

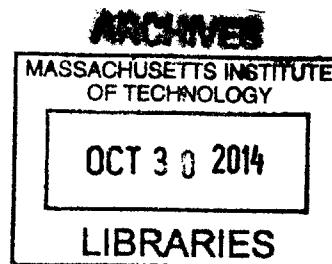
Computational Visual Reality

by

Matthew Waggener Hirsch

B.S., Tufts University (2004)

M.S., MIT Media Lab (2009)



Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2014

© Massachusetts Institute of Technology 2014. All rights reserved.


Signature redacted

Author _____

Program in Media Arts and Sciences
July 11, 2014

Signature redacted

Certified by _____

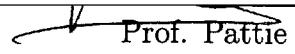
 Henry Holtzman
Research Scientist
Program in Media Arts and Sciences
Thesis Supervisor
Signature redacted

Certified by _____

Ramesh Raskar
Associate Professor of Media Arts and Sciences
Program in Media Arts and Sciences
Thesis Supervisor

Signature redacted

Accepted by _____

 Prof. Pattie Maes
Interim Academic Head
Program in Media Arts and Sciences

Computational Visual Reality

by

Matthew Waggener Hirsch

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on July 11, 2014, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences

Abstract

It is not so far-fetched to envision a future student working through a difficult physics problem by using their hands to manipulate a 3D visualization that floats above the desk. A doctor preparing for heart surgery will rehearse on a photo-real replica of his patient's organ. A visitor to the British Museum in London will sketch a golden Pharaoh's headdress, illuminated by a ray of sunlight pouring in the window, never aware that the physical artifact is still in Egypt. Though such scenarios may seem cut from the pages of science fiction, this thesis illuminates a path to making them possible.

To create more realistic and interactive visual information, displays must show high quality 3D images that respond to environmental lighting conditions and user input. The availability of displays capable of addressing the full range of visual experience will improve our ability to interact with computation, the world, and one another.

Two of the many problems that have impeded previous efforts to design high-dimensional displays are the need to:

1. process large amounts of information in realtime; and
2. fabricate hardware capable of conveying that information.

Light field capture and display is enormously data-intensive, but by applying compressive techniques that take advantage of multiple data redundancies in light transport, it is possible to overcome these challenges and make use of hardware available in the near-term.

This thesis proposes display and capture frameworks that use non-negative tensor factorization and dictionary-based sparse reconstruction, respectively, in conjunction with the co-design of algorithms, optics, and electronics to allow compressive, simultaneous, light field display and capture.

Thesis Supervisor: Henry Holtzman

Title: Research Scientist, Program in Media Arts and Sciences

Thesis Supervisor: Ramesh Raskar

Title: Associate Professor of Media Arts and Sciences, Program in Media Arts and Sciences

Computational Visual Reality

by

Matthew Waggener Hirsch

The following people served as readers for this thesis:

Signature redacted

Thesis Reader _____



Michael Hawley
Director and Curator, EG

Signature redacted

Thesis Reader _____

V. Michael Bove
Principal Research Scientist
Program in Media Arts and Sciences

Contents

| | |
|--|-----------|
| Abstract | 3 |
| 1 Introduction | 11 |
| 1.1 What is an 8D Display? | 13 |
| 1.1.1 Mathematical Model | 13 |
| 1.1.2 Scale of the Problem | 15 |
| 1.2 What is Compressive Capture and Display? | 17 |
| 1.3 Motivation | 18 |
| 1.4 Applications | 19 |
| 1.4.1 Passive Entertainment | 20 |
| 1.4.2 Gaming | 22 |
| 1.4.3 Medical | 23 |
| 1.4.4 Interaction and Interfaces | 25 |
| 1.4.5 Data Visualization | 26 |
| 1.4.6 Optical Computing | 27 |
| 1.4.7 Abstract Representation or Stimuli | 28 |
| 1.5 Dissertation Overview | 28 |
| 2 Background | 31 |
| 2.1 The Human Visual System | 31 |
| 2.1.1 Physiology | 32 |
| 2.1.2 Depth Perception | 37 |
| 2.1.3 Temporal Perception | 39 |
| 2.2 Light Transport | 40 |
| 2.3 Taxonomy of Display Systems | 40 |
| 2.3.1 Glasses-based Displays | 40 |
| 2.3.2 Glasses-free Displays | 42 |
| 2.4 Taxonomy of Camera Systems | 47 |
| 2.4.1 Light Field Cameras | 47 |
| 2.4.2 Depth Cameras | 47 |
| 2.5 Combined Systems | 49 |
| 2.6 Beyond Dirac Representations | 49 |
| 2.7 Interactive Techniques | 51 |
| 3 Compressive Methods for Visual Display | 53 |

| | | |
|----------|--|------------|
| 3.1 | Requirements for a Compressive Display System | 54 |
| 3.2 | Optically Efficient Methods | 55 |
| 3.2.1 | Tomographic Synthesis | 55 |
| 3.2.2 | Liquid Crystal Displays | 59 |
| 3.2.3 | Modeling Multi-Layer LCDs | 61 |
| 3.2.4 | Synthesizing Polarization Fields | 64 |
| 3.2.5 | Implementation | 66 |
| 3.2.6 | Assessment | 71 |
| 3.3 | High-Rank 3D | 76 |
| 3.3.1 | Content-Adaptive Parallax Barriers | 83 |
| 3.3.2 | Implementation | 92 |
| 3.3.3 | Assessment | 94 |
| 3.4 | Tensor Displays: A Compressive Display Framework | 101 |
| 3.4.1 | Tensor Display Framework | 101 |
| 3.4.2 | Analysis | 111 |
| 3.4.3 | Implementation | 118 |
| 3.4.4 | Assessment | 126 |
| 3.4.5 | Understanding Tensor Displays | 127 |
| 3.4.6 | Light Field Tensors | 129 |
| 3.4.7 | Light Field Tensor Factorization | 129 |
| 3.4.8 | Limitations | 137 |
| 4 | Compressive Methods for Visual Capture | 141 |
| 4.1 | Requirements for a Compressive Capture System | 142 |
| 4.2 | Optically Efficient Methods | 142 |
| 4.2.1 | Angle Sensitive Pixels | 142 |
| 4.3 | Sparse Reconstruction | 143 |
| 4.3.1 | Dictionary Learning | 144 |
| 4.3.2 | Reconstruction | 145 |
| 4.4 | A Switchable Light Field Camera | 145 |
| 4.4.1 | Light Field Acquisition with ASPs | 145 |
| 4.4.2 | Synthesis | 149 |
| 4.4.3 | Analysis | 150 |
| 4.4.4 | Implementation | 155 |
| 4.4.5 | Results | 158 |
| 5 | Compressive 8D Display | 161 |
| 5.1 | About This Chapter | 161 |
| 5.2 | A Classical Method | 162 |
| 5.2.1 | Implementation | 162 |
| 5.2.2 | Assessment | 164 |
| 5.2.3 | Prototyped Applications | 166 |
| 5.3 | Architecture | 169 |
| 6 | Applications and Extensions | 173 |

| | | |
|----------|--|------------|
| 6.1 | Focus 3D: Accomodation | 173 |
| 6.1.1 | Focus 3D Architecture | 177 |
| 6.1.2 | Implementation | 192 |
| 6.1.3 | Assessment | 194 |
| 6.2 | Compressive Light Field Projector | 200 |
| 6.2.1 | Compressive Light Field Synthesis | 202 |
| 6.2.2 | Implementation | 209 |
| 6.2.3 | Assessment | 214 |
| 6.3 | Soundaround: An 8D Display for Audio | 218 |
| 6.3.1 | Related Work | 220 |
| 6.3.2 | Multi-View Audio | 221 |
| 6.3.3 | Implementation | 225 |
| 6.3.4 | Assessment | 226 |
| 7 | Conclusion | 229 |

Thanks

I owe a tremendous debt of gratitude to my Ph.D. advisers and those with whom I have collaborated on the constituent work comprising this thesis. Ramesh Raskar has been a force of creativity and inspiration in this work, always pushing me to see beyond what is in front of me to what might be. Henry Holtzman, lover of great technology in many forms, has helped me center my work on human experience.

Douglas Lanman and Gordon Wetzstein have been collaborators, mentors, friends, and leaders. The compressive display and compressive capture frameworks presented in this thesis are built around their mathematical insights.

None of this could have happened without my closest friends: my fiancée Louise Flannery and my technical conscience, Tom Baran. They have kept me sane, and provided invaluable advice of every kind.

I thank my parents, Ellen Waggener and Paul Hirsch, and my grandfather, James Pendleton Waggener, Jr., for instilling in me a lasting sense of wonder and optimism for the world, without sufficient reserves of which no Ph.D. would be possible.

Members of the Camera Culture, Information Ecology, and Design Ecology groups at the Media Lab, as well as students and staff at the Center for Bits and Atoms have provided me with invaluable discussions, questions, advice, assistance, and more.

Chapter 1

Introduction

It is not so far-fetched to envision a future student working through a difficult physics problem by manipulating a 3D visualization that floats above his desk—perhaps using a finger to trace lines of integration through a field. A doctor preparing for heart surgery will rehearse on a photo-real replica of his patient’s organ. He may use a scalpel-like light-pen over a mock surgical table, while the scene before him is annotated by simulated vital readings and guides derived from medical imaging data. The owner of an upscale club, while preparing for an evening event previews materials and textures from a “Versailles” catalog on a tablet, tilting the screen to get the full effect of the ambient lighting on the bas relief tiles shown there. Finally settling on blue velvet wall coverings and a coffered ceiling, 8D displays covering the walls and ceiling of the club render photo-real, three-dimensional details that respond to the light sources in the room. A visitor to the British Museum in London will sketch a golden Pharaoh’s headdress, illuminated by a ray of sunlight pouring in the window, never aware that the artifact is still in Egypt.

These, and countless science fiction accounts of future displays like them, are compelling because they blur the line between visual reality and rendering in a way that current, planar, purely emissive displays cannot. The displays in the hypothetical scenarios above require general modulation of light transport to achieve a high level of realism. However, going from a 2D display to a glasses-free 3D display or a light transport display, is not as simple

as adding one or two additional dimensions. Moving along this trajectory, the underlying problem goes from two, to four, to eight dimensions. While the current brute-force Dirac sampled methods of display and capture work well for 2D problems, they fail to achieve the same apparent quality for high dimensional problems, where designs must resort to multiplexing. Compressive light transport displays will take advantage of natural redundancy in 4D and 8D light transport problems to greatly simplify their data and fabrication requirements, at the cost of increased computation.

In the next sections of this chapter, it may seem odd to work as hard as this thesis does to justify the development of 3D, 4D, or 8D displays. Science fiction and fantasy are famously rife with display technologies that explore the heights of human imagination and plumb the depths of technological impossibility (See Figure 1-1). Furthermore, popular interest in new display technologies is consistent and strong. However, the litany of disappointments in the fields of virtual reality, augmented reality, and glasses-based and glasses-free 3D over periods of decades have soured the commercial and academic appetite for advanced displays. The long history of display technologies that have failed to surpass the convenience and comfort of standard 2D screens is now inescapable when proposing a display with new capabilities. In some cases, it seems the prevailing public opinion has shifted to believe that binocular depth cues and other features absent in today's display technology are not useful, or worse, detrimental to experiencing immersive content or transmitting information.

This thesis is about display technologies that can reshape what we mean when we think of a display: its fundamental capabilities and applications. However, in recognition of the long history of the field, we will attempt to frame an argument justifying renewed attention in this area. The argument, though nuanced, comprises two key points:

- 3D, 4D and 8D display technology is indeed useful for a multitude of tasks including entertainment, human-computer interfaces, data visualization, and scientific research.
- The reason that previous attempts to expand the capabilities of display technology have been disappointing is that they have attempted to directly, ray-by-ray, scan out or capture light, following the formula for success to-date in 2D display. This

motivates the core contributions of the thesis to compressive light field capture and display.

1.1 What is an 8D Display?

An 8D display is a generalized light transducer. In abstract such a display is capable of reproducing any visual phenomena observed in the physical world—and many phenomena that go beyond what is possible with linear optics. In practice, as discretized physical devices, 8D displays will be subject to constraints on spatio-angular resolution (depth-of-field), intensity, contrast, color gamut (wavelength), temporal update rate, and field-of-view, and also be subject to various optical aberrations and imperfections. However, with these constraints, 8D displays offer the tantalizing possibility of moving beyond the flat and unresponsive screens of today, and creating a truer window into the digital world.

1.1.1 Mathematical Model

In this thesis, we adopt the methods of computer graphics to analyze the design space of 8D displays. The most fundamental description of the problem of computer graphics is embodied in the rendering equation [106].

$$L_o(\mathbf{x}, \omega_o, \lambda, t) = L_e(\mathbf{x}, \omega_o, \lambda, t) + \int_{\Omega} f_r(\mathbf{x}, -\omega_i, \omega_o, \lambda, t) L_i(\mathbf{x}, -\omega_i, \lambda, t) (\omega_i \cdot \mathbf{n}) d\omega_i \quad (1.1)$$

where λ is the wavelength of light, t is time, \mathbf{x} is a spatial coordinate, ω_o is an outgoing light direction, and ω_i is an incoming light direction, and Ω is the hemisphere of possible light direction vectors. Equation 1.1 represents a high-dimensional, complex model that is sufficient to capture all visual information generated by a point in physical space. While modern computer graphics has standardized around a geometric model that simplifies the

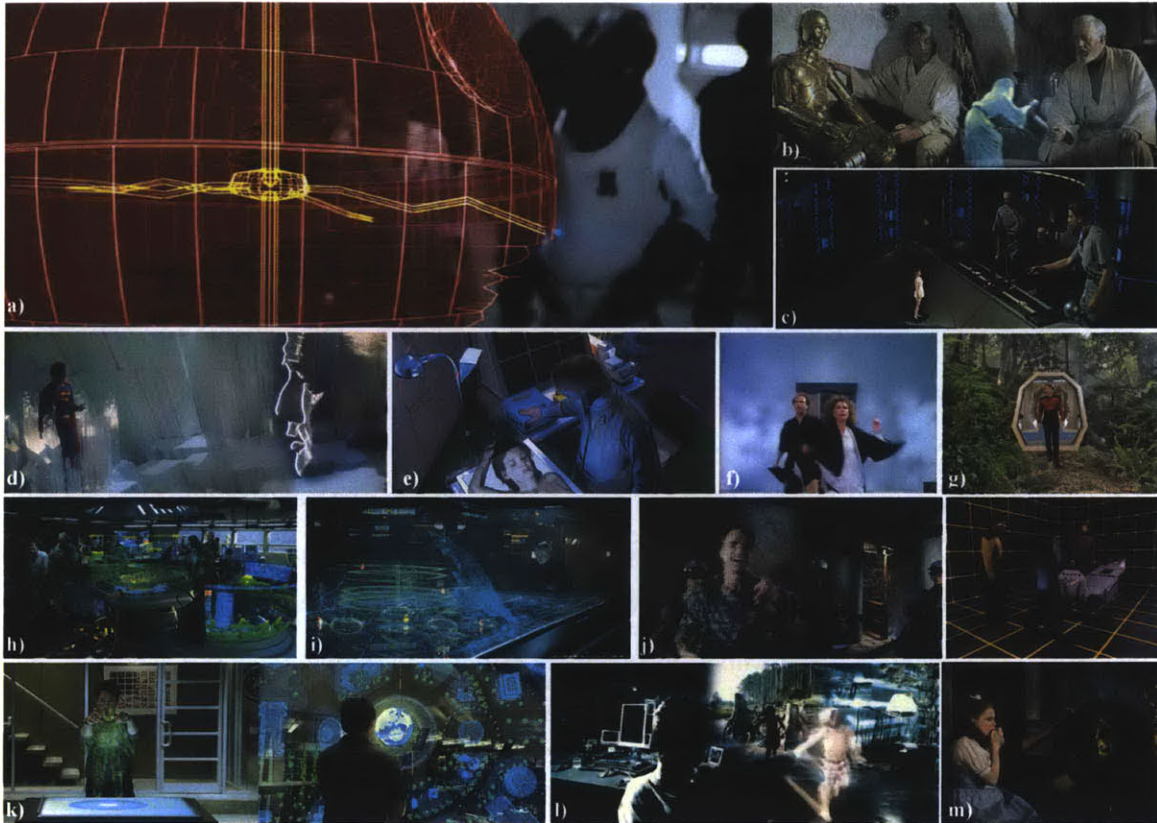


Figure 1-1: Novel display technologies proposed in science fiction films. a) *Star Wars IV: A New Hope*. A 3D Death Star map floats in the center of a command station and b) Princess Leia delivers a message. Perhaps the most well known depictions of advanced displays in fiction, these have inspired considerable follow-on work in film, literature, and research. c) *The Forbidden Planet*. An alien display renders live images that interact optically with the environment, driven by a brain interface. d) *Superman*. A ghostly, dimensional, disembodied head conveys messages from a stored consciousness. e) *A Scanner Darkly*. A hologram-like table-top display can project captured 3D images of people and objects. The agents wear a suit with an ever-shifting appearance to mask their identities. f) *The Veldt*. A Bradbury short story, adapted for television. Perhaps the first description of a physically and optically real virtual world created by advanced display technology. In a special room, characters experience computer renderings in perfect, life-like detail. The renderings can inflict physical harm. g) *Star Trek: The Next Generation*. The Holodeck is perhaps the best known example of a completely immersive display technology. Much like in *The Veldt*, characters in the Holodeck experience anything they can imagine to program, in complete optical and physical realism. The interactive displays of h) *Avatar*, i) *Prometheus*, and k) *Iron Man* represent a modern trend: depicting advanced gestural interfaces as mediated through 4D display. They share the characteristics of being semi-transparent (appear volumetric, additive). They depict 3D objects without a mediating display surface between the object and the observer's eye. In j) *Total Recall* a dimensional rendering of a person is projected non-line-of-sight, and reacts believably to scene lighting conditions. Observers accept that the person is physically present. l) *Minority Report* shows a 2D video re-rendered for a glasses-free 3D display. m) *The Wizard of Oz*. An oft-used trope in fantasy stories is the crystal ball. The observer looks in upon a life-like view of a spatially or temporally remote scene, presumably with all the clarity and depth of reality.

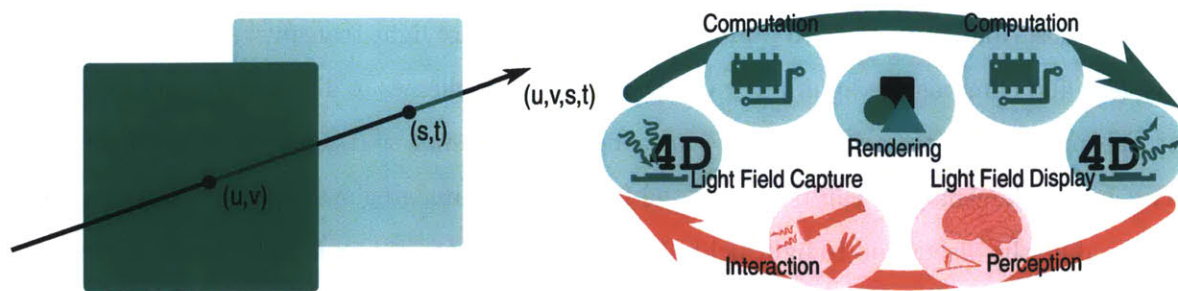


Figure 1-2: A mathematically convenient way to parameterize a light field is the two plane parameterization, shown here. A ray is identified in 3D space by two points of intersection with each of two planes, in this case (u, v) and (s, t) . This yields a 4D representation of light transport. When considering both capture and display problems simultaneously, the problem becomes 8D, but enables new forms of rendering and interaction.

rendering equation, a branch of computer graphics—image based rendering (IBR)—has instead developed a discretized, linear algebraic approach that requires sampling a simplified four-dimensional quantity derived from the rendering equation known as a light field [126, 63] (See Figure 1-2).

Though IBR has fallen out of favor for pure computer graphics applications, light fields have proven to be an invaluable tool in modeling advanced camera [147, 194, 135] and display [156, 99, 119, 206] systems. In these cases, the 4-dimensional quantity corresponds only to ω_i or ω_o over Ω and \mathbf{x} , as the light field is taken to represent only received or emitted light in a camera or display system, respectively, and not both.

The notion of an 8D display comes from simultaneously considering both input and output ray fields from the surface of the 8D display. This corresponds to modeling two unique sets of rays, ω_o and ω_i over (Ω, \mathbf{x}) .

1.1.2 Scale of the Problem

When considering Equation 1.1, it is clear that an 8D display will demand a great deal more data throughput than a 2D display; there are many more rays in a 3D volume than there are pixels on a 2D screen.

In Chapter 2 we describe a variety of existing approaches to creating glasses-free 3D systems.

We also describe approaches to build cameras that capture light transport. These light field displays and cameras predominantly use Dirac sampled schemes to directly convert 2D pixels at one moment in time into 4D rays in a light field, using spatio-temporal multiplexing. Though similar analyses can be conducted for light field imaging systems, in this section we will confine the discussion to light field displays. In order to achieve a high degree of realism, a light field display must support all visual stimuli recognized by the human visual system. Perhaps the most difficult effects to achieve are the related effects of accommodation and retinal blur; the refocusing of the lens of the eye creates spatial blur effects across the depth of a displayed scene. In order to achieve this effect with a ray-based light field display it has been shown that at least two angular samples are required across the pupil of the eye [180]. One type of display to achieve this effect is the Supermultiview lenticular display [181].

In order to build intuition about the scale of this problem, let us consider the bandwidth requirements of an iPhone “retina” display, and compare it to a hypothetical Supermultiview display with sufficient angular sampling rate and field of view to create accommodation when held in the hand at a comfortable distance.

The iPhone 5, released September 2012, has a screen resolution of 640×1136 and a refresh rate of 60 Hz. Assuming a 24-bit color space, the space-time bandwidth product for this display is over 1 GB/s.

For the purpose of this comparison, we will make some assumptions in order to determine the angular sampling requirements of the proposed Supermultiview display. First, we assume the pupil of the human eye is 5 mm, even though it varies with lighting conditions. Also, we assume a 45 cm viewing distance, as this is the approximate target viewing distance calculated for “retina” displays. Finally, we assume that the viewer will want to turn the device up to 45° in any direction, and that the 3D effect should work in any orientation (requiring a full 4D light field).

Under these assumptions, the angular extent subtended by the human pupil will be approximately 0.6° . Since the system requires a 90° field-of-view on each axis, and two samples across the pupil, the light field angular dimension will be 300×300 . At the native iPhone

5 refresh rate, color depth, and image resolution, this will require a space-time-angle bandwidth product of almost 100 TB/s.

The iPhone 5 display has 326 ppi, meaning each pixel is approximately $78\mu\text{m}$. To use a naive lenticular approach to achieve the hypothetical device of the last paragraph and maintain the iPhone 5 form factor, each pixel would need to be about 260 nm, while the wavelength of visible light is centered around 500 nm. Scaling pixel sizes below the wavelength of light will create severe diffraction blur effects.

It is clear from this back-of-the-envelope analysis that new approaches will be required from both optics and algorithms perspectives to meet the challenges of convincing, true-to-life display systems. Though the challenge laid out above may still be beyond near-term technical limits, in this thesis we lay the groundwork for methods that will mitigate the optical and computational challenges of advanced display systems.

1.2 What is Compressive Capture and Display?

Compressive light field acquisition refers to a set of techniques inspired by recent developments in compressive sensing [50, 14] and sparse data modeling. The intuition that motivates this problem is that, though light fields are dense and high-dimensional constructs, the information they contain is well structured. By creating a robust model for the data typically contained in light fields, it is possible to exploit the structure of natural scenes in order to capture a dense light field with fewer samples than predicted by the Nyquist sampling theorem. These techniques are described in detail in Chapter 4.

Compressive light field display is conceptually very close to compressive light field acquisition. The problem is the reverse: to emit light rays from a display in order to represent a desired light field by setting fewer pixel values than predicted by the Nyquist sampling theorem. Again, this can be accomplished by exploiting the structure of the light field to be displayed. These techniques are described in detail in Chapter 3.

1.3 Motivation

By many accounts we are living in a technological gilded age, in which the computational devices that surround and pervade our lives are ever more numerous and interconnected. The connections between our minds and our devices, by way of our sensory inputs, has not kept pace. The gap between the capabilities of the human visual system (HVS), and those of current display technologies is particularly wide. Though our visual systems are capable of disambiguating complex three dimensional scenes, disentangling the complex interplay of illumination and pigmentation patterns, perception across many stops of dynamic range and many wavelengths, the fundamental capabilities of our displays have not changed for a century—since it was possible to represent the illusion of motion. If we consider the problem of displaying text and images, an LCD screen is little better than a 4000 year old parchment.

