011). The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of vision*, 11(8):11.

Son, J.-Y., Saveljev, V. V., Kim, D.-S., Kwon, Y.-M., and Kim, S.-H. (2007). Three-dimensional imaging system based on a light-emitting diode array. *Optical Engineering*, 46(10):103205–103205–4.

Steptoe, W. (2014). Ar rift. http://willsteptoe.com/post/66968953089/ar-rift.

Sutherland, I. E. (1968). A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, AFIPS '68 (Fall, part I), pages 757–764, New York, NY, USA. ACM.

Świrski, L., Bulling, A., and Dodgson, N. A. (2012). Robust real-time pupil tracking in highly off-axis images. In *Proceedings of ETRA*.

Travis, A., Large, T., Emerton, N., and Bathiche, S. (2013). Wedge optics in flat panel displays. *Proceedings of the IEEE*, 101(1):45–60.

Wang, Y. Z., Morale, S. E., Cousins, R., and Birch, E. E. (2009). Course of development of global hyperacuity over lifespan. *Optom Vis Sci*, 86(6):695–700.

Westheimer, G. (2008). Directional sensitivity of the retina: 75 years of stiles-crawford effect. *Proceedings of the Royal Society of London B: Biological Sciences*, 275(1653):2777–2786.

Wetzstein, G., Lanman, D., Heidrich, W., and Raskar, R. (2011). Layered 3D: Tomographic image synthesis for attenuation-based light field and high dynamic range displays. *ACM Trans. Graph.*, 30(4).

Wetzstein, G., Lanman, D., Hirsch, M., and Raskar, R. (2012). Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 31(4):1–11.