There is significant evidence that, at least for spatial tasks, additional visual cues from advanced displays can increase understanding and competence. This has been shown definitively in the case of surgical robots [144], visualization [208], and CAD applications [140]. It is hard to enumerate now, as we embark on the journey to build the tools that will enable truly engaging advanced display systems, the ways in which human cognition and human computer interaction may benefit. With the correct toolbox, we argue that the addressable application space is large. Indeed, in Section 1.4 we sample some of the promising elements from the field of possible applications for this suite of technologies, though the bulk of this thesis is dedicated to the methods—electro-optical and algorithmic—for creating the next generation of light transport technologies.

It is well established that the human visual system is both the most informative and most costly of our senses. In addition to the dense packing of optically sensitive neurons in the eye, there are many stages of processing on the retina, in the optic nerve, and in the visual cortex of the brain. This is explained further in Section 2.1. This complexity has been finely tuned over millions of years to enable, from a signal processing perspective, a high-bandwidth sensory channel into the conscious mind. The unique capacity of this channel to

inform and measure the world justifies the great metabolic cost and evolutionary complexity of the visual system.

However, the HVS is not as simple or general as a man-made camera. Because vision is mediated by many levels of special-purpose neural processing, each level being adapted through evolution to historical features of our environment and survival needs, the information that reaches the brain cannot be considered as merely a bit stream. This will ring true to anyone who has attempted to read a bar code, or memorize the random distribution of grass in a field or sand on a beach. Rather, visual information that we hope to absorb easily and recall readily must be *coded* to avail itself of the existent and pre-determined special-purpose processing of the HVS. And as we shall see in later chapters, these processing layers are keenly tuned to shape, motion, depth, and the interplay of light and matter.

Today we find ourselves in the position of demanding an ever richer stream of information from our computational systems, primarily through visual channels. However, we continue to deliver this information using displays that present the HVS with an impoverished slice of visual reality. In this thesis we hope to provide the tools necessary for building the next generation of computer driven displays that can provide interactive imagery in the native format of the HVS.

1.4 Applications

Broadly, applications for glasses-free 3D display fall on a spectrum from direct spatial representation, to spatial metaphor, to abstract representation, to direct stimulation of the visual system to create non-physical visual effects. While 3D display systems coupled with optical input capabilities have not been widely examined to date, much of the application space overlaps with traditional glasses-free 3D display systems. The following exploration of applications has been ordered along the direct-indirect spectrum, starting with direct representations of real or imagined physical worlds for passive entertainment, and ending with the representation of abstract information, through direct stimulation of distinct layers of the HVS. The evolution of applications for advanced displays will be as much about

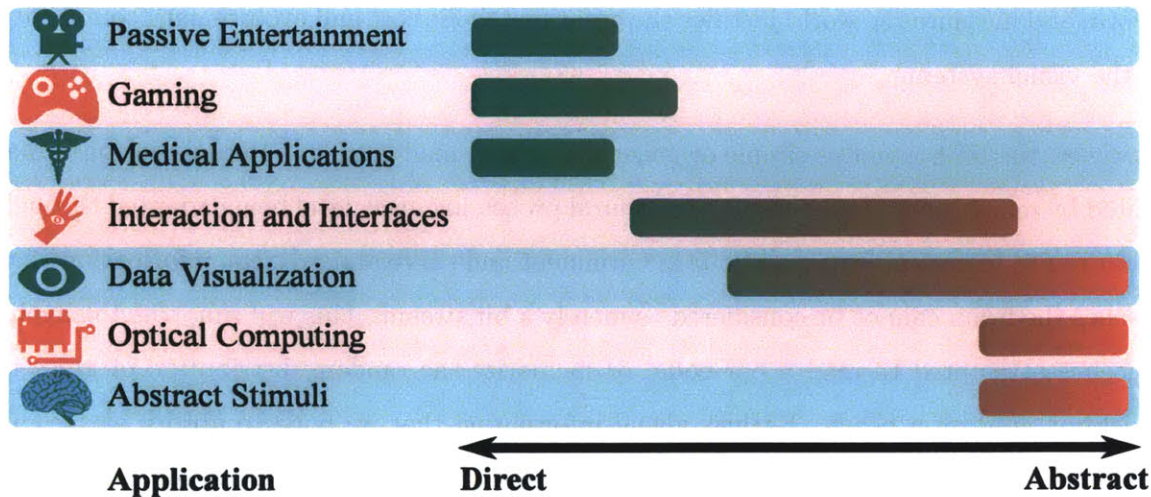


Figure 1-3: Application spectrum, from direct to indirect representation. 3D displays have been applied primarily to direct representation applications in the past. One of the promising directions for the future of advanced display systems is to more quickly and accurately convey abstract information.

our evolving understanding of human perception and cognition as it will be about the technologies that support arbitrary visual stimuli. The primary contributions of this thesis are technological, but we hope to motivate these technological pursuits by observing that there are multitudes of applications beyond the direct spatial representations of medical imaging, CAD, and 3D maps, especially once displays take on collocated light field input and output capabilities with sufficient spatio-temporal resolution.

1.4.1 Passive Entertainment

Since the inception of television the primary use case of display technologies has been in support of passive entertainment—watching film and animation to convey a story or events to a non-participating audience. Certainly the advent of ubiquitous general purpose computation has drastically changed the landscape of display applications. However when considering use cases for all types of 3D displays, the most often proposed are passive entertainment applications.

Since the 1950s stereo movies have flirted with audiences every two decades, never quite moving beyond gimmick status. More recently LCD panel manufacturers have sought to

differentiate their offerings in a fiercely competitive market by adding various forms of glasses-based and glasses-free 3D features. The lukewarm reception of 3D displays in the service of passive entertainment applications is rooted not only in the visual quality problems suffered by popular technologies, but more deeply in perceptual effects of adding additional depth cues.

If the goal of a display intended for passive entertainment is to allow an audience to lose themselves in an immersive story experience, in many cases binocular depth cues can stand in opposition to this goal. Instead 3D displays can push viewers into an “uncanny valley” of visual perception wherein the un-reality of the scene presented to the viewer becomes difficult to ignore. The most serious contributing issues are the loss of scale invariance, and limited depth-of-field, which necessitates depth compression. In the case of a typical flat display, the 2D images presented on the screen are invariant to screen size and distance, leaving the HVS to fill in expected sizes from features of the images. In contrast, binocular depth cues provide sufficient information to ground the size of the displayed virtual objects relative to the physical size of the bounding box of the screen. This makes it more challenging to ignore the physical size of the screen presenting the images, which may not be appropriate or reasonable for every scene of a movie.

By way of example, we present four common scenes in film that cause problems for various displays that present binocular depth cues.

- In order to increase the intimacy of a scene in a movie viewed by audiences in a theater using stereo glasses, a director chooses a tightly cropped close up on the lead actors’ faces. This appears to the uncomfortable audience as two giants pushing their faces through the front of the theater.
- A stereo movie is adapted for display on a stereo television. It contains a sweeping landscape fly-over filmed from a widely spaced stereo rig on a helicopter. Because binocular disparity does not occur over great distances, the appearance of disparity in the scene causes it to look like a diorama landscape the size of the television.

- A cartoon about dinosaurs is rendered for a stereo television. However, to a viewer the scene appears more like a puppet show created by marionettes.
- A movie contains a scene looking down a long hallway. The limited depth-of-field of a glasses-free 3D display requires the scene to be depth compressed in preparation for the screen. To the viewer this presents as an unnaturally flat looking world.

This analysis should not suggest that it is impossible to present properly formatted passive entertainment on an appropriate 3D display. However, it should suggest that such an undertaking will require a great deal of care, and may not be achievable by adapting content created for 2D displays. The analysis should also not suggest that content suffering from these problems is un-watchable. In fact personal preference around these issues varies widely, and many people find the experience of viewing 3D content enjoyable in spite of the many challenges in delivering such content appropriately. Finally, this analysis should suggest that as the size of the screen shrinks— for example, in going from a theater setting to a television in a living room—the problems of scale invariance and depth-of-field will only be exacerbated.

1.4.2 Gaming

Gaming is a promising application for 3D and other advanced display systems in that

1. content is rendered, rather than captured, meaning enormous flexibility in creating the right content for the right display, and
2. experiences can more easily be tailored to the capabilities of the display. This means that, unlike captured images, or images that seek to realistically represent physical environments, it is more easily possible to constrain the placement of virtual objects to the depth-of-field of the display, and work within the temporal, spatial, and angular resolution limits of the display.

The gaming industry has invested quite a bit in 3D gaming, with each of the major players releasing consoles and games that support the latest generation of commercial 3D televisions. Adoption faces the classic chicken-and-egg problem of requiring users to first own a 3D capable display, for which there is little available content, before they can play a 3D console game. 3D gaming on computer platforms has followed a similar trend, where many titles support 3D, but are crippled in that they must always cater to the overwhelmingly larger audience consuming 2D content.

A lesson to take away from this situation is that there is another related chicken-and-egg problem at play in the case of the adoption of advanced displays. 3D capable displays will not become widely distributed until there are compelling applications available that motivate consumers to overcome their cost of adoption. However, application developers tend to seek backwards compatibility with 2D displays in their application in order to ensure a minimum of available audience. This, by definition, means that the 3D effect in these backwards-compatible applications will always be unnecessary to its core functions. It is little wonder that few truly compelling applications exist for 3D in the consumer space when they are also required to be functional without 3D capabilities.

As display abilities increase beyond 2D, 3D, and 4D to true 8D displays the space of possible gaming applications will increase dramatically. Game designers working with such displays will have the ability to blend the real and the virtual seamlessly.

1.4.3 Medical

Medical diagnostics has been one of the early adopters of advanced displays [171], though progress has not been steady. Being a life-critical field, there is the potential that institutions responsible for treatment will pay a higher price than consumer markets for even small performance gains over commodity display equipment. This has led to research and development efforts to target advanced displays to medicine.

Another reason for early 3D display applications to have targeted medical imaging is that in many cases the data collected by medical imaging devices are naively 3D, or volumetric. A



Figure 1-4: A small cross-section of 3D gaming devices from the past decades. a) Early VR gaming platform Nintendo VirtualBoy. Released in 1995 it was a commercial failure. b) Autostereoscopic gaming console Nintendo 3DS. Released in 2011 it has achieved only moderate commercial success. c) Nvidia 3D Vision shutter glasses kit, version 1 and 2. Stereo glasses that work with desktop computers have been available since the 1990s. Nvidia’s offering was released in 2008. Such glasses have maintained a strong following in a niche market but have not become mainstream. d) LG Optimus 3D P920 and e) HTC EVO 3D are smartphones with parallax barrier autostereoscopic displays, both released in 2011. 3D capable smartphones have not been widely adopted and new models have not been produced. f) Oculus Rift VR display. Having improved upon long-standing pain problem areas in VR, such as latency, image resolution, and cost, the platform currently holds promise, but no products have been released or announced.

significant component of the analysis of this data is dedicated to interpreting 3D structures in a patient’s body and classifying those structures to determine treatments. This means that the sub-problem of serving doctors seeking to render diagnosis or treatment via medical imaging is a straightforward direct representation problem. The display system seeks to create as true-to-life a representation as possible.

However, as motivated at the opening of this section, applications for displays that spatially and temporally collocate optical transducers will be able to move beyond direct spatial representations of stored medical imaging data. Ultimately, advanced displays will lead to new and improved forms of interaction—covered in more detail in Section 1.4.4. This can lead to new forms of medical instruments which give doctors realistic live views of patient physiological information in real time across large size scales.

One example of such a device would be an advanced form of an otoscope for measuring a patient’s inner ear. Traditional otoscopes face a number of challenges, including illumination, and magnification. An electronic version composed of two linked 8D displays at different

size scales could present a novel solution to these problems. The practitioner would be able to view larger-than-life images of important structures in a patient's ear. The probe end of the device, a second 8D display, would be inserted into the ear in place of the tip of a traditional Otoscope. While capturing 4D information for viewing, the tip can also emit light, controlled by the light recorded by the viewing end. This will allow arbitrary, well controlled illumination.

1.4.4 Interaction and Interfaces

Many applications exist, such as the Otoscope above, which map the intuitions and skills developed by everyday existence into domains that are typically, by way of size, temporal scale, or complexity, remote. Coupled with insights about the nature of the Human Visual System summarized in Section 1.3, mediating remote domains through computation and advanced display systems will provide powerful new tools for Human Computer Interaction (HCI).

In Section 5.2.3 we explore an application developed on a prototype 8D display that allows a user to explore a medical imaging data set using a standard lamp as a virtual x-ray. In the demo, the intensity of incident light is mapped to the opacity of structures from a volumetric scan of a patient. Thus, as the user moves the lamp above a 3D rendering of the patient data, she can select different segments and layers of the data for viewing by positioning the lamp.

Optical information collected by an 8D display can also be used more indirectly to extract structure information about objects in front of the screen—for example the position and pose of a human hand. This means that displays with light field input capabilities can double as gestural computing peripherals [81].

Interaction scenarios employing advanced displays are not limited to interaction with computers and data. Interaction between people, mediated by advanced displays will play an enormous role in the future of this technology. Though it initially sounds like a small detail, creating a display that can link two speakers and conveys all the nuance entailed in

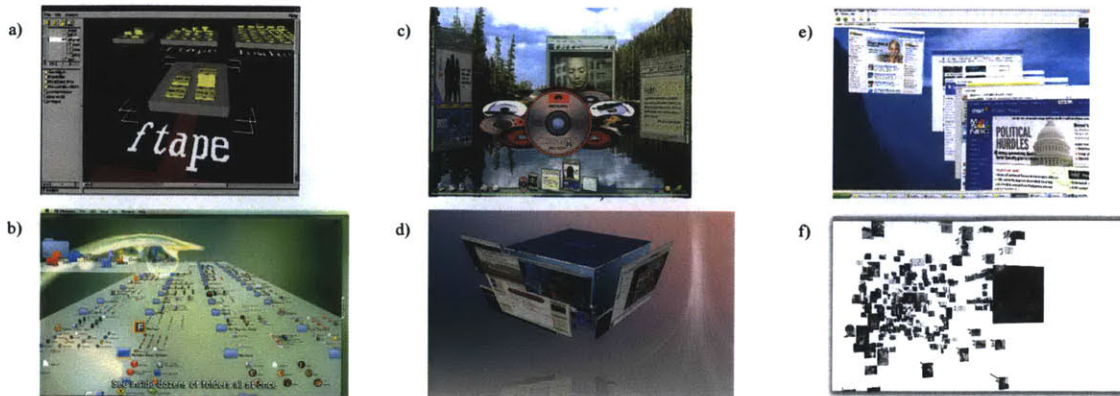


Figure 1-5: Examples of 3D human-computer interfaces. File browsers such as a) FSV in Unix and b) 3D fileSpace on Mac OS provided spatially distributed metaphors for file hierarchies. 3D desktop environments seek to use spatial metaphors for distributing operating system elements. Pictured are c) Sun Microsystems' Project Looking Glass, and the Linux compiz compositing window manager 3D desktop cube plugin. Similar ideas have been prototyped for web browsing applications as shown in e) SphereXPlorer and for organizing web content as shown in f) E15:FB, which allows the user to browse facebook data in a 3D environment.

human social interaction, not the least of which includes the ability to make eye contact while speaking, will greatly increase the efficiency of remote interaction. An 8D display, which collocates 4D light field input and 4D light field output can solve this problem.

However, often when interaction is considered it is in the context of HCI. With the advent of computer graphics hardware capable of cheaply rendering down-projected 3D onto 2D screens, there have been many attempts to create 3D environments for desktop computing and web browsing. Some of these are depicted in Figure 1-5. It is not clear that the advantages afforded by such interfaces outweigh the encumberment of navigating a 3D environment with the tools of a 2D interface. These environments have not been successful in finding adoption in the mainstream: a cautionary tale for those seeking to implement 3D interfaces.

1.4.5 Data Visualization

The problem of data visualization is one of transmitting information from a display into the head of a human being. Of all the applications listed in this chapter, the application of data visualization most clearly avails itself of the bandwidth, or channel capacity of the human

visual system. The area of information visualization has been well studied in the space of 2D displays [202], using both down-projected 3D and 2D representations. Less attention has been paid to data visualization on 3D displays [208].

To be sure, there is little that can be presented on a higher dimensional display that cannot be conveyed *eventually* on a lower dimensional display. As the display dimensionality is increased we expect cognition tasks related to digesting presented information to become more rapid, and more intuitive. For example, to extract 3D information from a 2D display, motion parallax over time, lighting cues, and more may be required—especially when a user is asked to identify the structure of unfamiliar or non-physical objects such as the data of a graph or chart. A 4D display adds parallax based on the users own head motion, and perhaps accommodation effects—all of which make the experience of viewing the display more like that of viewing an object in the real world. Finally an 8D display can add optical interactivity to the process of consuming data from a screen. A viewer can observe the imagery on the display under natural lighting conditions created by a hand-held lamp. All of this provides an additional level of realism. But when observing data that have no obvious physical analogy, it is possible to break physically analogous models, such as was described in the virtual x-ray demo in Section 5.2.3.

1.4.6 Optical Computing

An 8D display is a general purpose optical transducer. Within the spatial, angular, and temporal sampling limits of the hardware such an apparatus can be used to directly link physical optics with general purpose computation. This presents the opportunity to create hybrid optical systems, where portions of the system are implemented virtually behind an 8D display and other portions are implemented physically in front of the display. Further, multiple 8D displays can be linked to achieve bi-directional pass-through systems.

8D display is closely linked to computational light transport. Recent work has shown that it is possible to create hybrid computational and optical systems in order to iteratively examine and operate upon a light transport matrix [152, 153]. A possible application for

future 8D displays will be as a valuable tool in creating and measuring arbitrary light transport in order to characterize the light transport matrix of a scene. Thus, 8D displays with sufficient dynamic range could be used to change the appearance of objects in a room, or quickly make precise measurements of material properties.

1.4.7 Abstract Representation or Stimuli

Many visual phenomena are addressable by advanced displays beyond the perception of depth through binocular or other cues. Exploring these corners of the HVS will doubtless be a valuable application for future 8D displays.

One example lies in the area of color perception. It is known that color information is encoded differentially in human vision, where yellow and blue are opposed, as are red and green. Thus, with one eye one cannot observe the color “yellowish-blue” or “reddish-green”. However, it has been shown that causing opposed colored patches to merge in the same region of the visual field of both eyes of an observer, such exotic colors can be observed [45]. Thus it may be that 4D or 8D displays can allow observers to see colors beyond the standard RGB combinations typically observed in nature.

It has also been shown that presenting different intensity stimuli to the same region of an observers eyes can result in the perception of specularity (glossiness) [203]. This seems a natural result, as specular objects have a narrow reflective lobe, often leading to dramatic intensity differences with small changes in perspective. This fact suggests that 4D and 8D displays may be able to induce the perception of albedo beyond what is naturally observable.

1.5 Dissertation Overview

The aim of this thesis is to develop a framework for driving displays that support capabilities beyond 2D output into the mainstream. The framework is primarily technical in nature, describing the optics, electronics, and algorithms that support the goal of overcoming the longstanding limitations of high dimensional display.

In this section we describe the arc of this thesis document.

As display systems become more advanced, the technology and methods for creating displays becomes increasingly entangled with human perception, the structure of the information being displayed, and the application served by the display. In this chapter, we have laid out a line of reasoning to motivate the technical work to come. We have presented a non-exhaustive list of applications (Section 1.4) that may benefit from advanced displays—from 3D, to 4D, to 8D. And we have explored some of the reasons that displays have been unsuccessful in the past. We have also motivated the need for displays that simultaneously emit and capture 4D light fields (Section 1.1), as a special class of advanced display which we term an 8D display. Capturing rich optical information about the environment in which the display exists is a key feature required to create seamless experiences in which the display functions more like a window than a screen. We have also shown that the information requirements for advanced displays under naive sampling schemes are prohibitive.

The line of reasoning in this chapter suggests that compressive displays will be the way forward as we seek to overcome the challenges presented by supporting novel display applications and capabilities.

In subsequent chapters of the thesis, we situate and then detail the technical contributions towards solving the problems motivated by this chapter. Chapter 2 provides a family tree of prior work in the related problems of light field capture and light field display.

Chapters 3 and 4 introduce the methods used for compressive light field display and compressive light field capture, respectively. Both chapters follow a similar structure, introducing the requirements for light field display and capture systems and then presenting algorithms and electro-optical systems, designed holistically, to meet the proscribed requirements. It is an interesting result that the methods for compressive capture and for compressive display, though similar in concept, share very little in practical implementation. This is due to the fundamental asymmetry of the ordering of linear optics and non-linear computation in the problems of light capture (measurement) and display (representation).

Chapter 5 delves into the details of 8D displays. It provides a recipe for a working 8D display

prototype which uses a simple Dirac (classical) sampling scheme. The goal of this chapter is to describe the methods by which compressive 8D displays can be created, from both an algorithmic and hardware perspective. Chapter 6 details a number of applications that extend the framework presented in Chapter 3. Specifically, Section 6.1 details a method for creating accommodation cues in an advanced display system, Section 6.2 details a method for creating glasses-free 3D projection, and Section 6.3 explores the extension of the ideas developed in this thesis to the domain of audio, which serves as an example of a medium that is not well suited to the geometric ray approximation.

In this thesis we present results from working prototypes that demonstrate the compressive display and compressive capture principles outlined herein. We also present results from a working prototype 8D display that uses a classical sampling method, but implements some representation and interaction methods described in Section 1.4. Finally, we propose future hardware and algorithmic schemes to combine all of these presented concepts together into a computational 8D display. With some limitations, such displays will blur the lines between simulated or captured and replayed information, and the physical world. We adopt the term Computational Visual Reality to describe the space occupied by computational 8D displays.

Chapter 2

Background

Here, we will describe the foundational work upon which the thesis will be built. In the Sections 2.3 and 2.4 we describe the long history in optics and more recently, computer graphics, of creating and capturing light transport. This is achieved by introducing a taxonomy of display and camera systems. We will focus on glasses-free 3D display and light field capture, as these areas are deeply related to the techniques developed in this thesis. In order to understand the prior work in the 3D display space, and motivate this thesis, Section 2.1 describes the physiology of human vision—what cues we use to understand the world and disambiguate objects. Section 2.2 describes the background of light fields, a ray-based method of modeling light transport that has its origins in computer graphics, and a tool that is used throughout this thesis to develop methods for advanced display systems. In section 2.6 we provide the background to support compressed light field representations for display and capture, and contrast the two applications. We also provide examples of previous work in support of human-computer interaction using advanced display systems.

2.1 The Human Visual System

The human visual system is complex and involves many layers of processing. Cells within the eye respond to incident light in different wavelength bands and intensities [117]. The visual

cortex is divided into regions V1 through V5, which are connected through feed-forward and feedback relationships, and process increasingly more complex and abstract visual data. Processing occurs in V1 to accomplish many low level tasks, for example, detecting oriented edges. At higher levels, a single neuron may fire when a particular familiar face enters the visual field [161].

2.1.1 Physiology

The neuronal processing systems involved in vision range from simple, specialized, parallel operations, such as oriented wavelet filters, to broadly general, serial operations, such as object recognition. Different visual problems are processed at different levels of the visual system. For example, one can readily detect a red fruit among green leaves, or a patch with a pattern that does not match the background, but finding a specific type of object, say pliers, among a jumble of other objects, or finding a specific face in a crowd, requires slow methodical search. In the latter case, the attention of the viewer must dwell on each item and consider whether it matches some criteria. In this section we include a brief overview of the visual system for reference purposes.

One of the most attractive reasons to use advanced display systems capable of providing physiological depth cues is that more of the burden of visual processing can be shifted to the fast, parallel, neural paths, rather than slower, serial paths.

The Eye The eye contains the optics and sensors of the visual system. The cornea, crystalline lens, and vitreous humor of the eye all provide refractive power to focus images onto the retina. The lens shape can be altered with musculature around the front of the eye in order to provide adjustable focusing power. The eye also has an adjustable aperture, the iris. The structure, which is camera-like in nature, can be seen on the left of Figure 2-1. Notably, the eye has a curved focal surface, unlike most man-made camera systems, greatly reducing the required complexity of the optical system.

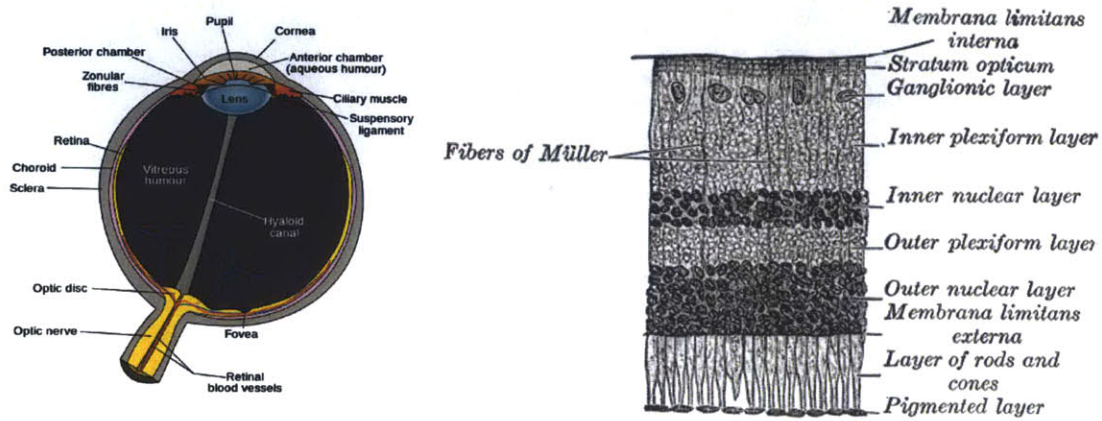


Figure 2-1: The eye and retina. The eye (left) contains the optics and sensors of the visual system. The cornea, crystalline lens, and vitreous humor of the eye provide adjustable focusing power, while the retina (detail, right) acts as an optical sensor. (Images public domain)



Figure 2-2: The point spread function of the optics of the eye (left). A letter 'E' shown at 20/20 acuity (center), is substantially blurred by the PSF of the eye (right). It is an area of active research to understand how the visual system recovers high fidelity imagery from this information. Images courtesy of Roorda Lab [164]

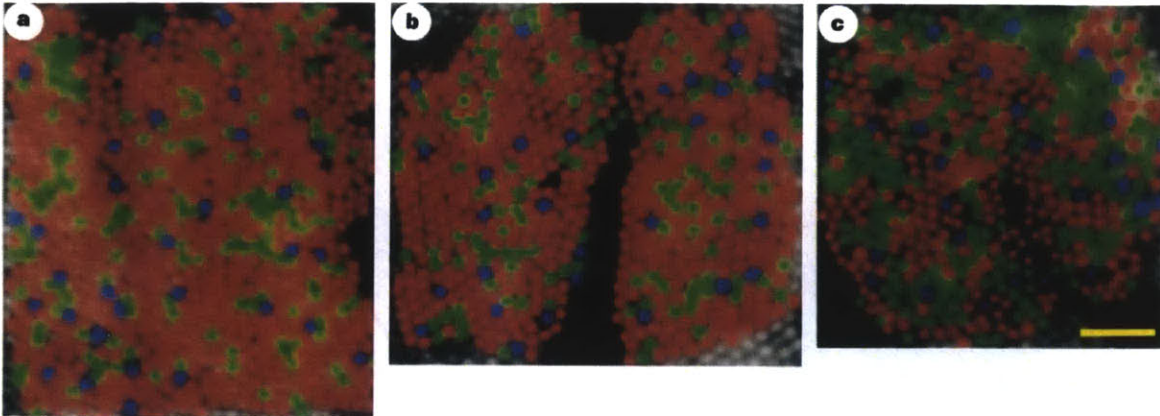


Figure 2-3: Pseudocolor images from various retinal regions of different subjects, courtesy of Roorda Lab [163]. Between different subjects and across regions of the retina the distribution of short (blue), medium (green), and long (red) receptive cells to varies randomly, with some general trends.

Retina The retina is the optical sensor of the visual system. Rod and cone cells are tuned to respond to low and high intensity light, respectively. Cone cells are further differentiated with long, medium, and short wavelength pigments, making them responsive to red, green, and blue wavelengths. Unlike a camera sensor, the receptive cells of the retina are not placed in a uniform or repeating pattern or with uniform density [162] (See Figure 2-3). Though the distribution of cells appears to be random, some generalizations can be made. The cells are most densely packed in the fovea — the center of the retina — which covers approximately two degrees of the visual field. This places the highest spatial resolution in the center of the visual field. Conversely, the highest temporal response appears at the periphery of the visual field.

Considerable early processing occurs on the retina. Bipolar cells create connections between neighboring photoreceptors in order to respond to bright and dark spots in the visual field. Horizontal cells create connections to larger areas of photoreceptors, and modulate the responses of the connected cells to enhance contrast. Varieties of retinal ganglion cells are instrumental in object detection, both spatially and temporally. These systems are still actively studied and not perfectly understood.

Visual Cortex The visual cortex contains multiple levels and pathways of processing, and a full description of the layout and function of this brain region is an active area of investigation in neuroscience. For the purpose of this section, it is sufficient to note the structure of the cortex, the macro-scale function of the pathways, and the abstraction levels at various processing stages, from simple image features calculated in parallel to high-level attentional processing that is abstract and conceptual.

The optic nerve, delivering impulses from the retinas of both eyes, enters the brain at the mid-brain in the Thalamus in a structure known as the lateral geniculate nucleus (LGN). The LGN combines the left and right visual fields of the eyes, and delivers them to the right and left brains, respectively. The majority of inputs into the LGN are modulatory, originating elsewhere in the brain. The LGN feeds into the Occipital lobe at the back of the brain, where the primary visual cortex is located.

The visual cortex is subdivided into distinct regions called cortical areas, named sequentially V1, V2, V3, etc. The first area of the visual cortex to receive visual stimuli, V1 primarily deals in low level stimuli such as intensity, color, and edge orientation. In fact, the V1 region contains a warped spatial map of the retina, known as a retinoptic map, as shown in Figure 2-4.

Later stages of processing in the visual cortex take place in V2, V3, and so on. The signals diverge into dorsal and ventral streams [62], with the dorsal stream hypothesized to be involved in object localization and motion planning, and the ventral stream involved object identification. Recent work has cast some doubt on the strict separation between the functions of these two areas [138].

Perhaps because the metabolic energy available to the brain is limited, or because the body is limited in the number of concurrent actions it can perform (for example, movement of eyes or arms), the brain has developed a neural activity suppression mechanism known as attention [75]. According to feature integration theory [190], attention allows low level visual features, such as color, motion, and edge orientation, to be fused into more complex, higher-level descriptions. Thus, simple distinctions, such as those between the texture of the region

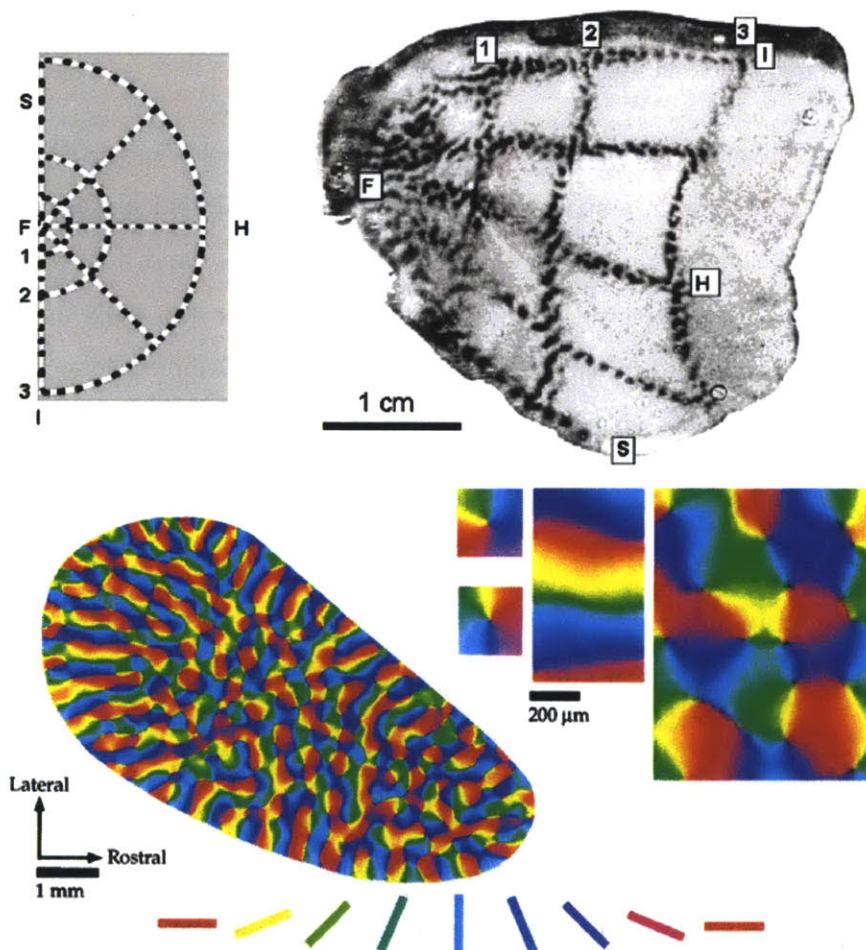


Figure 2-4: Images of the V1 cortex of a macaque monkey. (Left) the monkey was injected with a radioactive substance, taken up in brain tissue with neural activity. The sedated monkey's eyes were exposed to the blinking target shown on the left. The monkey was sacrificed and the brain photographed with a device sensitive to radioactivity. The warped spatial map of the retina can be observed. [185] (Right) In a similar experiment, a map of V1 response to oriented edges was produced [23].

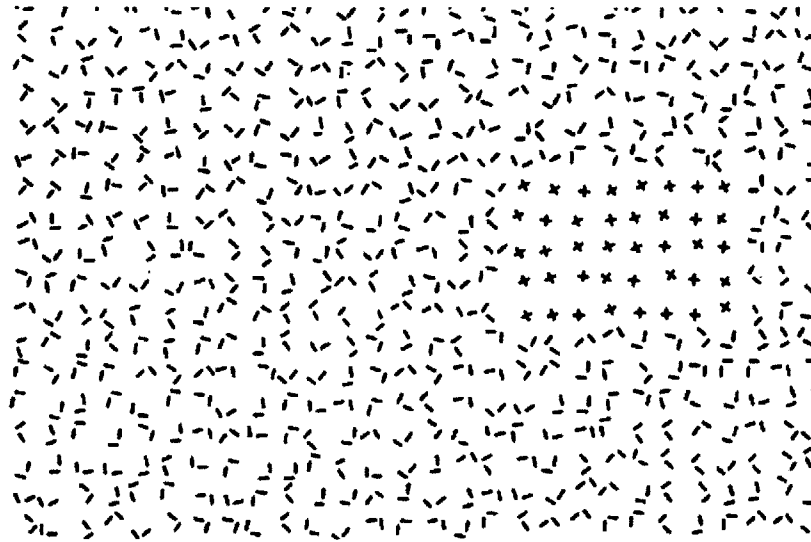


Figure 2-5: An example comparing parallel and serial visual search [75]. The + stands in contrast to the disjoint L, as it can be identified pre-attentionally, early in the visual cortex. The distinction between the L and T symbols requires attentional processing.

composed of + symbols in Figure 2-5, and the region composed of disjoint L symbols, can be quickly recognized with low-level parallel processing. More complex distinctions, such as those between the disjoint L and disjoint T symbols in Figure 2-5, require feature fusion, and therefore serial, element-by-element attentional processing.

Binocular depth cues are processed at a pre-attention stage, and therefore represent a powerful set of visual cues for rapid visual processing.

2.1.2 Depth Perception

When the visual system is functioning correctly, the overlapping visual fields of the two eyes are stitched together into a seamless single view of the world, with depth information extracted from an expansive set of cues. This section will focus on the cues used by the visual system to extract depth information from an arbitrary scene [80].

Psychological Cues These cues rely on high-level scene interpretation and prior experience. Thus they can be conveyed in a flat picture, do not require binocular vision, and therefore are supported by existing 2D displays.

Perspective The lens of the eye creates a perspective projection of the world onto the retina. Parallel lines vanish at a point at infinity.

Relative Size Objects of apparently similar size appear smaller as they recede into the distance.

Known Size Objects of known size appear smaller as they recede into the distance. The size of familiar objects such as other people can be used as a depth indicator without a reference.

Atmospheric Effects Distant objects are obscured by additional atmospheric haze.

Occlusion Distant objects are obscured as they pass behind nearer objects. This cue yields scene ordering.

Texture Gradients Repeating patterns diminish in size as they recede from the viewer.

Lighting and Shading The pattern of shaded and illuminated features on an object provides a strong depth cue.

Motion Parallax As the head is moved from side to side, regions of the background are revealed or obscured.

Physiological Cues [80] These cues are nearer to direct physical measurements taken by various components of the visual system. They gauge the true depth of a scene, and therefore give misleading or inconsistent cues when viewing a 2D image of a scene. There is a complex interplay between the physiological and psychological cues that is not yet fully characterized, although it is known that providing conflicting cues in some circumstances can lead to viewer discomfort [85].

Binocular Disparity The relative offsets of objects viewed from the different perspectives of the right and left eyes yields depth information for objects within a few meters of the viewer.

Convergence The eyes are aimed at an object of interest. The degree to which the eyes are “toed in” – towards being cross-eyed – gives the visual system depth information about the object.

Accommodation The muscle tension required to focus the eye on a nearby object is proportional to its distance.

Retinal Blur Related to accommodation, the degree to which the image of an object is blurred on the retina is proportional to distance.

2.1.3 Temporal Perception

The human eye has a finite temporal response rate, which manifests as temporal averaging. Thus, high frequency temporal stimuli will appear increasingly uniform as the frequency is increased. The rate above which the average person will perceive a flickering light source to be uniform is known as the flicker fusion rate [74]. As the eye is not a simple camera, the flicker fusion rate is a function of six inputs:

- The modulation frequency of the lighting source.
- The peak-to-peak value of the modulation.
- The mean intensity of the modulation.
- The wavelength of the light source.
- The position of the source in the field-of-view of the of the observer.
- The degree of dark adaptation of the viewer (i.e. previous recent exposure to bright or dark).

This variation is primarily explained by the distribution of rods and cones across the eye, and the relative wavelength, intensity, and temporal sensitivity of these types of cells. By understanding the temporal parameters of human vision, we are able to make displays that are tuned to best deliver only the important components of a scene.

2.2 Light Transport

Light transport has been studied from physically derived principles, both as a wave and a particle, from the time of Newton’s *Opticks*, 1704, and Huygens *Traité de la Lumière*, 1678. More recently approximations more suitable to computational representation have been developed for analysis of camera [2] and display [70] systems.

One computationally tractable representation of light transport known as the light field (Figure 1-2, Left) has become a valuable tool in rendering [126], and more recently in developing next-generation cameras [147, 194] and displays [156, 99]. Light fields have also seen limited use for HCI, where depth information, and gesture can be extracted from a light field captured through a SIP screen and combined mask [81]. Such systems can also support interaction using light emitting widgets [81, 184]. The benefit of this higher dimensional geometric optics formulation is an intuitive and mathematically concise set of operators to describe light transport through optical systems and free space. The intuition developed through light field analysis can benefit interaction researchers in understanding emerging display and capture technologies.

In computer graphics the rendering equation [106] encapsulates the complexity of light transport, capturing both reflection and emission. An efficient representation of reflection known as the Bidirectional Scattering Distribution Function (BRDF) [201], or more generally the Bidirectional Subsurface Scattering Distribution Function (BSSDF) [97], is closely related to the eight dimensional light transport described in this thesis (Chapter 5), as it describes the light transport to and from an object.

2.3 Taxonomy of Display Systems

2.3.1 Glasses-based Displays

The simplest type of 3D display, which has seen episodes of popularity in various forms since its invention in 1838 [207] is the stereoscope. Presenting two different images to the

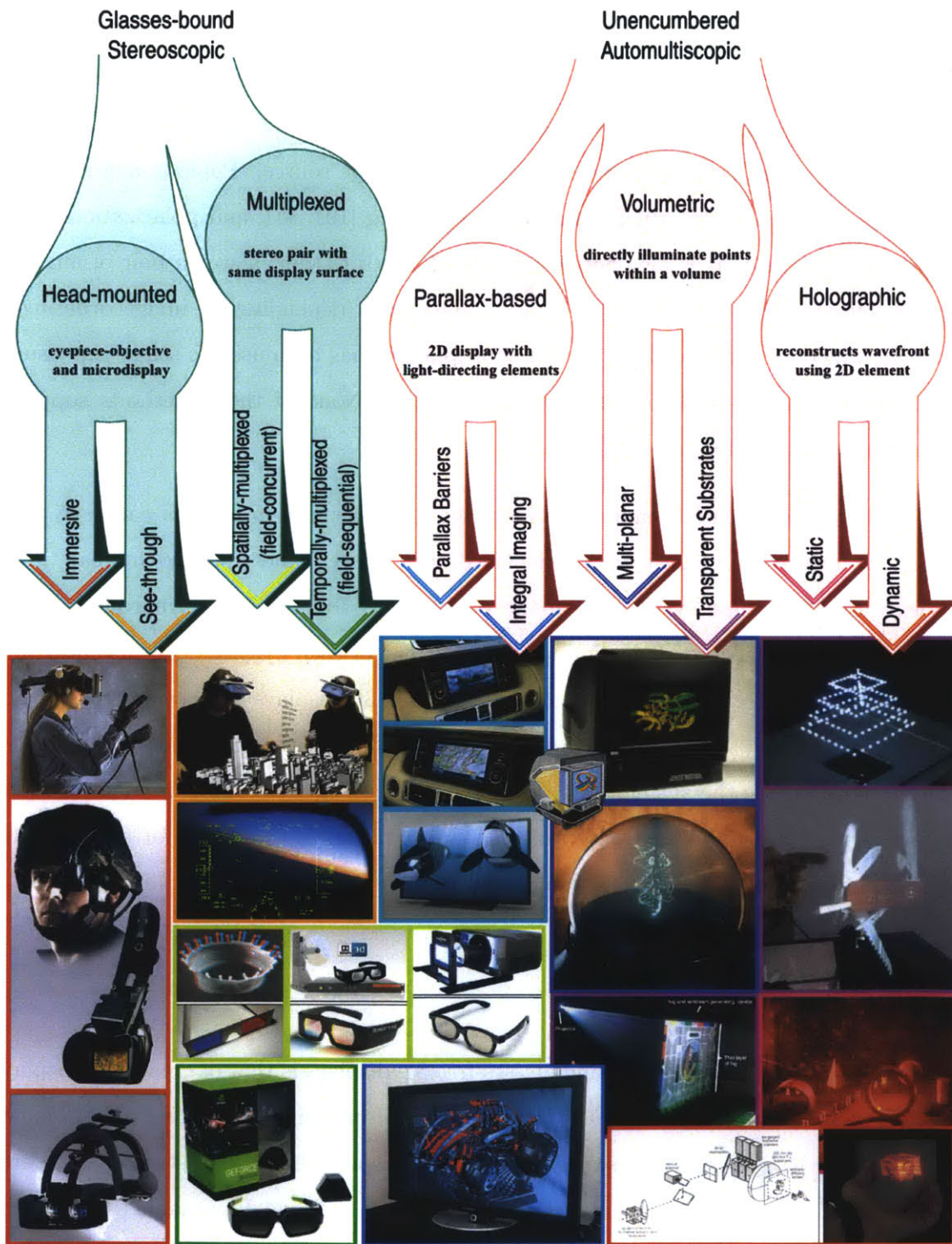


Figure 2-6: A taxonomy of display systems capable of creating physiological depth cues. The systems are detailed in the sections below.

two eyes can add the depth cues of binocular disparity and convergence, if done correctly. Many methods to deliver a pair of images to the eyes have been developed, most of which rely on a head-worn apparatus, such as glasses that include optical filters. Stereo viewers like the stereoscope and View-Master [207, 68] have mostly given way to these glasses-based approaches that rely on a remote multiplexed image source. Popular methods are wavelength multiplexing [15, 101], temporal multiplexing [191, 41], and polarization state multiplexing [218, 129]. It is also possible to train oneself to view a stereo pair of aligned right and left images by crossing or diverging one's eyes (depending on image ordering). This effect, which can be considered spatial multiplexing, has been used to study the visual system, in the case of random dot stereograms [102]. None of these methods support accommodation, or motion parallax.

Virtual reality or augmented reality systems place small displays in front of each eye [29], which allows arbitrary imagery to be displayed to a viewer, and updated in real time. With head tracking these systems have the potential to add the motion parallax depth cue. Placing a volumetric or light field display in front of each eyepiece can also add the accommodation depth cue [139]. However, glasses-based 3D displays will always suffer from the problem of immediacy. Users have to think of putting on the glasses or other device when they want to use the system. Such displays also cannot take a measurement of light transport at the location where they project virtual imagery, so they cannot support the light based interaction methods explored in this thesis.

2.3.2 Glasses-free Displays

Fundamentally, the problem of creating glasses-free 3D displays is that of creating a display that can control the spatial *and angular* variation of light intensity, as different images must be steered into the eyes of viewers. Glasses-free 3D displays have been studied for more than a century, with early works including those of Ives [93] and Lippmann [128]. Such methods trade spatial resolution for angular resolution, which creates a trade-off between “pop-out” effect, and blurriness of each view. With advances in computation and display technology, researchers have integrated viewer tracking [156, 159], image compression [137],

electronically-switchable displays [96], and temporal multiplexing [111] to overcome some limitations of glasses-free displays. With the increasing area and density of displays, limited success has been achieved in simply scaling up lenticular designs [6, 131]. However, these designs suffer from basic data bandwidth problems, as light field data is very redundant in the space and angle bases. In the course of this thesis work we have shown that by exploiting temporal multiplexing with multi-layer displays, in combination with low-rank matrix or tensor factorizations, these architectures can be optimized in terms of image fidelity, brightness, and frame rate [119, 205, 121, 206, 134, 82]. We have shown that including additional optical components, such as directional backlighting, can further improve display quality by adding additional degrees of freedom.

Volumetric Displays Another common glasses-free display paradigm is represented by volumetric displays [54] and stacks of light-emitting, rather than light-attenuating, layers [5]. Volumetric devices usually require mechanically moving parts [43, 99, 182] or time-multiplexed diffusers [179]. The majority of such volumetric displays can only depict 3D content that is confined within the physical device enclosure, excluding the light field displays proposed by Cossairt et al. and Jones et al. In contrast to the additive image formation model inherent to most volumetric displays, the optical designs examined in this thesis exploit multiplicative attenuation of light to allow synthesized 3D objects to extend outside the enclosure of the display. Further, the class of displays based on multiple layers of attenuators support specularities, occlusions, and global illumination effects, without requiring moving parts, or encumbering the user.

Directional Backlighting Directional backlights are an emerging trend in display technology. The combination of a fast-switching LCD and a rear-illuminating light guide allows stereoscopic [187, 186, 38, 36, 26] and multiscopic image synthesis [136, 188]. Stolle et al. [178] and Kwon and Choi [115] implement multidirectional backlighting using lenslet arrays.

Holography Holographic [17, 113] or hybrid holographic [3, 166] displays exploit the wave nature of light to create directional variation. Holography in general produces exquisite still images, but fabrication and bandwidth have proved challenging when attempting to create moving imagery. Holographic displays share the bandwidth limitations of directly sampled light field displays. In addition, it is difficult to create photo-polymers with sufficient temporal resolution [160], while direct electro-optical solutions lack spatial resolution, requiring optically tiling many devices to achieve moderate display resolution [211]. Computing full holographic fringe patterns for high spatio-temporal resolution devices is a challenging computational problem, requiring clusters of computers and GPU processors [155]. Devices capable of creating holographic stereograms with horizontal parallax at high spatio-temporal bandwidth have emerged recently [174], and are a promising area of research. In the thesis, however, we are primarily concerned with devices that can both exploit near-term or currently available electro-optical devices to go beyond directly-sampled light fields to simultaneous compressive light field display and capture.

Supporting Accommodation Displays supporting correct accommodation are able to create a light field with enough angular resolution to allow subtle, yet crucial, variation over the pupil. Such displays utilize three main approaches. Ultra-high angular resolution displays, such as super multiview displays [180, 181, 154], take a brute-force approach: all possible views are generated and displayed simultaneously, incurring high hardware costs. In practice, these drawbacks have limited the size, field of view, and spatial resolution of the devices. Multi-focal displays [5, 83, 170], virtually place conventional monitors at different depths via refractive optics. This approach is effective, but requires encumbering glasses. Volumetric displays [54] physically generate light rays at the perceived 3D position, but are limited to small volumes and cannot reproduce occlusion. Closely related light field displays with anisotropic diffusion surfaces [99, 43] can reproduce small volumes with occlusion, but accommodation has been demonstrated in the horizontal dimension only within a limited depth range [99].

Large Scale Projection Large-scale autostereoscopic and multiscopic projection systems have been actively investigated throughout the last century. Most of the proposed systems are variants of integral imaging or parallax barriers; by combining active projection and large barrier screens, theater-sized installations have been built in France and Russia starting in the 1940s [59]. Today, barrier-type light field projection systems are still an active area of research (e.g., [213, 110]). The fundamental limitation of these displays, as every other integral imaging or barrier-based method, is the loss of spatial resolution. 2D/3D switchable solutions for projectors have been proposed [88] but require multiple devices.

Resolution limits can be overcome using multiple projectors combined with front or rear-projected lenticular screens [137, 90] or unidirectional diffusers [12, 98]. In these systems, the number of devices roughly matches the number of viewing zones. Projector arrays can also be directly observed [103], but require one device per pixel. Dodgson et al. [49] investigate multi-projector devices combined with time-multiplexed image synthesis for 3D display. Compared to existing multi-device solutions, we present a new optical configuration that is well-suited for compressive light field synthesis with a single device.

Single-device configurations have been explored [148, 42, 141, 22]. These methods project individual viewing zones sequentially and at a high-speed onto special screens. Unfortunately, these screens require mechanically-moving parts that translate in unison with the high-speed projection. We present a compressive light field projection system that requires only a single device (it can be enhanced using a few additional devices), operates at the full display resolution, and does not require active components in the screen. The proposed display combines a novel screen design based on Keplerian angle expansion with high-speed light field projection and compressive factorizations. In addition, we show applications to 2D superresolution and high dynamic range projection.

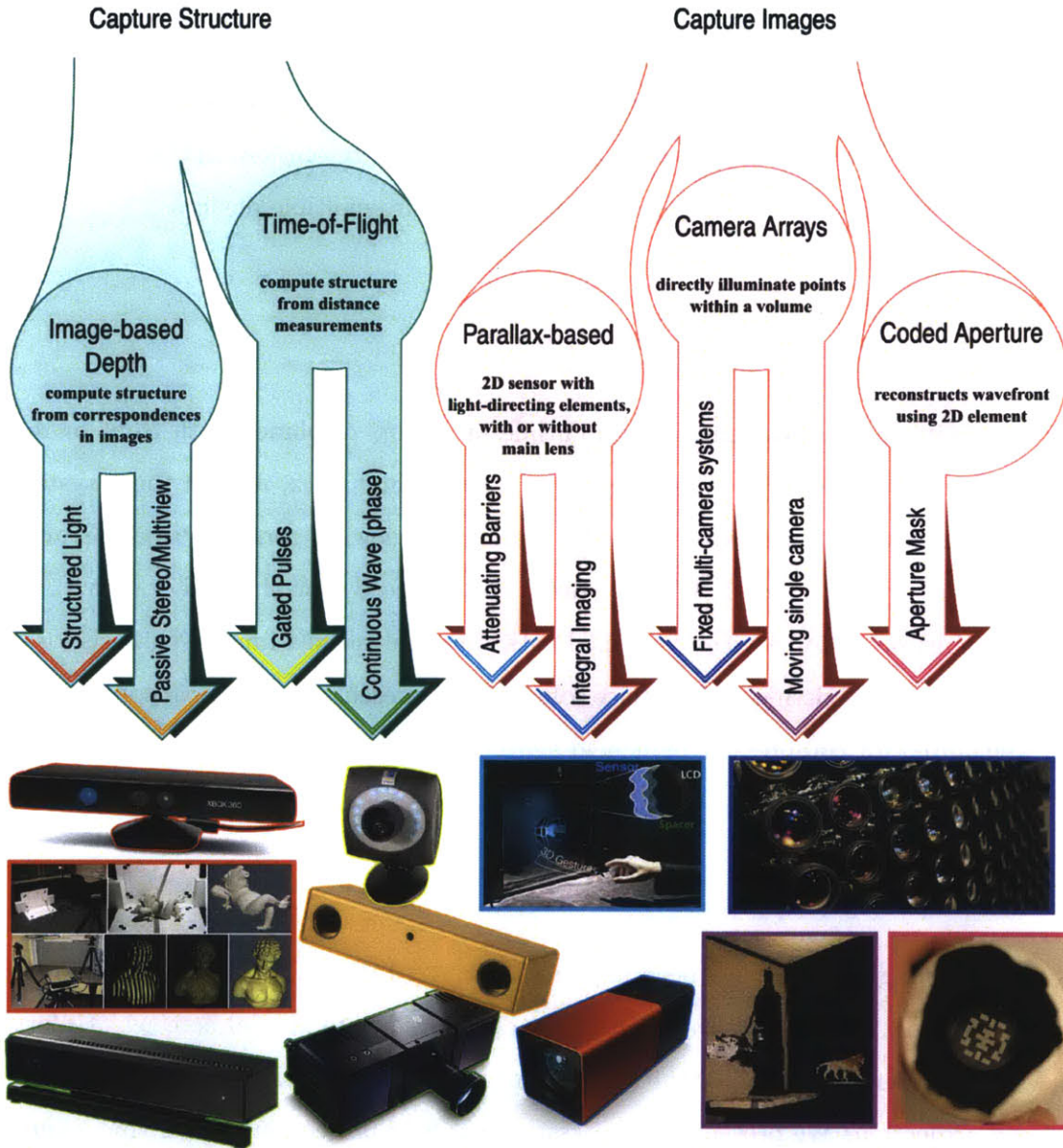


Figure 2-7: A taxonomy of camera systems capable of capturing imagery suitable for advanced display systems. The systems are detailed in the sections below.

2.4 Taxonomy of Camera Systems

2.4.1 Light Field Cameras

Light field cameras were invented more than a century ago. Early prototypes either used a microlens array [128] or a light-blocking mask [93] to multiplex the rays of a 4D light field onto a 2D sensor. In the last decades, significant improvements have been made to these basic designs, i.e. microlens-based systems have become digital [2, 147] and mask patterns more light efficient [194, 120]. However, the achievable resolution is fundamentally limited by the spatio-angular resolution tradeoff: spatial image resolution is sacrificed for capturing angular information with a single sensor. Detailed discussions of this topic can be found in the literature (e.g., [125, 204]).

Two common approaches seek to overcome this tradeoff: using camera arrays [209, 196] or capturing multiple images of the scene from different perspectives [126, 63, 127]. However, camera arrays are usually bulky and expensive whereas multi-shot approaches restrict photographed scenes to be static. It is also possible to combine a regular camera and a microlens-based light field camera [130]; again, multiple devices are necessary. In this paper, we present a new camera architecture that uses a single device to recover both a conventional 2D image and a high-resolution 4D light field from a single image.

2.4.2 Depth Cameras

Though this thesis is focused on image-based methods for scene capture other methods exist to extract explicit geometric representations. Broadly known as depth cameras, such systems seek to reconstruct a point cloud representing measured surface boundaries. These systems can broadly be separated into two categories: stereoscopic and structured light systems, and time-of-flight systems. Explicit geometry information can be adapted to be displayed on a light field display, though issues such as hole filling [149] arise, and are not always easily resolved.

Stereoscopic and Structured Light Methods In a method closely inspired by human vision, binocular, trinocular [192] and multi-view stereo [72] systems capture images of a scene from multiple offset cameras, yielding multiple perspectives of the scene. Correspondences are identified in the set of captured images, and triangulate depth from the disparity between correspondence points on the image plane of each camera. The problem of matching correspondence points is still a challenging research area [167]. As such the performance of stereoscopic systems is scene dependent.

It is also possible to exploit the perspective variation across the aperture of a camera to extract depth information. Techniques such as shape from focus [146] exploit this effect as it manifests through depth of field in a standard large aperture camera system. Another approach is to modify the aperture of the camera with an amplitude [124, 53] or phase [66] mask to create an invertible point spread function (PSF), capable of preserving scene depth information.

Structured Light In order to obtain robust measurements in scenes with a wide variety of texture patterns and surface structure, active illumination can be employed to optically paint correspondence patterns onto a scene. A review of methods can be found in [19]. Commercial products such as the Microsoft Kinect have used this technique to produce reliable depth maps for computer interaction and entertainment. LIDAR systems, which use a laser to provide active illumination, are commonly used in industrial settings—for example on the Google Street View cars, which have mapped many of the world’s roads.

Time-of-flight and time-resolved methods probe a scene with ultra-fast light pulses and recover the returned pulses from the scene with ultra-fast sensors. Such systems are revolutionizing the fields of computer vision and computer graphics. Commercial products such as the Microsoft Kinect 2, Canesta, and the PMD cameras employ a phase-based technique to determine the delay in a reflected continuous wave optical signal due to scene depth. It is even possible to resolve non-line-of-sight geometry with such systems [195, 105].

2.5 Combined Systems

A promising result from this thesis comes from considering display systems that collocate light sensitive and light emissive elements. Though these types of combinations are still uncommon, a few other examples can be found in the areas of optical computing, and user interaction.

Glasses-free 3D parallax barrier [93] and lens-array [128] displays have existed for over 100 years. Nayar et al. [145] create a lighting sensitive display, though it cannot accurately map shadows and specularities. BRDF displays can simulate flat surfaces with a particular Bi-Directional Reflectance Distribution Function [91]. 6D displays that demonstrate 4D relighting of 2D images have been shown in both active [81] and passive [58] modes. A recently shown 7D display [184] tracks a single light point as input. In a closely related work, Cossairt et al. [44] implement a 7fps 8D display, but focus on rendering illumination effects for a 2D camera, rather than 3D perception for a live viewer. Our work contributes a hardware approach to real-time 8D display that is compatible with emerging display technologies and a new GPU rendering and capture pipeline to make simultaneous, interactive 4D lighting and 4D capture feasible.

Combinations of camera and projector systems have also been used for optical computing purposes to probe the light transport matrix of an arbitrary scene [152]. This allows applications such as transposing the position of the camera and projector [168], allowing a photo to be taken from the location of a projector. It has further been shown that, using optical coding, operations can be performed on the light transport matrix without first capturing it. This allows, for example, imaging only the direct or scattered component of a scene [153].

2.6 Beyond Dirac Representations

Displays It is well-understood that light fields of natural scenes contain a significant amount of redundancy. Most objects are diffuse; a textured plane at some depth, for

instance, will appear in all views of a captured light field, albeit at slightly different positions. This information can be fused using super-resolution techniques, which compute a high-resolution image from multiple subpixel-shifted, low-resolution images [183, 169, 18, 132, 158, 196, 28, 200].

We show throughout Chapter 3 that through casting the problem of light field display as a matrix approximation problem it is possible to exploit these redundancies. For a two layer system, we pose the problem as

$$\mathbf{L} = \mathbf{FG} \tag{2.1}$$

where \mathbf{L} is the light field to be displayed, and \mathbf{F} and \mathbf{G} are, respectively $N \times T$ and $T \times N$ sized matrices, representing the T time multiplexed patterns displayed on each display layer. Because incoherent light is non-negative, this problem can be solved using non-negative matrix factorization, posed as

$$\arg \min_{\mathbf{F}, \mathbf{G}} \frac{1}{2} \|\mathbf{L} - \mathbf{FG}\|_{\mathbf{W}}^2, \text{ for } \mathbf{F}, \mathbf{G} \geq 0, \tag{2.2}$$

Cameras With the discovery of compressed sensing [32, 50], a new generation of compressive light field camera architectures is emerging that goes far beyond the improvements offered by super-resolution. For example, the spatio-angular resolution tradeoff in single-device light field cameras [9, 10, 212, 135] can be overcome or the number of required cameras in arrays reduced [108]. Compressive approaches rely on increased computational processing with sparsity priors to provide higher image resolutions than otherwise possible.

The problem typically called compressed sensing, and first explored by Candès et al. [32] is posed as follows:

$$\arg \min_{\tilde{x} \in \mathbb{R}^N} \|\tilde{x}\|_{\ell_1} \text{ subject to } \Phi \tilde{x} = y \tag{2.3}$$

Here, Φ represents the measurement matrix for the system. However, as described in more detail in Chapter 4, we have found that a sparsity constrained least squares approach, where the ℓ_1 term is taken as a sparse regularizer, has yielded good results when applied to the problem of compressed light field recovery. Posed as follows the problem is known as Basis Pursuit De-Noise (BPDN) [35]:

$$\arg \min_{\tilde{x} \in \mathbb{R}^N} \frac{1}{2} \|y - \mathbf{A}\tilde{x}\|_{\ell_2}^2 + \lambda \|\tilde{x}\|_{\ell_1} \quad (2.4)$$

2.7 Interactive Techniques

Light pens and widgets have been previously used for interaction [8]. In recent years, lighting widgets have been integrated into tabletop computing systems [123], and novel optics and computer vision have been used for interaction with screens, tables, and physical surfaces over a screen [95, 165, 57]. Augmented reality gaming on systems such as the Nintendo 3DS and PlayStation Vita has created entertaining user experiences by rendering on top of stereoscopic images captured live from a scene. Mixing rendered and live captured stereoscopic content on a hand held gaming device is closely related to the concept of an 8D Display, introduced in Chapter 5. Though the concept is less general it provides an insight into some types of applications that will become available with widespread general purpose light transport devices.

Tompkin et al. [184] demonstrated an application known as light field painting, which comprises a general light field display device multiplexed with an optical system designed to track a single point at the tip of a light pen. This system enables the user to draw on a glasses free 3D display using an optical input. The hardware implemented is nearly identical to that of the classical 8D Display prototype presented in Chapter 5, though it is used as a “7D Display”, in that on the input side only a point (x, y, z) in 3D space is captured.

Chapter 3

Compressive Methods for Visual Display

In this chapter we consider the problem of 4D light field display. A subset of this problem—often called glasses-free 3D, or automultiscopic display—is solved for 3D, horizontal-only light fields, which contain no variation in the vertical angle dimension (see Section 2.3). The methods developed in this chapter apply to this 3D problem as well as the 4D problem, but for most applications it is most instructive to consider the full 4D problem.

In Section 3.1 we lay out the requirements for compressive display systems, in contrast to conventional, Dirac sampled systems (see Section 2.6 for background in this area). We begin with the intuition that, when navigating the world, the observations we make with our eyes obey relatively predictable models and generally vary smoothly with small changes in perspective. This implies that 4D light transport is in some way encoding a great deal of redundant data. Though the task of recreating light transport at the fidelity of the real world is a daunting one, such an observation should lead us to suspect that it may be possible to create a *compressive display*, capable of exploiting the redundancy inherent in light transport to reduce the number of degrees of freedom required to represent high fidelity light field scenes. In this chapter we show multiple approaches to realizing this goal by exploiting an abundant resource: computation. The framework developed over the

course of this chapter is flexible, in the sense that it is capable of efficiently mapping light fields into the available degrees of freedom of a display system in many cases.

In the following sections, we describe methods for adding degrees of freedom to display systems, spatially (Section 3.2), and temporally (Section 3.3). Section 3.4 presents the Tensor Display Framework, a general framework incorporating spatiotemporal degrees of freedom, that additionally allows the analysis of systems that create angular variation through refractive optics. Finally, Chapter 6 looks at applications of the Tensor Display Framework to challenging problems in high-dimensional display: Creating accommodation effects on a TV sized screen (Section 6.1) and creating a glasses-free 3D display system for theater-scale applications.

3.1 Requirements for a Compressive Display System

The framework developed over the course of this chapter represents a flexible and general method for driving advanced display systems capable of manipulating light intensity over a region of space and angle. However, in order to apply the framework to a given data-set on a given display device it is necessary to satisfy the following basic requirements.

Structured Data The light field imagery to be displayed must be compressible—for the purposes of the Tensor Display Framework (Section 3.4), the data must be low rank as parameterized by the geometry of the display device.

N-to-M map An output light ray intensity must be a function of many display elements (e.g. pixels), and each display element must influence multiple ray intensities. A one-to-one mapping, such as the Dirac representation found in most light field displays to date, does not allow for efficient exploitation of data redundancy.

Non-linearity The interaction between display output elements and ray intensities must contain a non-linearity in order for the display to represent discontinuous effects in the light field such as specularities and occlusion. In the case of a display comprising

two multiplicative layers—a simple Tensor Display—the non-linearity is imposed by the multiplicative, bilinear form of the interaction between display elements driving the intensity of each ray. In this example the intensity of ray \mathbf{L}_{ij} in the output light field is determined by the modulation value of pixels ξ_i , and ξ_j as $\mathbf{L}_{ij} = \xi_i \xi_j$.

Suitable Optical Basis The display hardware defines an optical basis, or parameterization of the light field, in which the data to be represented by the display must be well represented. If this condition is not satisfied, the data will not be sufficiently redundant to meet the first requirement. This requirement is closely related to the *Structured Data* requirement. Taken together, the requirements impose both that there exists a parameterization of the light field to be represented in which the representation is low rank, and that the display hardware represents such a parameterization.

3.2 Optically Efficient Methods

Before diving into the Tensor Display Framework (Section 3.4) we will use this section to build intuition about the performance of advanced display systems that derive all degrees-of-freedom from the spatial distribution of modulating layers. Here we recap an intuitive tomographic synthesis approach to light field display (Section 3.2.1), give an overview of Liquid Crystal Displays (LCDs), and show how the tomographic synthesis method can be adapted to LCDs in an optically efficient manner (Section 3.2.2).

3.2.1 Tomographic Synthesis

Wetzstein et al. [205] showed that multi-layer devices, comprised of two or more attenuating layers, can represent light fields compressively using a tomographic reconstruction technique to embed a light field into the layers of the display device. In that work, the authors describe the forward image synthesis for a back-lit attenuating volume, and show that it is equivalent to the problem of computed tomography [78]—the light field emitted from an attenuation volume is equal to the negative Radon transform of the attenuation map. We



Figure 3-1: Dynamic light field display using polarization field synthesis with multi-layered LCDs. (Left) We construct an optically-efficient polarization field display by covering a stack of liquid crystal panels with crossed linear polarizers. Each layer functions as a polarization rotator, rather than as a conventional optical attenuator. (Right, Top) A target light field. (Right, Bottom) Light fields are displayed, at interactive refresh rates, by tomographically solving for the optimal rotations to be applied at each layer. (Middle) A pair of simulated views is compared to corresponding photographs of the prototype on the left and right, respectively. Inset regions denote the relative position with respect to the display layers, shown as black lines, demonstrating objects can extend beyond the display surface.

recap that work here, and show that iterative back-projection algorithms such as SART [7, 107] are applicable to solving the tomographic light field problem, even when the attenuation volume is discretized into thin layers. Fast methods such as SART make such light field decompositions suitable for real-time display applications. The following sections comprise a brief summary of Section 3 of Wetzstein et al.[205]. A more complete discussion can be obtained from that paper.

Volumetric Attenuation

The intensity I of a light ray \mathcal{C} through an attenuation map $\mu(x, y)$ is governed by the Beer-Lambert law

$$I = I_0 e^{-\int_{\mathcal{C}} \mu(r) dr} \quad (3.1)$$

where I_0 is the incident intensity [73].

In Section 3.3 we show that, when using a relative two-plane light field parameterization [33], the light field $l(u, a)$ emitted when a volumetric attenuator is illuminated by a backlight

producing the incident light field $l_0(u, a)$ is given by:

$$\bar{l}(u, a) = \ln \left(\frac{l(u, a)}{l_0(u, a)} \right) = -p(u, a) \quad (3.2)$$

where \bar{l} denotes a normalized logarithm, and $p(u, a)$ denotes the Radon transform of $\mu(x, y)$.

Light Field Synthesis

An estimate of the attenuation map $\bar{u}(x, y)$ is recovered from $p(u, a)$ using the inverse Radon transform. However, the traditional method of filtered backprojection does not yield stable reconstructions with limited view angles and does not provide positive only solutions required for display fabrication from attenuating layers.

The authors adopt an iterative reconstruction method that better accounts for inconsistent projections over limited angles. They employ a series expansion method for which attenuation is modeled by a linear combination of N_b non-negative basis functions $\bar{\phi}_k(x, y)$:

$$\mu(x, y) = \sum_{k=1}^{N_b} \alpha_k \bar{\phi}_k(x, y) \quad (3.3)$$

In the case of Layered 3D the authors chose to use a series expansion into a set of normalized linear basis functions rather than a discrete voxel representation for $\bar{\phi}_k$. It is shown that when considering a discrete light field \bar{l}_{ij} the above choice of reconstruction algorithm leads to a linear system of equations such that

$$\bar{l}_{ij} = - \sum_{k=1}^{N_b} \alpha_k P_{ij}^{(k)} \quad (3.4)$$

where (i, j) are the discrete indices corresponding to the continuous coordinates (u, a) . The structure of the projection matrix $P_{ij}^{(k)}$ is given in Wetzstein et al. [205]. The system can be expressed in matrix-vector form as $\mathbf{P}\alpha = -\bar{\mathbf{l}} + \bar{\mathbf{e}}$, where $\bar{\mathbf{e}}$ is the approximation error. The

attenuation map synthesis can be cast as the following non-negative linear least squares problem:

$$\arg \min_{\alpha} \|\bar{\mathbf{l}} + \bar{\mathbf{P}}\alpha\|^2, \text{ for } \alpha \geq 0 \quad (3.5)$$

This formulation as a convex optimization problem yields an optimal attenuation map, in the least-squares sense, that emits a target light field with consistent views.

Layered Attenuators

So far we have considered a continuously varying attenuation volume. However, we are interested in modern display hardware such as LCD panels that are better represented by a finite number of discrete attenuation layers.

Wetzstein et al. [205] extend their analysis to such multi-layered attenuators. They show that, analogously to Equation 3.2, a ray (u, a) is modulated by N_l layers such that

$$l(u, a) = l_0(u, a) \prod_{k=1}^{N_l} t_k(u + (d_k/d_r)a) \quad (3.6)$$

where $t_k(\xi)$ is the transmittance of mask k (separated by a distance d_k). Taking the logarithm gives the forward model

$$\bar{l}(u, a) = - \sum_{k=1}^{N_l} a_k(u + (d_k/d_r)a) \quad (3.7)$$

where $a_k(\xi) = -\ln t_k(\xi)$ is the absorbance. Analogously to Equation 3.4, the discretized linear system is $\bar{l} = -\sum_{k=1}^{N_l} a_k P_{ij}^{(k)}$ with discretized rays (i, j) . The projection matrix $P_{ij}^{(k)}$ is modified to encode the intersection of every ray with each attenuating layer. Because in practice layers have a finite contrast, Equation 3.5 is solved as a *constrained* least-squares problem.

3.2.2 Liquid Crystal Displays

Of course, in Section 3.2 we would like to formulate the problem of light field decomposition such that it can be implemented efficiently, both optically and computationally, on a stack of LCD panels. As a pixel on an LCD panel can, under some circumstances, be considered a voltage-controlled linear polarization state rotator, we observe in this section that it is possible to create a stack of linear polarization state rotating layers, rather than a stack of attenuating layers.

Constructing polarization field displays requires an accurate characterization of the optical properties of LCDs. The transformation of polarized light due to passage through layered materials is modeled by the Jones calculus [100]. Orthogonal components of the electric field are represented as a complex-valued Jones vector. The optical action of a given element (e.g., a birefringent layer or polarizing film) is represented by a Jones matrix, with the product of this matrix and a Jones vector encoding the polarization state transformation. Yeh and Gu [214] formally characterize the polarization properties of LCDs, providing analytic Jones matrices for common technologies, including twisted nematic (TN), vertical alignment (VA), and in-plane switching (IPS) panels. In this paper we consider a unifying, but simplified, Jones matrix model, wherein LCDs are approximated as spatially-controllable polarization rotators.

Applying a more detailed Jones matrix model for our modified off-the-shelf panels has the potential to reduce visible artifacts in the prototype, possibly at the cost of decreased refresh rates due to increased computational complexity (see Section 3.2.6). Moreno et al. [143] estimate the Jones matrix of an LCD using seven irradiance measurements, two linear polarizers, and a single quarter-wave plate. Ma et al. [133] propose a simplified calibration using only three measurements. A promising alternative to these model-based refinements is to directly engineer LCD panels to act as polarization rotators. Davis et al. [48] implement such panels using a custom parallel-aligned LCD covered by a pair of crossed quarter-wave plates. Moreno et al. [142] construct a polarization rotator using a conventional TN panel. In both works, the liquid crystal is operated as a voltage-controlled

wave plate to produce polarization state rotations. Layered constructions of such panels are ideally suited to implement practical polarization field displays.

A liquid crystal display (LCD) contains two primary components: a backlight and a spatial light modulator (SLM). The backlight is designed to produce uniform illumination, typically by conditioning the light produced by a cold cathode fluorescent lamp (CCFL) or a light-emitting diode (LED) array, using a light guide and various diffusing and brightness-enhancing films. The spatial light modulator is a thin layer of liquid crystal, enclosed between glass sheets with embedded, two-dimensional electrode arrays. This stack is further enclosed by a pair of crossed linear polarizers.

Applying a voltage across an electrode pair alters the polarization properties of a pixel. We assume the effect can be approximated as inducing a rotation of the polarization state of light rays traversing the pixel. This holds to varying degrees of accuracy for off-the-shelf LCDs (see Section 3.2.6). Yet, following Davis et al. [48] and Moreno et al. [142], such polarization rotators can be constructed by modifying existing LCDs. Under this model the transmitted intensity I is given by Malus' law:

$$I = I_0 \sin^2(\theta), \tag{3.8}$$

where I_0 is the intensity after passing through the first polarizer and θ is the angle of polarization after passing through the liquid crystal, defined relative to the axis of the first polarizer [73]. By controlling the voltages applied across the electrode array, two-dimensional images are rendered with varying shades of gray depending on the induced rotation. The rotation angle θ must vary only over the interval $[0, \pi/2]$ radians to reproduce all shades of gray—the range afforded by most commercial LCD panels, including widespread twisted nematic (TN) architectures. We note that this model only strictly applies for rays oriented perpendicular to the display surface. At oblique angles, light leakage occurs through crossed polarizers and birefringence of the liquid crystal produces elliptical, rather than linear, polarization states [214]. However, as experimentally verified in Section 3.2.6, this model is sufficient for the viewing angles considered in the prototype.

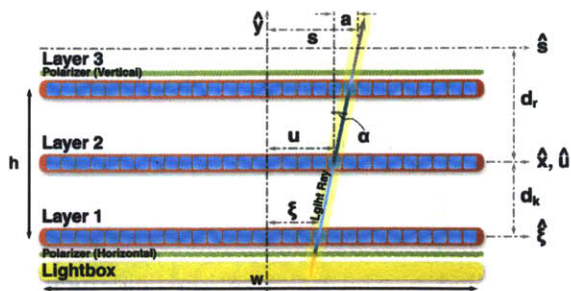


Figure 3-2: Polarization field displays. A K -layer display is constructed by separating multiple liquid crystal panels. The light field $l_0(u, a)$ emitted by the backlight is linearly polarized by the rear polarizer. The polarization state of ray (u, a) is rotated by $\phi_k(\xi)$ after passage through layer k , where $\xi = u + (d_k/d_r)a$. The emitted light field $\tilde{l}(u, a)$ is given by applying Equation 3.9 to the emitted polarization field $\tilde{\theta}(u, a)$ upon passage through the front polarizer.

Two design alternatives enable color LCDs: color filter arrays and field sequential color. In current LCDs, a color filter array is deposited on the glass sheet closest to the viewer. Each pixel is divided into three subpixels by an array of filters with spectral transmittances corresponding to three color primaries. This requires the resolution to be tripled along one display axis, increasing fabrication complexity and cost. Color filter arrays also decrease brightness, typically to 30% of the backlight intensity. Rather than brightening the backlight, which reduces power efficiency, field sequential color (FSC) can be employed. With FSC, a strobed backlight successively illuminates a high-speed monochromatic LCD with varying color sources. If strobing occurs faster than the human flicker fusion threshold [71], a color image is perceived. While yet to be widely commercially available, FSC LCDs are an active area of research [177, 34].

3.2.3 Modeling Multi-Layer LCDs

In this section we consider how multi-layer LCDs can be constructed to emit a four-dimensional light field, rather than a two-dimensional image. As shown in Figure 3-4, we consider the following architecture: a backlight covered by multiple, disjoint spatial light modulators. First, to maximize the optical efficiency, we assume field sequential color illumination; this eliminates K layers of color filters that would otherwise cause severe moiré [16] and brightness attenuation by a factor of approximately 0.3^K (e.g., 2.7% transmission for

a three-layer LCD). Second, we observe that only two polarizing films are necessary, one on the top and bottom of the multi-layer stack. This creates a *polarization field display*, wherein each spatial light modulator consists of a liquid crystal layer functioning as a spatially-addressable, voltage-controlled polarization rotator.

Such displays must be controlled so the polarization field incident on the last polarizer accurately reproduces the target light field. In this section we present our analysis in flatland, considering 1D layers and 2D light fields, with a direct extension to 2D layers and 4D light fields. As shown in Figure 3-2, we consider a display of width w and height h , with K layers distributed along the y -axis such that $d_k \in [-h/2, h/2]$. A two-plane light field parameterization $l(u, a)$ is used [33]. The u -axis is coincident with the x -axis and the slope of ray (u, a) is defined as $a = s - u = d_r \tan(\alpha)$, where the s -axis is a distance d_r from the u -axis.

The emitted light field $l(u, a)$ is given by applying Equation 3.8 to the polarization field $\theta(u, a)$ incident on the front polarizer:

$$l(u, a) = l_0(u, a) \sin^2(\theta(u, a)), \quad (3.9)$$

where $l_0(u, a)$ is the light field produced by the backlight after attenuation by the rear polarizer. The backlight is assumed to be uniform such that $l_0(u, a) = l_{max}$ and the light field is normalized such that $l(u, a) \in [0, l_{max}]$. This expression is used to solve for the necessary target polarization field $\theta(u, a)$, as follows.

$$\theta(u, a) = \pm \sin^{-1} \left(\sqrt{\frac{l(u, a)}{l_0(u, a)}} \right) \bmod \pi \quad (3.10)$$

Under these assumptions, the principal value of the arcsine ranges over $[0, \pi/2]$. Note, with full generality, the target polarization field is multi-valued and periodic, since a rotation of $\pm\theta \bmod \pi$ radians will produce an identical intensity by application of Malus' law.

Each layer controls the spatially-varying polarization state rotation $\phi_k(\xi)$, as induced at point ξ along layer k . Ray (u, a) intersects the K layers, accumulating incremental rotations

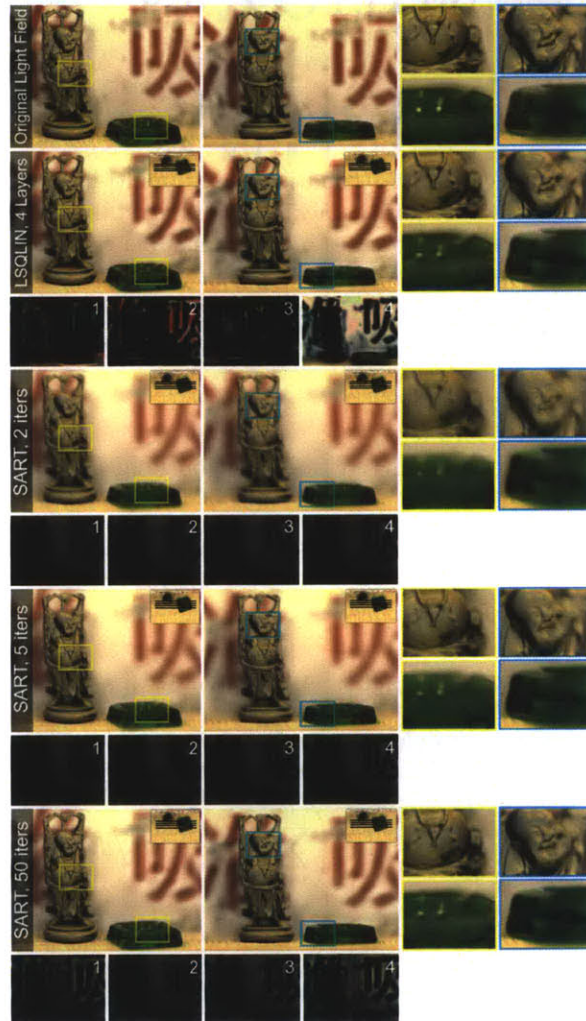


Figure 3-3: GPU-based SART allows real-time multi-layer optimization approaching the fidelity of the off-line solver. The first and second columns show different target views. Polarization-rotating layers are shown below each example. The off-line reference solver [40] produces sharp reconstructions (second row). A small number of SART iterations causes blurring (third row). Additional iterations converge to the reference (bottom row), with five iterations yielding similar quality (fourth row). Note that simulated views are shown, rather than prototype results.

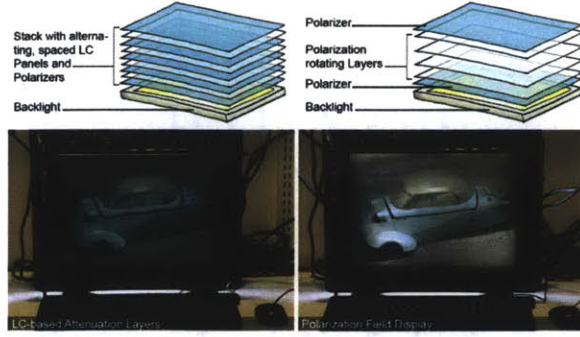


Figure 3-4: Polarization-based vs. attenuation-based multi-layer LCDs. (Top, Left) An attenuation-based light field display requires stacking liquid crystal panels with polarizers between each layer. This construction effectively creates a programmable transparency stack. (Top, Right) Polarization-based light field displays improve optical efficiency using a single pair of crossed polarizers. (Bottom) Corresponding photographs of the prototype configured as an attenuation-based vs. polarization-based multi-layer LCD.

at each intersection, such that the emitted polarization field $\tilde{\theta}(u, a)$ is given by

$$\tilde{\theta}(u, a) = \sum_{k=1}^K \phi_k(u + (d_k/d_r)a). \quad (3.11)$$

Combining Equations 3.9 and 3.11 yields the following model for the light field $\tilde{l}(u, a)$ emitted by a K -layer polarization field display:

$$\tilde{l}(u, a) = l_0(u, a) \sin^2 \left(\sum_{k=1}^K \phi_k(u + (d_k/d_r)a) \right). \quad (3.12)$$

3.2.4 Synthesizing Polarization Fields

This section describes the optimization of multi-layer LCDs for polarization field display. We consider a discrete parameterization for which the emitted polarization field is represented as a column vector $\tilde{\theta}$ with M elements, each of which corresponds to the angle of polarization for a specific light field ray. Similarly, the polarization state rotations are represented as a column vector ϕ with N elements, each of which corresponds to a specific display pixel in

a given layer. Under this parameterization, Equation 3.11 yields a linear model such that

$$\tilde{\theta}_m = \sum_{n=1}^N P_{mn} \phi_n, \quad (3.13)$$

where $\tilde{\theta}_m$ and ϕ_n denote ray m and pixel n of $\tilde{\theta}$ and ϕ , respectively. An element P_{mn} of the projection matrix \mathbf{P} is given by the normalized area of overlap between pixel n and ray m , occupying a finite region determined by the sample spacing.

An optimal set of polarization state rotations ϕ is found by solving the following constrained linear least-squares problem:

$$\arg \min_{\phi} \|\theta - \mathbf{P}\phi\|^2, \text{ for } \phi_{min} \leq \phi \leq \phi_{max}, \quad (3.14)$$

where each layer can apply a rotation ranging over $[\phi_{min}, \phi_{max}]$. Similar to Wetzstein et al. [205], Equation 3.14 can be solved using a sparse, constrained, large-scale trust region method [40]. However, we observe that this problem can be solved more efficiently by adapting the simultaneous algebraic reconstruction technique (SART). As proposed by Andersen and Kak [7] and further described by Kak and Slaney [107], SART provides an iterative solution wherein the estimate $\phi^{(q)}$ at iteration q is given by

$$\phi^{(q)} = \phi^{(q-1)} + \mathbf{v} \circ (\mathbf{P}^T(\mathbf{w} \circ (\theta - \mathbf{P}\phi^{(q-1)}))), \quad (3.15)$$

where \circ denotes the Hadamard product for element-wise multiplication and elements of the \mathbf{w} and \mathbf{v} vectors are given by

$$w_m = \frac{1}{\sum_{n=1}^N P_{mn}} \quad \text{and} \quad v_n = \frac{1}{\sum_{m=1}^M P_{mn}}. \quad (3.16)$$

After each iteration, additional constraints on $\phi^{(q)}$ are enforced by clamping the result to the feasible rotation range. Building upon the Kaczmarz method for solving linear systems of equations [104], SART is shown to rapidly converge to a solution approaching the fidelity of that produced by alternative iterative methods, including trust region and conjugate

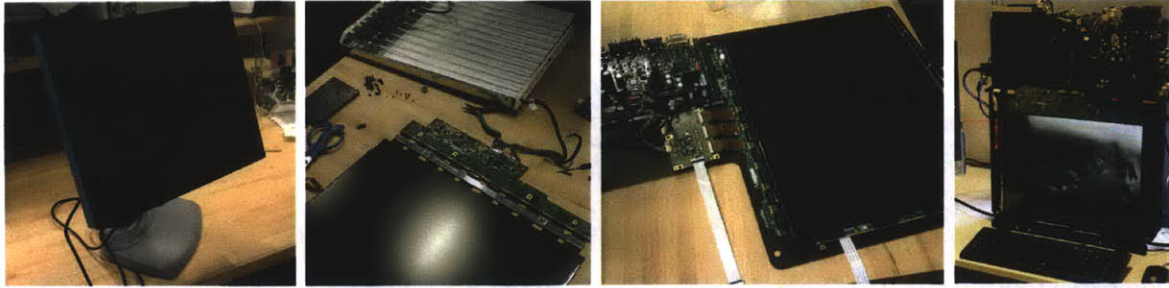


Figure 3-5: Constructing the polarization field display prototype. Four monochromatic LCDs were modified to create a single multi-layer LCD. Photographs depict from left to right: an unmodified Barco E-2320 PA LCD, the liquid crystal panel and backlight after removing the case and power supply, a modified panel mounted on an aluminum frame, and the assembled prototype.

gradient descent techniques [107] (see Figure 3-3). In Section 3.2.5 we show that SART allows for real-time optimization for interactive polarization field displays.

In summary, polarization fields present both an optically and computationally efficient architecture for dynamic light field display using multi-layer LCDs. We briefly contrast this architecture to that required for a direct extension of the attenuation-based method proposed by Wetzstein et al. [205]. As shown in Figure 3-4, a multi-layered, attenuation-based display is fabricated by placing a polarizer on the backlight and additional polarizers after each liquid crystal layer, effectively creating a set of dynamically-programmable transparencies; however, such a design reduces the display brightness by a factor of 0.8^{K-2} compared to the proposed polarization field display, assuming a maximal transmission of 80% through each polarizer (as measured for those used in the prototype). Yet, we observe our adaptation of SART can similarly be applied to attenuation layers by substituting the logarithm of the emitted light field intensity \tilde{l}_m and the logarithm of the transmittance t_n for $\tilde{\theta}_m$ and ϕ_n in Equation 3.13, respectively; thus, we provide the first implementation for achieving interactive frame rates with such designs.

3.2.5 Implementation

This section describes the construction and performance of the prototype. First, we summarize the modifications made to commercial LCD panels to create a reconfigurable multi-layer display. Second, we review the off-line and real-time software for light field rendering, an-

tialiasing, and optimizing layer patterns. Third, we assess the prototype, evaluating our image formation model and illustrating the practical benefits and limitations of polarization field displays.

Hardware

Given that we require monochromatic layers and field sequential color, a custom prototype was necessary. PureDepth [16] offers dual-layer LCDs, but no supplier was found for multi-layer configurations. Each layer of the prototype consists of a modified Barco E-2320 PA LCD, supporting 1600×1200 8-bit grayscale display at 60 Hz, and an active area of 40.8×30.6 cm. As shown in Figure 3-5, the liquid crystal layer was separated from the case, backlight, and power supply. Polarizing films were removed and the adhesive was dissolved with acetone. By design, the driver board is folded behind the panel, blocking a portion of the display when used in a stacked configuration. An extended ribbon cable was constructed to allow the board to be folded above the display using a pair of 20-pin connectors and a flat flexible cable. The exposed panel, driver boards, and power supply were mounted to a waterjet-cut aluminum frame. Four such panels were constructed and stacked on a wooden stand. Arbitrary layer spacings are supported by translating the frames along rails. Acrylic spacers hold the layers at a fixed spacing of 1.7 cm for all experiments described in this paper, yielding a total display thickness of 5.1 cm. The prototype is illuminated using an interleaved pair of backlights and controlled by a 3.4 GHz Intel Core i7 workstation with 4 GB of RAM. A four-head NVIDIA Quadro NVS 450 graphics card synchronizes the displays. See Figure 3-40 for additional details on the construction of the prototype.

As shown in Figure 3-4, the display operates in either attenuation-based or polarization-based modes. The original polarizers were discarded and replaced with American Polarizers AP38-006T linear polarizers. By specification, a single polarizer has a transmission efficiency of 38% for unpolarized illumination. Transmission is reduced to 30% through a pair of aligned polarizers, yielding an efficiency of 80% for polarized light passing through a single, aligned polarizer. Five polarizers are required for attenuation-based display, with a pair of crossed polarizers on the rear layer followed by successively-crossed polarizers on each

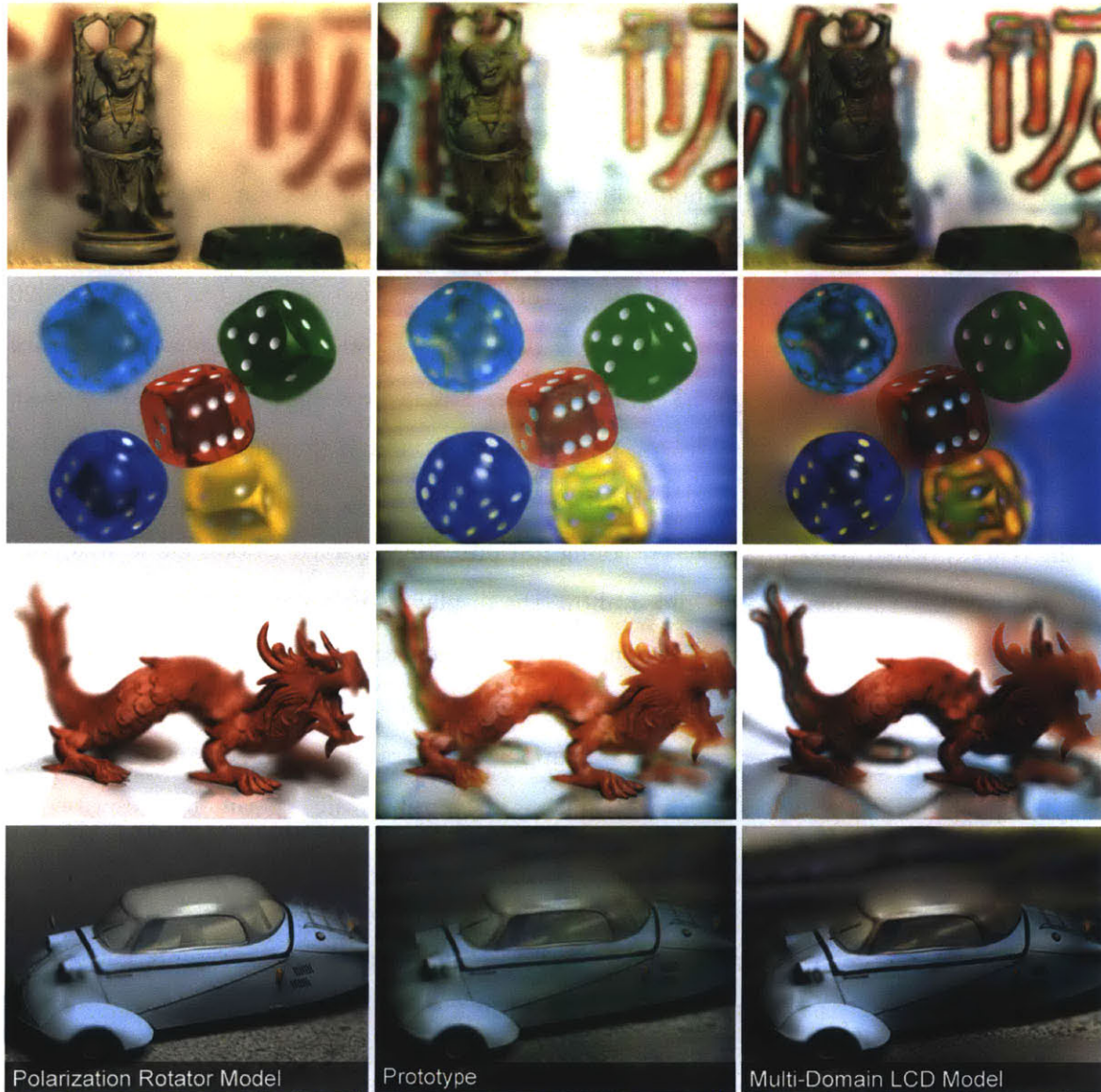


Figure 3-6: Polarization field display using the multi-layer prototype. The central views for the “Buddha”, “dice”, “dragon”, and “car” scenes are shown. Views predicted by the polarization rotator model (Equation 3.12) and the multi-domain LCD model (Equation 3.22) are compared in the left and right columns, respectively. Photographs of the prototype are shown in the middle. Section 3.2.6 and Section 3.2.6 quantitatively assess performance and artifacts.

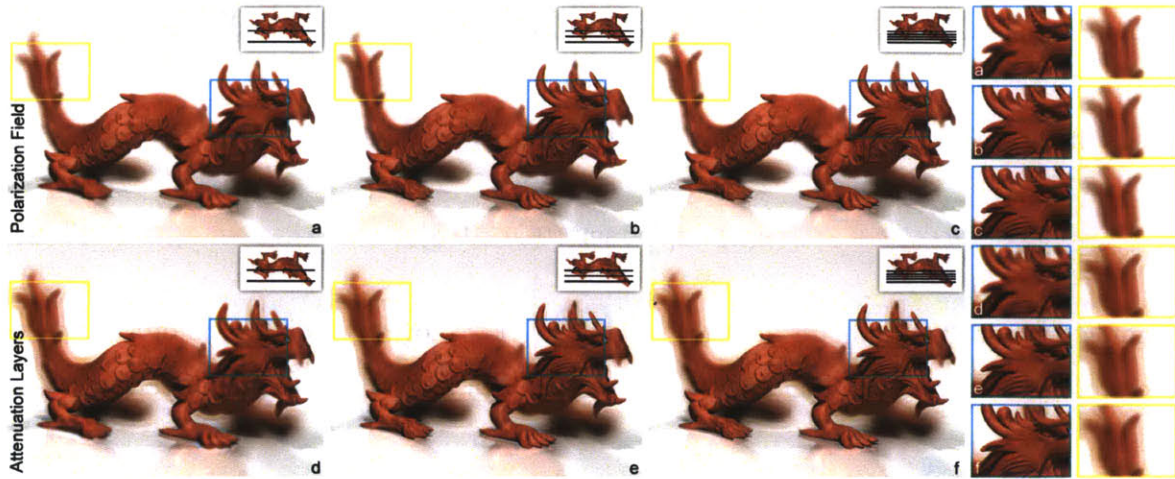


Figure 3-7: Simulated light field reconstructions using polarization fields (top row) and attenuation layers (bottom row) are shown for two, three, and five layers from left to right. Layer positions with respect to the scene are illustrated in the insets. Note that the reconstruction fidelity of objects within and outside the physical display extent increases for a larger number of layers, as highlighted by the cyan and yellow regions, respectively. Due to bias in the least-squares solution for a log-domain objective, optimized tomographic reconstructions for attenuation-based displays suffer from halo artifacts around high-contrast edges, which is not the case for polarization field displays.

remaining layer. A polarization field display is implemented by enclosing the stack by a single pair of crossed polarizers. Field sequential color is simulated, for still imagery, by combining three photographs taken while alternating the color channel displayed on each layer. To assist registration, examples in this paper use the color filters included in the Bayer mosaic of the camera, whereas the video summarizes experiments using Roscolux filters (#26, #91, and #80) placed on the backlight. The video shows dynamic examples in grayscale.

Each panel must be radiometrically calibrated to allow an accurate mapping from optimized rotation angles to displayed image values. The Barco E-2320 PA is intended for medical diagnostic imaging and replicates the DICOM Grayscale Standard Display Function. The normalized displayed intensity $I \in [0, 1]$ was measured as a function of the 8-bit image value $v \in [0, 255]$ using a photometer held against an unmodified panel. The resulting radiometric response curve is approximated by a gamma value of $\gamma = 3.5$ such that $I = (v/255)^\gamma$. Thus, gamma compression maps optimized pixel transmittances to image values when operating in the attenuation-based mode. When operated as a polarization field display, optimization

yields the polarization state rotation ϕ for each pixel. For an unmodified panel we model this mapping by Equation 3.8 such that $I = \sin^2(\phi)$. Equating this with the gamma curve yields the following mapping between rotations and image values.

$$v(\phi) = \lfloor 255 \sin^{2/\gamma}(\phi) + 0.5 \rfloor \quad (3.17)$$

Figures 3-1 and 3-6 compare modeled light field views to corresponding photographs of the prototype. Figure 3-4 compares the attenuation-based mode to the polarization-based mode.

Software

The light fields in this paper are rendered with a spatial resolution of 512×384 pixels and depict 3D scenes with both horizontal and vertical parallax from 7×7 viewpoints within a field of view of 10 degrees. POV-Ray is used to render the scenes shown in Figure 3-6. Following Levoy and Hanrahan [126] and Zwicker et al. [219], we apply a 4D antialiasing filter to the light fields by rendering each view with a limited depth of field. As analyzed by Wetzstein et al. [205], this antialiasing filter simultaneously approximates the limited depth of field established for multi-layer light field displays.

The Matlab LSQLIN solver serves as the reference solution to Equation 3.14, implementing a sparse, constrained, large-scale trust region method [40]. This solver converges in about 8 to 14 iterations for three to five attenuating or polarization-rotating layers. Solutions are found within approximately 10 minutes on the previously-described Intel Core i7 workstation.

The SART algorithm given by Equation 3.15 is implemented in Matlab and on the GPU. We observe SART is well-suited for parallel processing on programmable GPUs [109]. Our code is programmed in C++, OpenGL, and Cg. Light fields are rendered and antialiased in real-time using OpenGL, followed by several iterations of the GPU-based SART implementation. We achieve refresh rates of up to 24 frames per second using one iteration for four layers running on the NVIDIA Quadro NVS 450. Figure 3-3 illustrates SART convergence, demonstrating that 2 to 5 iterations minimize reconstruction artifacts. Estimates for the

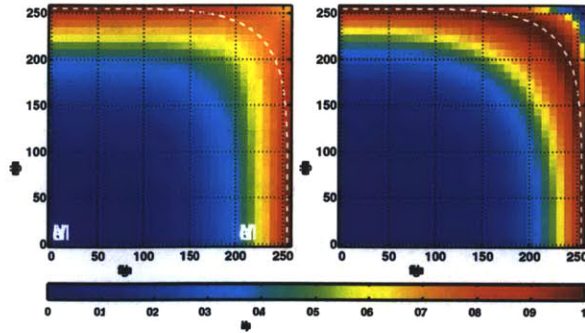


Figure 3-8: Radiometric calibration of the prototype. The measured (left) and modeled (right) normalized intensity I is plotted as image values v_1 and v_2 are displayed on the rear and front layer, respectively. The model is a least-squares fit of Equation 3.18 to the measured intensities. Note that the prototype only uses rotations corresponding to values located on the lower left of the white lines.

previous frame may seed the optimization for the current frame. For static scenes, this effectively implements an increasing number of SART iterations over time, while providing a suitable initialization for successive frames in a dynamic environment.

3.2.6 Assessment

Prototype Performance

As shown in Figure 3-1, polarization fields accurately depict multiple perspectives of the “Buddha” scene. Viewpoint variations capture highlights on the incense burner and occlusions of the background characters. Figure 3-11 demonstrates faithful reproduction of translucency for the dice and through the windows of the car. Detailed results for each scene is included in Figures 3-12,3-13,3-14, and 3-15. Smooth motion parallax is achieved and demonstrated in video provided with the 2011 SIGGRAPH Asia paper [121].

While confirming the prototype achieves automultiscopic display, photographs exhibit artifacts not predicted by simulations. Moiré is present, although it could be mitigated using the method of Bell et al. [16]. We attribute intensity artifacts, visible in Figure 3-6, to discrepancies between the prototype and the ideal construction using polarization-rotating layers. As analyzed in Section 3.2.6, the primary discrepancy is the presence of multiple

liquid crystal domains in our panels. Furthermore, as characterized by Yeh and Gu [214], commercial panels do not operate precisely as two-dimensional polarization rotators, particularly at oblique angles. To this end, we used photometric measurements to assess our model. As shown in Figure 3-8, a photometer measured the normalized intensity I as differing image values v_1 and v_2 were displayed on the rear and front layer, respectively. Substituting Equation 3.17 into Equation 3.12 yields the following prediction.

$$I(v_1, v_2) = \sin^2 \left\{ \sin^{-1} \left[\left(\frac{v_1}{255} \right)^{\frac{7}{2}} \right] + \sin^{-1} \left[\left(\frac{v_2}{255} \right)^{\frac{7}{2}} \right] \right\} \quad (3.18)$$

Measured intensities are nearly identical upon interchanging v_1 and v_2 , validating the additive model in Equation 3.11—upon which our tomographic optimization relies. Measured contrast is limited when v_1 and v_2 are large. This is confirmed in the supplemental video; overlaying a pair of white images produces a darker image, but with reduced contrast. Thus, artifacts persist in the prototype due to differences between our off-the-shelf panels and ideal polarization rotators. Additional measurements are summarized in Figure 3-10.

In Figure 3-7, polarization fields perform comparably to attenuation layers in terms of reconstruction fidelity. Yet, halo artifacts are noticeably reduced. We attribute this primarily to different biases introduced by least-squares optimization of transformed objective functions. As proposed by Gotoda [64] and Wetzstein et al. [205], attenuation-based displays optimize an objective, reminiscent of Equation 3.14, defined for the logarithm of the target intensities. This penalizes artifacts in dark regions, leading to the observed halos. By comparison, polarization fields optimize an objective defined for target intensities transformed by Equation 3.10; this transformation is more linear than for attenuation, thereby mitigating halos. This is confirmed by the average peak signal-to-noise ratio (PSNR) plots shown in Figure 3-9, in which polarization fields slightly outperform attenuation layers. Based on these trials, we conclude that polarization fields present an optically-efficient alternative to attenuation layers optimally-suited to multi-layer LCDs, closely mirroring the PSNR trends and dependence on the layer numbers and display thickness previously established for attenuation-based displays.

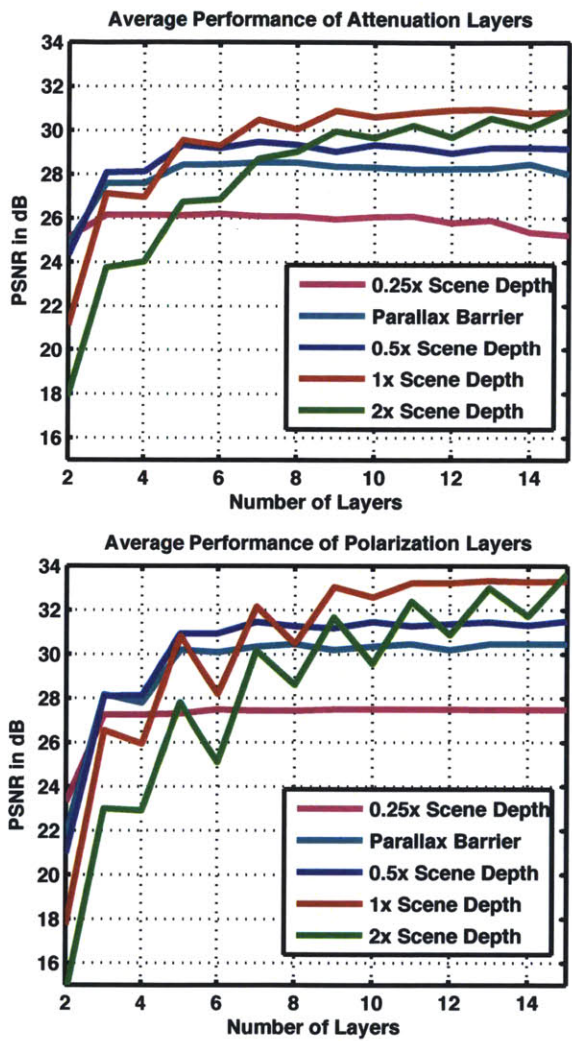


Figure 3-9: Average PSNR for attenuation layers vs. polarization fields. The PSNR was averaged for the four scenes in Figure 3-6 (and for two more in the video) depending on the number of layers and the relative display thickness. Note that polarization fields can accurately present objects beyond the display, but can also be operated in a volumetric mode enclosing the scene for reduced errors.

Multi-Domain LCDs

Artifacts observed in the prototype are not predicted by the polarization rotator model. We show artifacts can be primarily attributed to the presence of multiple liquid crystal domains in the in-plane switching (IPS) panels used in the prototype. By applying the Jones calculus, we introduce a multi-domain LCD model that accounts for artifacts and provides a formal means to assess model limitations.

As described by Yeh and Gu [214], the Jones matrix modeling an LCD depends on its architecture. Yet, as described by Date et al. [47], all LCDs are fundamentally retardation-based and can be approximated as *rotated half-wave plates*, with Jones matrix:

$$\mathbf{J}_{\text{HWP}}(\alpha) = \begin{pmatrix} \cos(2\alpha) & \sin(2\alpha) \\ \sin(2\alpha) & -\cos(2\alpha) \end{pmatrix}, \quad (3.19)$$

where α is the liquid crystal director angle. Compared to a true polarization rotator, each LCD acts as a *pseudo-rotator*: reversing the polarization state and doubling the rotation angle. The following expression models the normalized intensity for K -layer compositions of single-domain LCDs enclosed by crossed linear polarizers.

$$\begin{aligned} I_{\text{HWP-K-1}}(\boldsymbol{\alpha}) &= I_0 \left| \begin{pmatrix} 0 & 1 \end{pmatrix} \left(\prod_{k=1}^K \mathbf{J}_{\text{HWP}}(\alpha_{K-k+1}) \right) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right|^2 \\ &= I_0 \sin^2 \left(\sum_{k=1}^K (-1)^{k-1} 2\alpha_k \right) \end{aligned} \quad (3.20)$$

For the choice $\alpha_k = (-1)^{k-1} \phi_k / 2$, this expression is identical to Equation 3.12. Thus, under this model, multi-layer, single-domain LCDs can approximate layered polarization rotators.

Following Date et al. [47], we assume every IPS pixel is divided into two domains. Each domain i in layer k is approximated as a rotated half-wave plate $\mathbf{J}_{\text{HWP}}(\alpha_k^{(i)})$ with symmetric directors such that $\alpha_k^{(1)} = -\alpha_k^{(2)} = \alpha$. When the angle between the linear polarizers is not a multiple of 90 degrees, the normalized intensity for a single multi-domain panel differs from Equation 3.20. In Figure 3-10, this fact is used to confirm the prototype panels contain

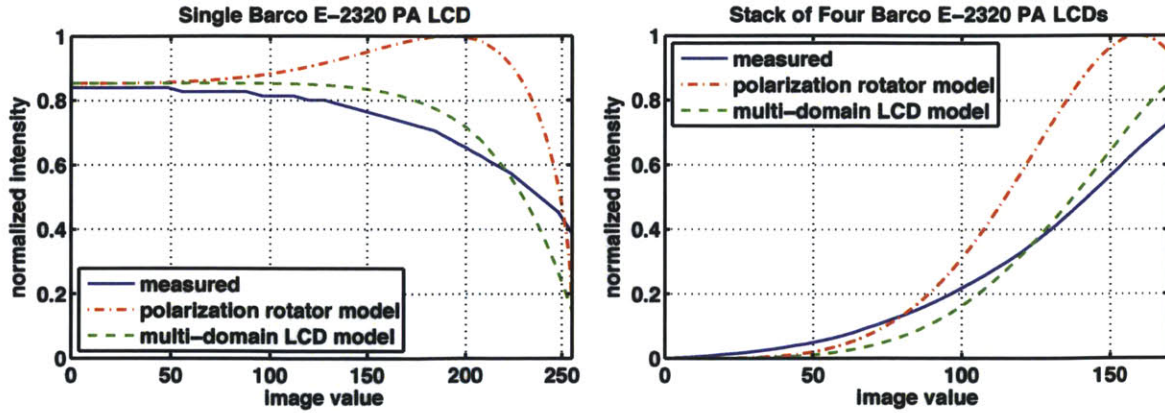


Figure 3-10: Radiometric measurements of the prototype. (Left) Normalized intensity for a single panel enclosed by polarizers with a relative rotation of 22.5 degrees. (Right) Normalized intensity for the four-layer prototype (each layer displays the same value). While the multi-domain model is more accurate, the polarization rotator model provides an approximation enabling real-time optimization.

multiple domains.

For a multi-layer, multi-domain LCD, rays emitted by the backlight will pass through a single domain in each layer. Considering a bundle of rays passing through a local region, the intensity will depend on the weighted average due to passing through all domain combinations. Summing over combinations yields the following expression for normalized intensity for two-layer, two-domain LCDs.

$$\begin{aligned}
 I_{\text{HWP-2-2}}(\boldsymbol{\alpha}) &= \frac{I_0}{4} \sum_{i=1}^2 \sum_{j=1}^2 \left| \begin{pmatrix} 0 & 1 \end{pmatrix} \mathbf{J}_{\text{HWP}}(\alpha_2^{(j)}) \mathbf{J}_{\text{HWP}}(\alpha_1^{(i)}) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right|^2 \\
 &= I_0 \left(\frac{\sin^2(2(\alpha_1 + \alpha_2)) + \sin^2(2(\alpha_1 - \alpha_2))}{2} \right) \quad (3.21)
 \end{aligned}$$

This expression provides intuition into how multi-layer, multi-domain LCDs deviate from polarization rotators. The first term is proportional to Equation 3.12, whereas the second term constitutes the error under a polarization rotator approximation. Extending this analysis to four layers yields the following expression.

$$I_{\text{HWP-4-2}}(\boldsymbol{\alpha}) = I_0 \left(\frac{1 - \prod_{k=1}^4 \cos(4\alpha_k)}{2} \right) \quad (3.22)$$

In Figure 3-10, we quantify how the polarization rotator approximation deviates from both

experiments and the multi-domain model (particularly for large image values). We observe, for small image values or cases for which values are large for a single layer, measurements and the multi-domain model are well approximated.

In conclusion, we identify the presence of multiple domains as the primary source of artifacts. This insight reveals potential solutions. Since the multi-domain model accurately predicts experimental artifacts (see Figure 3-6), one may consider it as a foundation for an enhanced optimization procedure; however, Equation 3.22 is non-linear and not directly amenable to real-time optimization via the SART algorithm. Alternatively, replacing panels with single-domain alternatives is predicted, via Equation 3.20, to better approximate polarization rotators. In practice we expect both strategies must be pursued, together with laboratory characterizations, to obtain the full performance afforded by polarization field displays.

3.3 High-Rank 3D

In this section we introduce a two layer temporally multiplexed light field display and the concepts necessary to drive such displays compressively. These concepts will be generalized to drive a wider set of displays in Section 3.4. To date, such dual-stacked LCDs have used heuristic parallax barriers for view-dependent imagery: the front LCD shows a fixed array of slits or pinholes, independent of the multi-view content. While prior works adapt the spacing between slits or pinholes, depending on viewer position, we show both layers can also be adapted to the multi-view content, increasing brightness and refresh rate. Unlike conventional barriers, both masks are allowed to exhibit non-binary opacities. It is shown that any 4D light field emitted by a dual-stacked LCD is the tensor product of two 2D masks. Thus, any pair of 1D masks only achieves a rank-1 approximation of a 2D light field. Temporal multiplexing of masks is shown to achieve higher-rank approximations. This insight allows us to cast light field display as a matrix approximation problem. Non-negative matrix factorization (NMF) minimizes the weighted Euclidean distance between a target light field and that emitted by the display.

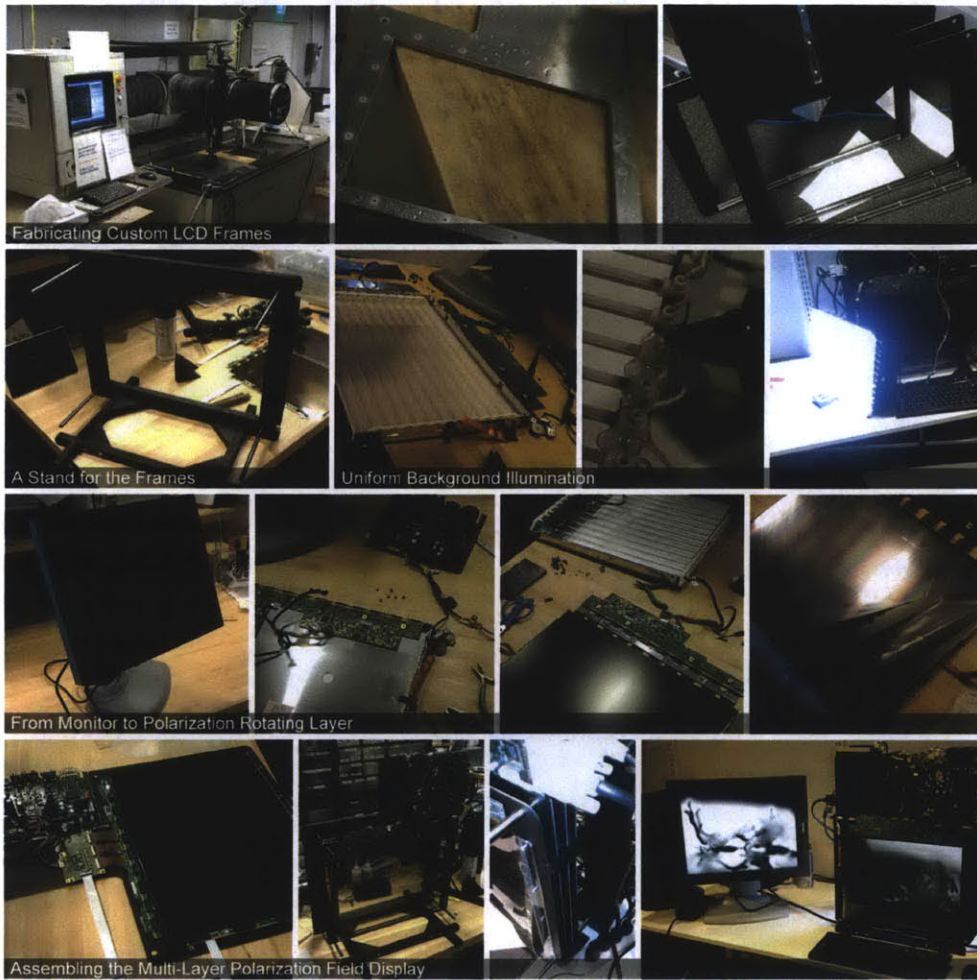


Figure 3-11: Photographic documentation of the prototype construction for Polarization Fields. Four monochrome, off-the-shelf medical LCDs were modified. Polarizing films were removed and electronics repositioned so the panels could be mounted on custom-fabricated frames. The layers are separated by acrylic spacers and illuminated by a uniform backlight.



Figure 3-12: “Buddha” scene, using the “Buddha” model from <http://graphics.stanford.edu/data/3Dscanrep/>. Simulated views are compared for polarization-rotating layers (first and second columns) and attenuating layers (fourth and fifth columns). Rows illustrate, from top to bottom: target views, reconstructions using the off-line solver for two, three, and four layers (for the same depth range), and SART reconstructions with two, five, and fifty iterations with four layers. Columns three and six present decomposed layers.

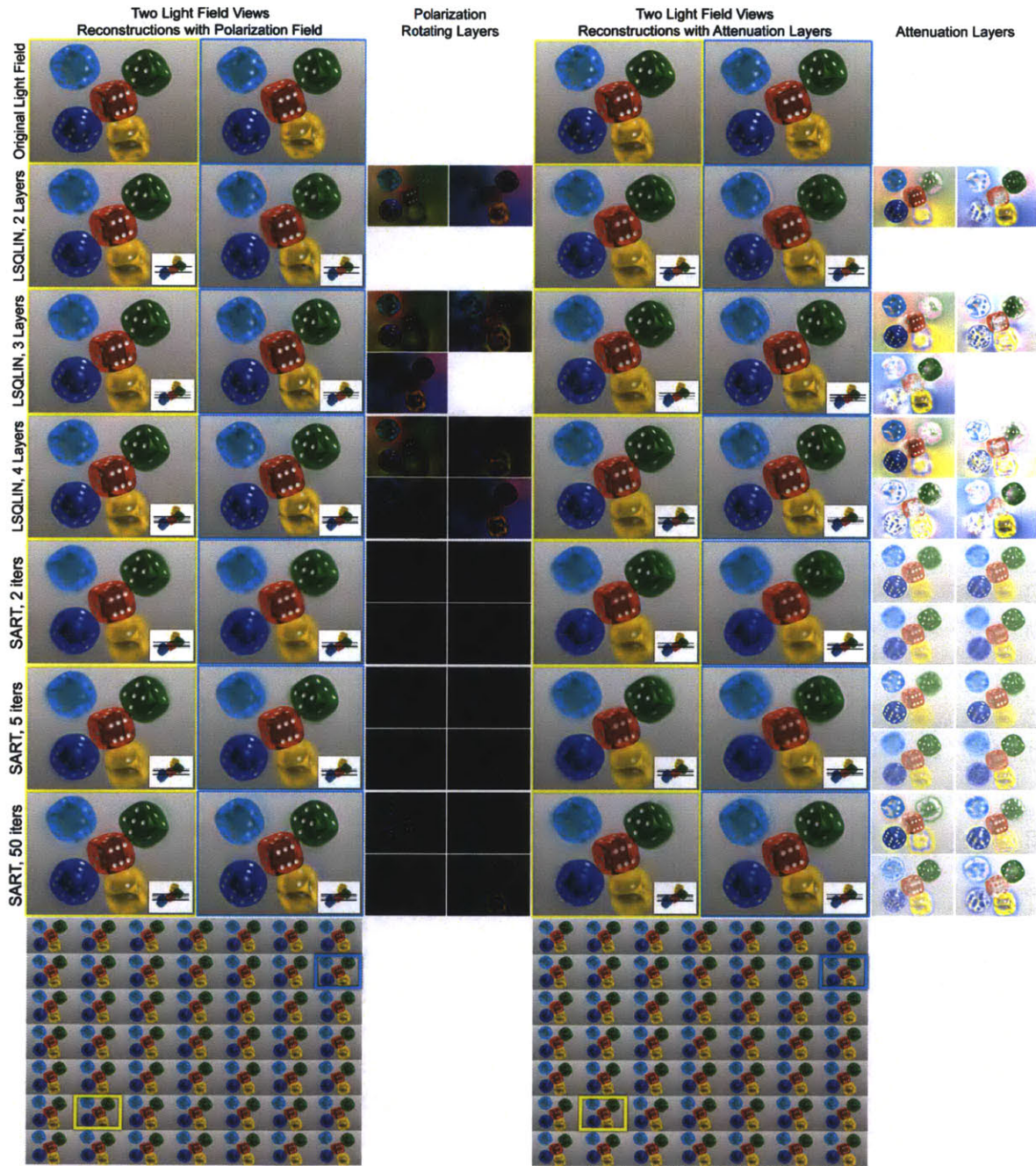


Figure 3-13: Additional results for the “dice” scene. The display dimensions and optimization parameters match that of the prototype and correspond with those used in Figure 3-12. The light field has a resolution of 512×384 spatial samples and 7×7 angular samples. The target imagery spans a field of view of 10 degrees.

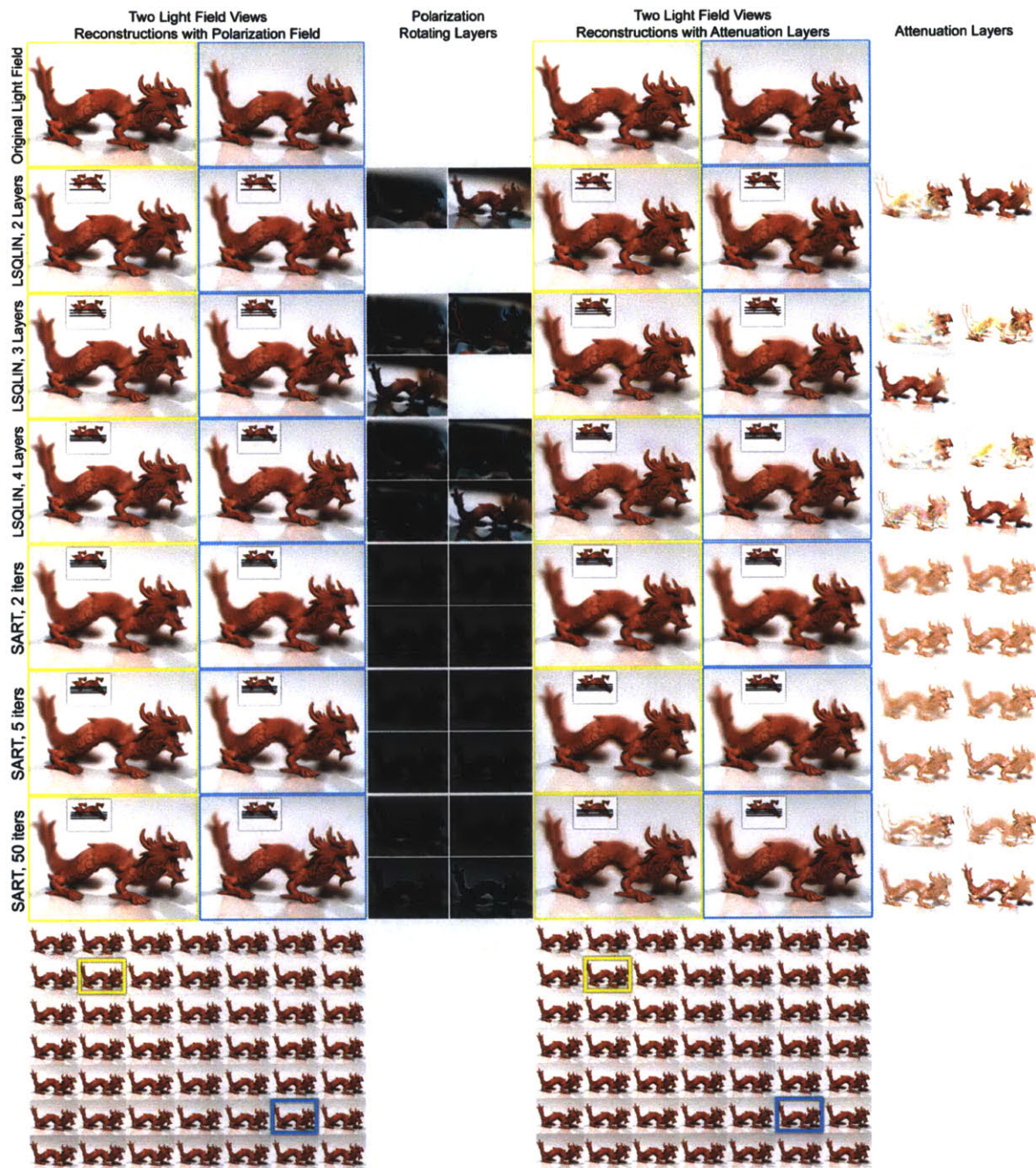


Figure 3-14: Additional results for the “dragon” scene, using the “dragon” model from <http://graphics.stanford.edu/data/3Dscanrep/>. The display dimensions and optimization parameters match that of the prototype and correspond with those used in Figure 3-12. The light field has a resolution of 512×384 spatial samples and 7×7 angular samples. The target imagery spans a field of view of 10 degrees.

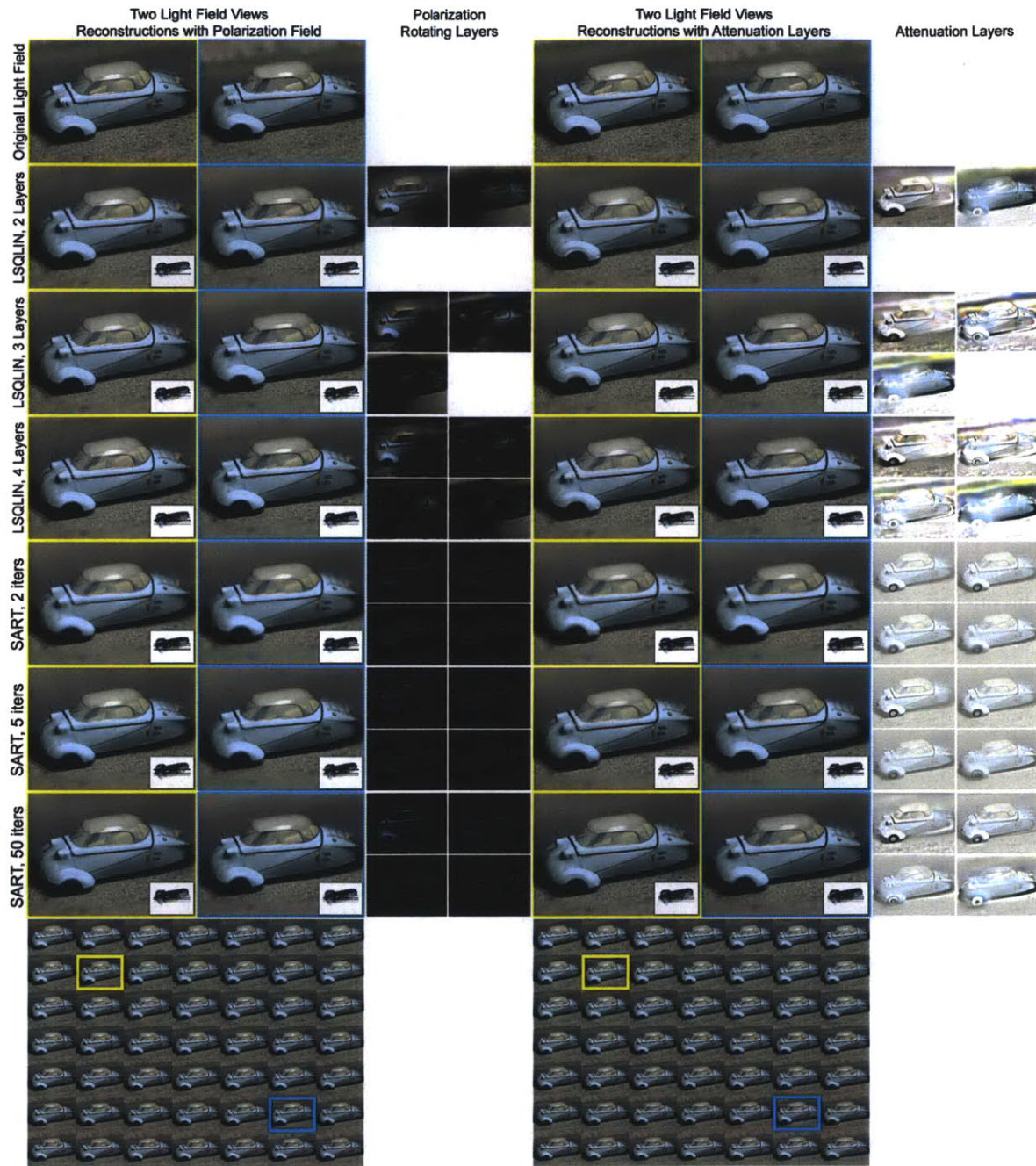


Figure 3-15: Additional results for the “car” scene. The display dimensions and optimization parameters match that of the prototype and correspond with those used in Figure 3-12. The light field has a resolution of 512×384 spatial samples and 7×7 angular samples. The target imagery spans a field of view of 10 degrees.

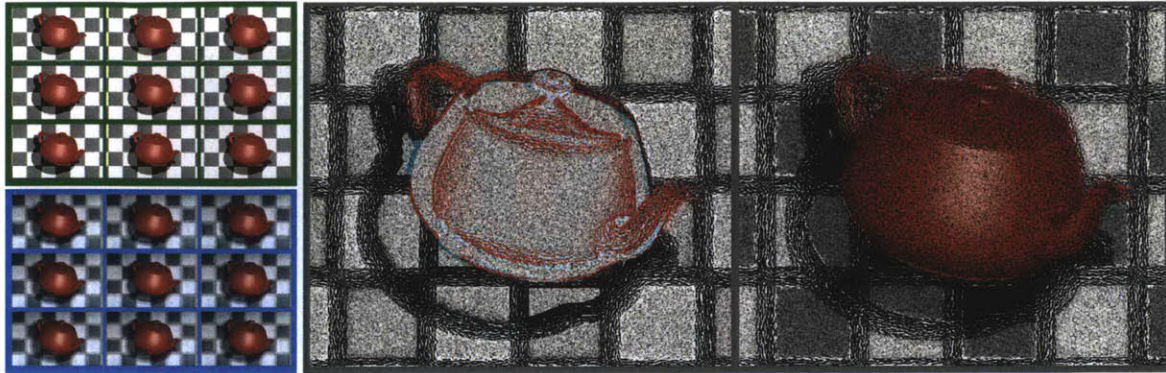


Figure 3-16: 3D display with content-adaptive parallax barriers. We show that light field display using dual-stacked LCDs can be cast as a matrix approximation problem, leading to a new set of *content-adaptive parallax barriers*. (Left, Top) A 4D light field, represented as a 2D array of oblique projections. (Left, Bottom) A dual-stacked LCD displays the light field using content-adaptive parallax barriers, confirming both vertical and horizontal parallax. (Middle and Right) A pair of content-adaptive parallax barriers, drawn from a rank-9 decomposition of the reshaped 4D light field matrix. Compared to conventional parallax barriers, with heuristically-determined arrays of slits or pinholes, content adaptation allows increased display brightness and refresh rate while preserving the fidelity of projected images.

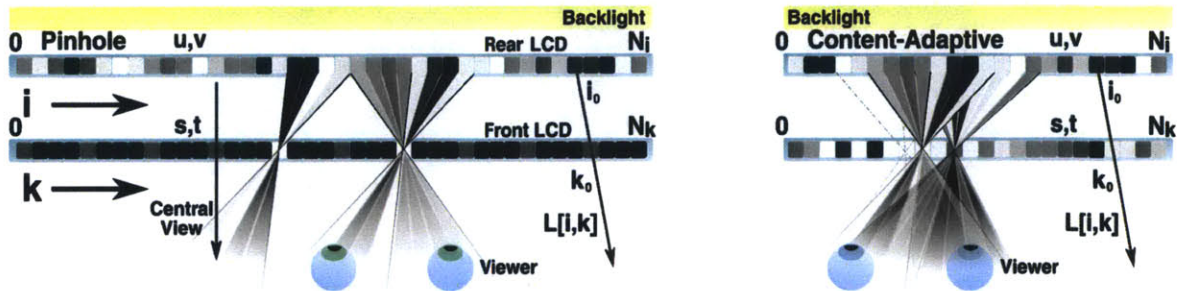


Figure 3-17: Conventional versus content-adaptive parallax barriers. (Left) In a conventional parallax barrier display the front panel contains a uniform grid of slits or pinholes. The viewer sees each pixel on the rear panel through this grid, selecting a subset of visible pixels depending on viewer location. A uniform backlight, located behind the rear layer, enables the rear layer to act as a conventional 2D display. (Right) Rather than heuristic barriers, we consider dual-stacked LCDs as general spatial light modulators that act in concert to recreate a target light field by attenuating rays emitted by the backlight. Unlike conventional barriers, both masks can exhibit non-binary opacities.

3.3.1 Content-Adaptive Parallax Barriers

In this section dual-stacked LCDs are analyzed as general spatial light modulators that act in concert to recreate a light field by attenuating rays emitted by the backlight. It is shown that any fixed pair of masks only creates a rank-1 approximation of a light field. Higher-rank approximations are achieved with time multiplexing. We optimize 3D display with dual-stacked LCDs using a matrix approximation framework. This leads to content-adaptive parallax barriers allowing brighter displays with increased refresh rates.

Light Field Analysis

A general parallax barrier display, containing two mask layers and a backlight, can be analyzed as a light field display device. The following analysis adopts an *absolute two-plane parameterization* of the 4D light field. As shown in Figure 3-17, an emitted ray is parameterized by the coordinates of its intersection with each mask layer. Thus, the ray (u, v, s, t) intersects the rear mask at the point (u, v) and the front mask at the point (s, t) , with both mask coordinate systems having an origin in the top-left corner.

In a practical automultiscopic display one is primarily concerned with the projection of optical rays within a narrow cone perpendicular to the display surface (see Figure 3-17), since most viewers will be located directly in front of the device. The distinct images viewable within this region are referred to as the “central views” projected by the display. As a result, a *relative two-plane parameterization* proves more convenient to define a target light field; in this parameterization, an emitted ray is defined by the coordinates (u, v, a, b) , where (u, v) remains the point of intersection with the rear plane and (a, b) denotes the relative offset of the second point of intersection such that $(a, b) = (s - u, t - v)$. As shown in Figure 3-16, a 2D slice of the 4D light field, for a fixed value of (a, b) , corresponds to a skewed orthographic view (formally an oblique projection).

A general pair of 2D optical attenuation functions, $f(u, v)$ and $g(s, t)$, is defined with the absolute parameterization. These functions correspond to the rear and front masks, respec-

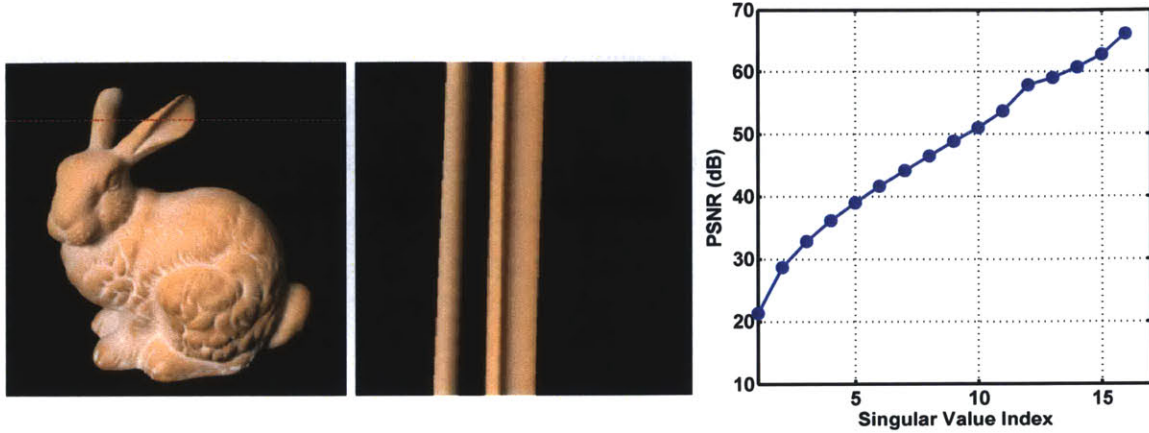


Figure 3-18: The rank of the *bunny* light field [116] is assessed. (Left) The central image captured by translating a camera within a 17×17 grid. (Middle) A 2D slice, along the dashed red line, of the 4D light field. (Right) The rank is assessed by the singular value decomposition of the 2D slice. The reconstruction error, measured in terms of the peak signal-to-noise ratio (PSNR), is plotted as a function of the number of singular values included in a given low-rank approximation. In this example, the numerical matrix rank is equal to 17 (i.e., the number of views contained in the light field slice). However, reconstruction with at least three singular values leads to a PSNR greater than 30 dB (generally accepted for lossy compression).

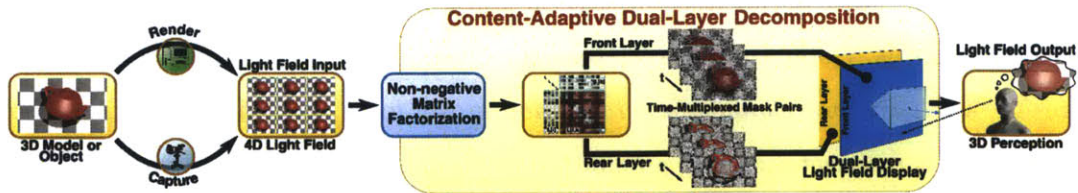


Figure 3-19: A thin, dual-layer display (e.g., a dual-stacked LCD) allows depth perception without special eyewear. Multi-view content is rendered or photographed and represented as a 4D light field. Content-adaptive parallax barriers are obtained by applying non-negative matrix factorization to the input light field, increasing display brightness and refresh rate compared to conventional barriers. These mask pairs are displayed using the dual-layer display, emitting a low-rank approximation of the input light field and enabling depth perception.

tively. The emitted 4D light field $L(u, v, s, t)$ is given by the product

$$L(u, v, s, t) = f(u, v)g(s, t), \quad (3.23)$$

assuming illumination by a uniform backlight. In practice, the masks and the emitted light field are discrete functions. The discrete pixel indices are denoted as (i, j, k, l) , corresponding to the continuous coordinates (u, v, s, t) , such that the discretized light field is $\mathbf{L}[i, j, k, l]$ and the sampled masks are $\mathbf{f}[i, j]$ and $\mathbf{g}[k, l]$. When considering only a 2D slice of the 4D light field, the resulting 2D light field matrix $\mathbf{L}[i, k]$ is given by the outer product

$$\mathbf{L}[i, k] = \mathbf{f}[i] \otimes \mathbf{g}[k] = \mathbf{f}[i]\mathbf{g}^T[k], \quad (3.24)$$

with the masks represented as column vectors $\mathbf{f}[i]$ and $\mathbf{g}[k]$. Note that Equation 3.23 can be compactly expressed as an outer product only by adopting an absolute two-plane parameterization. For 4D light fields, Equation 3.24 can be generalized so the light field is given by the tensor product of the 2D masks as follows.

$$\mathbf{L}[i, j, k, l] = \mathbf{f}[i, j] \otimes \mathbf{g}[k, l] \quad (3.25)$$

Rank Constraints

From Equation 3.24 it is clear that a fixed pair of 1D masks can only produce a rank-1 approximation of any given 2D light field matrix. Similarly, a fixed pair of 2D masks also produces a rank-1 approximation of the discrete 4D light field tensor via Equation 3.25. To our knowledge, this restriction has not been previously described for parallax barrier displays and provides an important insight into their inherent limitations. Figure 3-18 and Section 3.3.3 evaluate the rank of several synthetic and captured light fields; except for the special case when all objects appear in the plane of the display, the rank is typically greater than one. Thus, dual-stacked LCDs employing fixed mask pairs produce rank-deficient approximations; however, perceptually-acceptable approximations can be obtained using

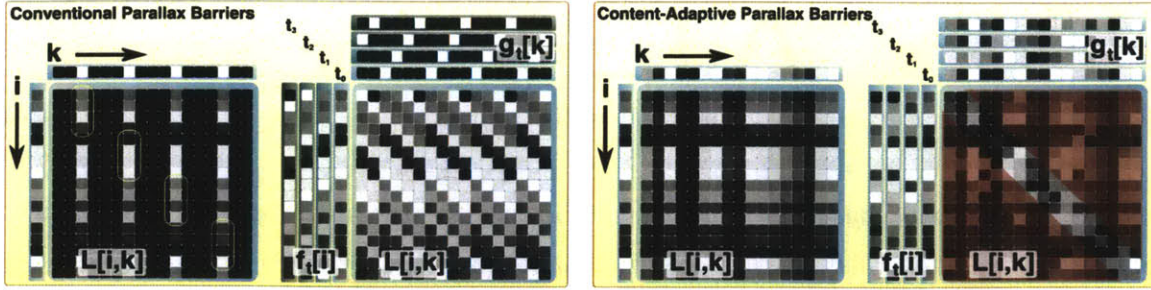


Figure 3-20: Rank constraints for parallax barriers. (Left) Conventional parallax barriers, following Equation 3.24, approximate the light field matrix (center) as the outer product of mask vectors (above and to the left). The resulting rank-1 approximation accurately reproduces the circled elements (corresponding to the central views in Figure 3-17). Note that most columns are not reconstructed, reducing display resolution and brightness. Periodic replicas of the central views are created outside the circled regions. (Middle Left) Time-shifted parallax barriers achieve higher-rank reconstructions by integrating a series of rank-1 approximations, each created by a single translated mask pair. (Middle Right) Content-adaptive parallax barriers increase display brightness by allowing both masks to exhibit non-binary opacities. Here a rank-1 approximation is demonstrated using a single mask pair. (Right) Rank-T approximations are achieved using temporal multiplexing of T content-adaptive parallax barriers via Equations 3.28 and 3.31. In practice, the light field will be full rank without enforcing periodic replication (as created by conventional parallax barriers). As a result, we do not constrain rays (shown in red) outside the central view in Equation 3.31.

conventional parallax barriers, at the cost of decreasing the achievable spatial resolution and image brightness.

As shown in Figure 3-20, a conventional parallax barrier display employs a heuristic front mask given by

$$\mathbf{g}_{pb}[k, l] = \begin{cases} 1 & \text{if } k \bmod N_h = 0 \text{ and } l \bmod N_v = 0, \\ 0 & \text{otherwise,} \end{cases} \quad (3.26)$$

where N_h and N_v are the number of skewed orthographic views along the horizontal and vertical display axes, respectively. Thus, the front mask is either a uniform grid of slits or pinholes. Under this definition, the rear mask $\mathbf{f}[i, j]$ is defined such that Equation 3.25 is satisfied for every ray passing through a non-zero outer mask pixel; thus, the rear mask is given by

$$\mathbf{f}_{pb}[i, j] = \mathbf{L}[i, j, N_h \lfloor i/N_h \rfloor, N_v \lfloor j/N_v \rfloor], \quad (3.27)$$

when the resolutions of the front and rear masks are equal. Note that, for regions outside the central field of view, periodic replicas of the skewed orthographic views will be projected. These replicas result from viewing neighboring regions of the rear mask through the parallax barrier [89]. While not correctly capturing the true parallax resulting from steep viewing angles, periodic replication remains a beneficial property of conventional parallax barriers, allowing viewers to see perceptually-acceptable imagery outside the central viewing zone.

In theory, conventional parallax barrier displays achieve perfect reconstruction for any light field ray passing through a non-zero front mask pixel (within the central viewing region). However, as shown in Figure 3-20, no rays are projected for dark pixels on the front plane. The reconstructed light field will have significant reconstruction errors, when measured using the Euclidean distance between corresponding elements of the target light field. In practice, however, a viewer is separated by a distance that is significantly larger than the slit or pinhole spacing. Spatial low-pass filtering, as performed by the human eye, minimizes perceptual artifacts introduced by parallax barriers (i.e., blending the region between neighboring parallax barrier gaps). As a result, the occluded regions between slits or pinholes are not perceptually significant; however, these occluded regions significantly reduce the display brightness.

Time Multiplexing for Higher-Rank Approximation

Despite their practical utility, parallax barriers remain undesirable due to severe attenuation through a slit or pinhole array, as well as reduced spatial resolution of the output light field. Recently, time-shifted parallax barriers have been proposed to eliminate spatial resolution loss [111]. In such schemes, a stacked pair of high-speed LCDs is used to sequentially display a series of translated barriers $\mathbf{g}_{pb}[k, l]$ and corresponding underlying masks $\mathbf{f}_{pb}[i, j]$. If the complete mask set is displayed at a rate above the flicker fusion threshold, no image degradation will be perceived.

We generalize the concept of temporal multiplexing for parallax barriers by considering all possible mask pairs rather than the restricted class defined by Equations 3.26 and 3.27.

Any sequence of T 1D mask pairs creates (at most) a rank- T decomposition of a 2D light field matrix such that

$$\mathbf{L}[i, k] = \sum_{t=1}^T \mathbf{f}_t[i] \otimes \mathbf{g}_t[k] = \sum_{t=1}^T \mathbf{f}_t[i] \mathbf{g}_t^T[k], \quad (3.28)$$

where $\mathbf{f}_t[i]$ and $\mathbf{g}_t[k]$ denote the rear and front masks for frame t , respectively. Time-multiplexed light field display using dual-stacked LCDs can be cast as a matrix (or more generally a tensor) approximation problem. Specifically, the light field matrix must be decomposed as the matrix product

$$\mathbf{L} = \mathbf{F}\mathbf{G}, \quad (3.29)$$

where \mathbf{F} and \mathbf{G} are $N_i \times T$ and $T \times N_k$ matrices, respectively. Column t of F and row t of G are the masks displayed on the rear and front LCD panels during frame t , respectively. Further observe that a similar expression as Equation 3.28 can be used to approximate 4D light fields as the summation of multiple tensor products of 2D mask pairs as follows.

$$\mathbf{L}[i, j, k, l] = \sum_{t=1}^T \mathbf{f}_t[i, j] \otimes \mathbf{g}_t[k, l] \quad (3.30)$$

Non-negativity

Each mask pair $\{\mathbf{f}_t[i, j], \mathbf{g}_t[k, l]\}$ must be non-negative, since it is illuminated by an incoherent light source (i.e., the rear LCD backlight). We seek a *content-adaptive* light field factorization $\tilde{\mathbf{L}} = \mathbf{F}\mathbf{G}$ that minimizes the weighted Euclidean distance to the target light field \mathbf{L} , under the necessary non-negativity constraints, such that

$$\arg \min_{\mathbf{F}, \mathbf{G}} \frac{1}{2} \|\mathbf{L} - \mathbf{F}\mathbf{G}\|_{\mathbf{W}}^2, \text{ for } \mathbf{F}, \mathbf{G} \geq 0, \quad (3.31)$$

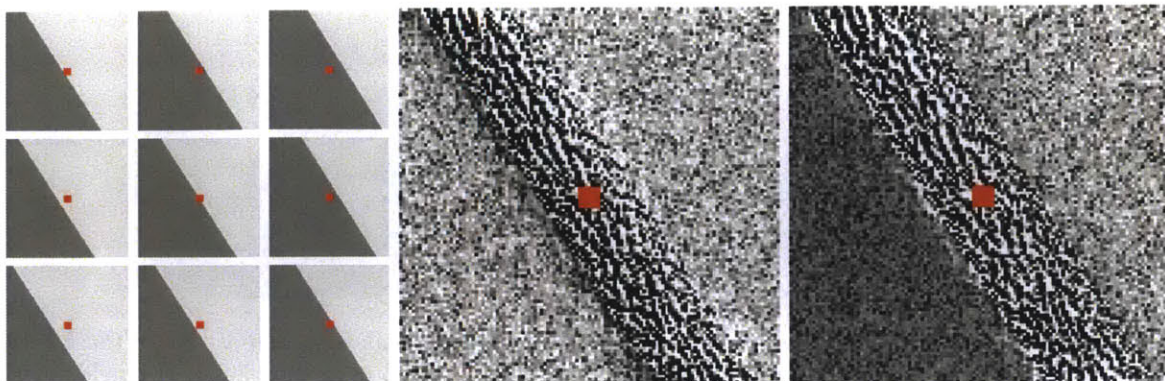


Figure 3-21: Intuition behind local parallax barriers. (Left) From left to right and top to bottom: oblique projections of a step edge seen as a viewer moves in similar directions. A rank-9 decomposition produces a set of mask pairs. (Middle) A rear-panel mask. (Right) A front-panel mask. Note that optimization appears to produce a local parallax barrier, rotated to align with the step edge.

where the reconstruction error is given by

$$\frac{1}{2} \|\mathbf{L} - \mathbf{F}\mathbf{G}\|_{\mathbf{W}}^2 = \sum_{ijkl} [\mathbf{W} \circ (\mathbf{L} - \mathbf{F}\mathbf{G}) \circ (\mathbf{L} - \mathbf{F}\mathbf{G})]_{ijkl}. \quad (3.32)$$

Here \circ denotes the Hadamard product for element-wise multiplication of matrices. Unlike conventional barriers, the field of view can be adapted to one or more viewers by specifying elements of the weight matrix \mathbf{W} (i.e., the Euclidean norm will be minimized where \mathbf{W} is large). The weight matrix plays a crucial role, ensuring a low-rank approximation can obtain high reconstruction accuracy by artificially reducing the rank of the target light field. General 4D light fields are handled by reordering as 2D matrices, with 2D masks reordered as vectors, allowing a similar matrix approximation scheme to be applied.

Equation 3.31 can be solved using non-negative matrix factorization. Prior numerical methods include the multiplicative update rule [122]. We use the weighted update introduced by Blondel et al. [20]. Initial estimates $\{\mathbf{F}, \mathbf{G}\}$ are refined as follows.

$$\mathbf{F} \leftarrow \mathbf{F} \circ \frac{[(\mathbf{W} \circ \mathbf{L})\mathbf{G}^{\top}]}{[(\mathbf{W} \circ (\mathbf{F}\mathbf{G}))\mathbf{G}^{\top}]} \quad \mathbf{G} \leftarrow \mathbf{G} \circ \frac{[\mathbf{F}^{\top}(\mathbf{W} \circ \mathbf{L})]}{[\mathbf{F}^{\top}(\mathbf{W} \circ (\mathbf{F}\mathbf{G}))]} \quad (3.33)$$

Typical mask pairs produced by the optimization procedure are shown in Figures 3-16,

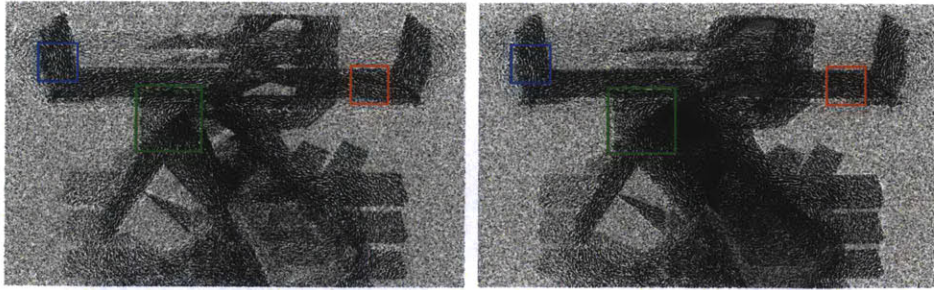


Figure 3-22: A content-adaptive parallax barrier mask pair. A rank-9 decomposition of the *blocks* light field, shown in Figure 3-24, was evaluated using Equation 3.33 in Section 3.3.1. A single mask pair is shown, with the rear and front masks to the left and right, respectively. To enhance the visibility of the emergent local parallax barriers, only the luminance channel of the light field is processed.

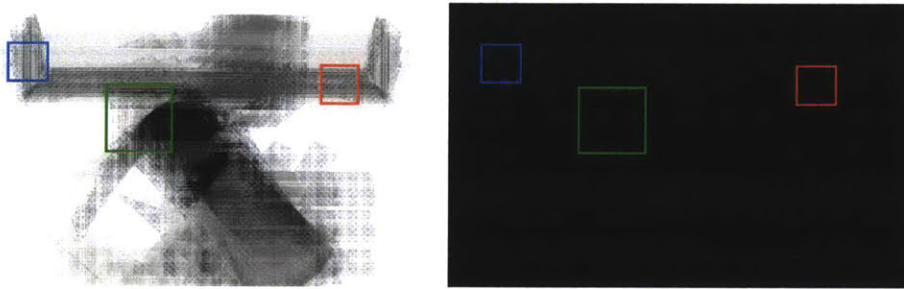


Figure 3-23: A conventional parallax barrier mask pair. A set of nine time-shifted conventional parallax barriers [111] were evaluated using Equations 3.26 and 3.27 in Section 3.3.1. A single mask pair is shown, with the rear and front masks to the left and right, respectively. For comparison with Figure 3-22, only the luminance channel of the light field is processed.

3-21, and 3-22. Note that, if Equation 3.31 was not constrained to weighted, non-negative factorizations, singular value decomposition (SVD) could be applied; however, Srebro and Jaakkola [176] have shown that solving a weighted SVD problem also requires an iterative algorithm with multiple local minima. In our implementation of Equation 3.33, the masks are initialized with random values uniformly distributed on $[0, 1]$; alternative strategies, including seeding with conventional parallax barriers, did not yield reconstructions with reduced errors or increased transmission. After each iteration the mask elements are truncated to the range $[0, 1]$. In conclusion, we propose the resulting non-negative, content-adaptive parallax barriers as a generalization of traditional parallax barrier displays, in which images displayed on both LCD layers are jointly optimized, independently for each target automultiscopic video frame.

Apparent Structure

Content-adaptive parallax barriers exhibit predictable structure. Consider the masks shown in Figures 3-16 and 3-22, as well as those in the supplementary materials: flowing, fringe-like patterns are consistently observed. We interpret that content-adaptive parallax barriers are locally-similar to conventional parallax barriers, but rotated to align to nearby edges in the light field. Intuitively, parallax is only perceived as a viewer moves perpendicular to an edge, thus a rotated *local parallax barrier* (i.e., an array of slits) is sufficient to project the correct 4D light field in such local regions. This is similar to the “aperture problem”, wherein a windowed, translated grating appears to move perpendicular to the stripe orientation.

Qualitatively, the front-panel masks exhibit flowing, slit-like barriers aligned perpendicular to the *angular gradient* of the 4D light field (see Figure 3-24), defined using a relative parameterization as

$$\nabla_{ab} L(u, v, a, b) = \left(\frac{\partial L}{\partial a}, \frac{\partial L}{\partial b} \right). \quad (3.34)$$

The rear-panel masks exhibit rotated spatially-multiplexed images similar to conventional parallax barriers. In Figure 3-21 we consider a region centered on a depth discontinuity. Locally, the scene is modeled by two fronto-parallel planes (i.e., a step edge). A 4D light field, containing 3×3 oblique projections, is rendered so the disparity between projections is 10 pixels. The front-panel masks contain perturbed lines that run parallel to the edge (i.e., perpendicular to the angular gradient). Their average spacing equals the angular resolution (3 pixels) and they span a region equal to the product of the disparity and the number of views minus one (± 10 pixels from the edge). The masks exhibit random noise away from the edge, approximating a scene without parallax. Following Lee and Seung [122], Equation 3.33 converges to a local stationary point, but not necessarily the global minimum; as a result, the observed local parallax barriers possess some randomization due to convergence to a local minima. Additional examples are shown in Figure 3-22.

Although we can predict mask structure, we cannot provide an analytic solution. This remains a promising future direction. However, the local parallax barrier interpretation gives intuition into the benefits and limitations of the method. Unlike 2D pinhole arrays,

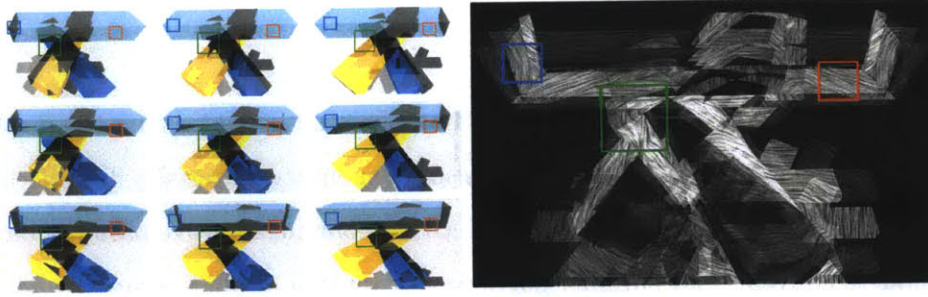


Figure 3-24: Predicting the structure of content-adaptive parallax barriers. (Left) The *blocks* light field containing a stack of three colored glass blocks. From left to right and top to bottom: oblique projections seen as a viewer moves from left to right and top to bottom. (Right) Streamlines of the angular gradient of the light field, evaluated following Section 3.3.1, visualized using line integral convolution [30]. Note that streamline direction predicts the orientation of the local parallax barriers appearing in the right-hand side of Figure 3-22. Consider the windowed region within the blue rectangle (rendered in the same position for all figures on this page). As shown on the left, the light field primarily exhibits horizontal parallax within this window. Thus, the streamlines run vertically on the right; similarly, the corresponding region on the right-hand side of Figure 3-22 exhibits vertically-oriented slits. As described in Section 3.3.1, the resulting local parallax barrier is sufficient to project this windowed region of the light field. Note similar correspondences within the red and green windows.

adaptation creates 1D slits that transmit more light. Consider $N_h \times N_v$ views of a sphere. With pinholes, each front mask is a grid of $N_h \times N_v$ tiles with one transparent pixel. We create local barriers following the angular gradient (e.g., the sphere boundary). Near discontinuities each $N_h \times N_v$ block of the front mask contains slits with an average of no less than $\min(N_h, N_v)$ transparent pixels. Thus, the average achievable brightness increase is $\min(N_h, N_v)$. We conclude that one significant benefit of content-adaptive parallax barriers is to allow simultaneous horizontal and vertical parallax, while preserving the brightness of conventional parallax barriers (i.e., arrays of slits) that support horizontal-only parallax.

3.3.2 Implementation

This section discusses the details of constructing a dual-stacked LCD using modified panels, validates its performance using conventional and content-adaptive parallax barriers, and assesses the performance compared to prior automultiscopic displays.

Hardware

As shown in Figure 3-26, a dual-stacked LCD was constructed using a pair of 1680×1050 Viewsonic FuHzion VX2265wm 120 Hz LCD panels. The panels have a pixel pitch of $282 \mu\text{m}$ and are separated by 1.5 cm. However, as described in Section 3.4.8, masks are displayed at half the native resolution. Thus, for a typical light field with an angular resolution $N_h \times N_v$ of 5×3 views, the prototype supports an $11^\circ \times 7^\circ$ field of view; a viewer sees correct imagery when moving within a frustum with similar apex angles.

The rear layer is an unmodified panel, whereas the front layer is a spatial light modulator (SLM) fashioned by removing the backlight from a second panel. The front polarizing diffuser and rear polarizing film are removed. The front polarizing diffuser is replaced with a transparent polarizer, restoring the spatial light modulation capability of the panel. Without such modifications, the polarizers in the front panel completely attenuate light polarized by the rear panel. Eliminating the redundant rear polarizer of the front panel increases light transmission. The LCD panels are driven separately via DVI links from a dual-head NVIDIA Quadro FX 570 display adapter, automatically synchronizing the display refreshes.

Software

Light fields are rendered with POV-Ray [157] and masks are represented by a series of texture pairs. Each color channel is factorized independently. The displays are driven at 120 Hz with a custom OpenGL application. Gamma compression is applied to ensure mask intensity varies linearly with the encoded value; a gamma value of $\gamma = 2.2$ was measured for our LCDs. Mask optimization uses a multi-threaded C++ implementation written with the POSIX Pthreads API; a single-threaded version is provided with the supplementary code. An Intel Xeon 8-core 3.2 GHz processor with 8 GB of RAM is used for optimization and display. For a typical light field with 5×3 views, each with a resolution of 840×525 pixels, the optimization takes approximately 10 seconds per iteration. In practice, at least 50 iterations are required for the PSNR to exceed 30 dB, leading to an average run-time of

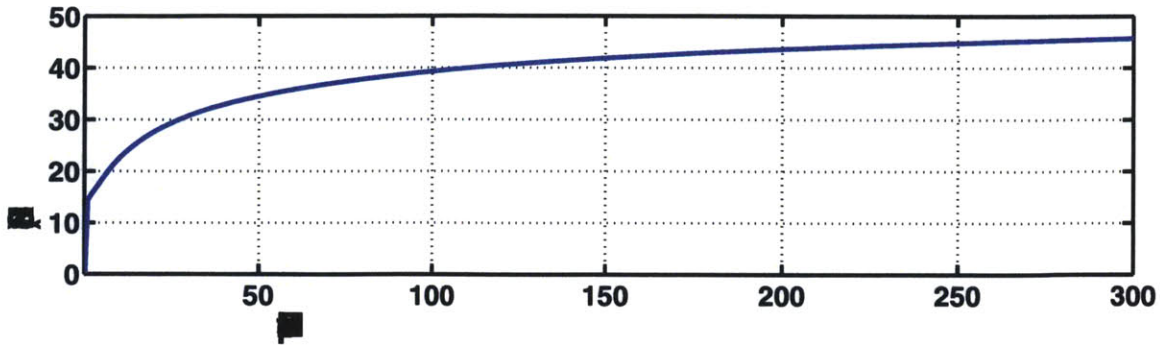


Figure 3-25: Approximation error as a function of NMF iteration. The average PSNR of the reconstruction is plotted for a rank-9 decomposition of the light fields shown in Figures 3-16 and 3-24.

eight minutes per frame (see Figure 3-25). As observed by Zwicker et al. [220], the target light field should be prefiltered to prevent aliasing. Such prefiltering was not applied in our implementation, causing additional artifacts in the right-hand column of Figure 3-27.

3.3.3 Assessment

As with any 3D display, a viewer is concerned with resolution (both spatial and angular), brightness, refresh rate, and reconstruction error. Experiments and simulations assess the performance of content-adaptive parallax barriers, as compared to time-shifted parallax barriers [111]. Two primary benefits result from content-adaptive parallax barriers: increased display brightness and increased display refresh rate; we analyze each in turn.

Increasing Display Brightness

Following Section 3.3.1, content-adaptive parallax barriers appear to exhibit local parallax barrier structure. Using this interpretation we previously predicted an average brightness increase by a maximum factor of $\min(N_h, N_v)$. The supplementary code was used to render a diverse set of light fields containing varying degrees of disparity, contrast, and geometric complexity. Select light fields are shown in Figures 3-16 and 3-24, with additional examples included in the supplementary material and video. As shown in Figure 3-29, the peak signal-

to-noise ratio (PSNR) of the reconstruction is measured as a function of the attempted increase in brightness (i.e., the target light field is multiplied by the desired gain).

Figure 3-29 demonstrates that content-adaptive parallax barriers can increase display brightness, in comparison to time-shifted parallax barriers. These examples use $T = N_h N_v$ time-multiplexed mask pairs (i.e., identical to the number of masks required with time-shifted parallax barriers). As predicted, when brightness is enhanced by the theoretically-predicted factor of $\min(N_h, N_v)$ (i.e., $3\times$ brighter in these examples), the PSNR of the reconstruction remains above 30 dB; for greater increases in brightness, artifacts become readily apparent. We observe that the PSNR is finite (i.e., artifacts are present), even when no increase in brightness is attempted. This indicates a limitation of the current optimization procedure. As described in Section 3.3.1, Equation 3.33 is not guaranteed to converge to the global minimum. Thus, artifacts persist even for the step edge in Figure 3-21. Furthermore, no local parallax barrier (i.e., an array of slits) can represent regions with both horizontal and vertical parallax. For such regions, increasing the brightness by any factor will lead to artifacts under the local parallax barrier interpretation (e.g., the checkerboard corners in Figure 3-27). However, since a PSNR greater than 30 dB is generally accepted for lossy compression, content adaptation achieves significant increases while presenting images that retain the fidelity of the target light field.

To confirm the predicted increase in brightness, a Canon EOS Digital Rebel XSi camera was used as a light meter to quantify brightness for patterns displayed by the dual-stacked LCD. The camera was placed directly in front of the screen. Experimental brightness measurements, with respect to baseline translated pinholes, confirmed an average of $3\times$ brighter for the light fields in Figures 3-16.

Increasing Display Refresh Rate with Compression

Content adaptation can also increase the effective refresh rate of the automultiscopic display. Consider the prototype system, supporting a native 120 Hz refresh rate. In this case, only five masks can be time-multiplexed before the effective refresh rate drops below 24 Hz and

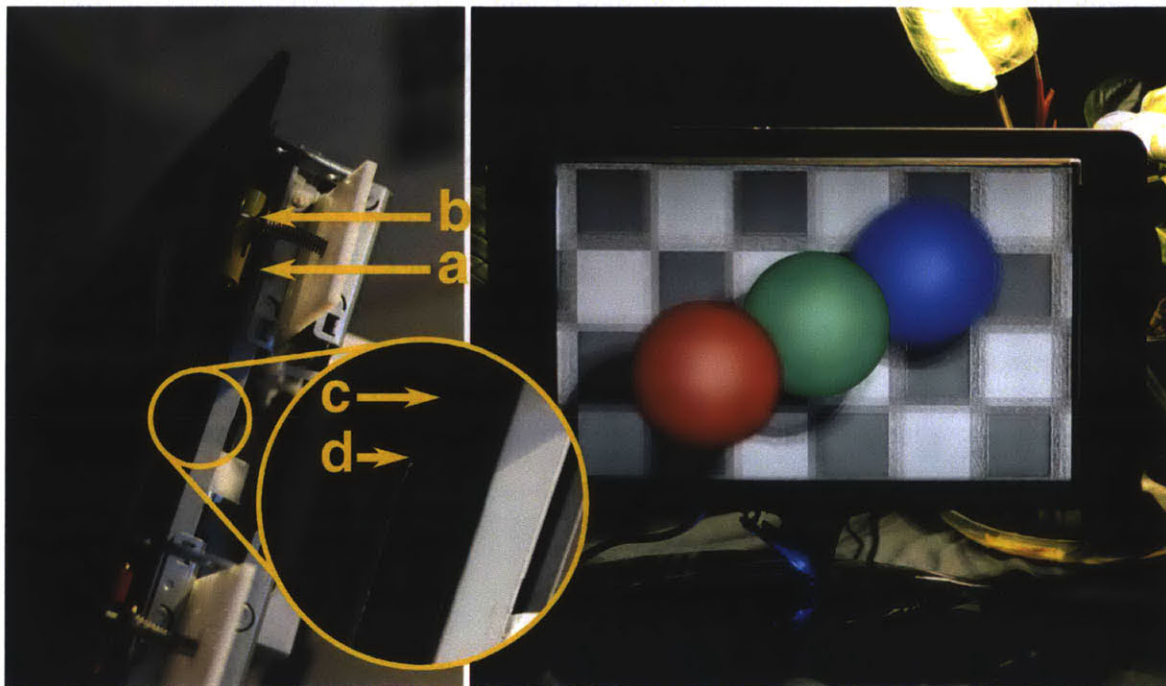


Figure 3-26: Prototype automultiscopic display using dual-stacked LCDs. (Left) Side view of the prototype. From right to left: (a) rear LCD with backlight, (b) spacer, (c) front LCD, and (d) replacement polarizing sheet. (Right) Central view of a synthetic scene rendered with content-adaptive parallax barriers. Video results using this prototype are included in the supplementary material.

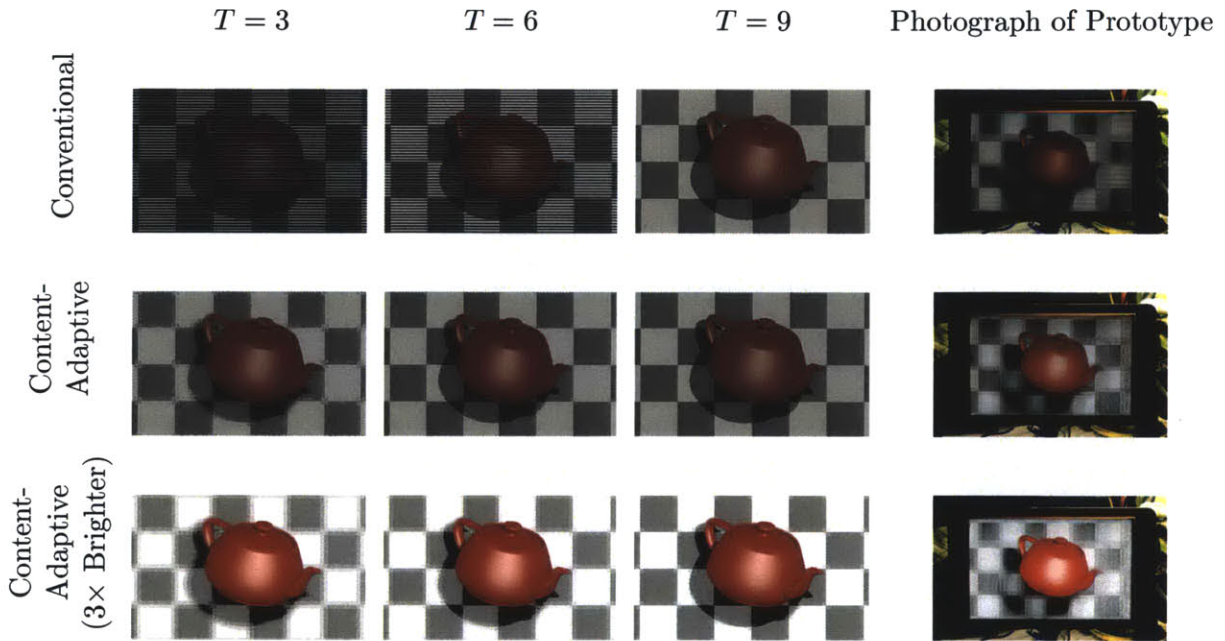


Figure 3-27: Increasing display brightness and refresh rate. Content-adaptive barriers are compared to time-shifted barriers [111], with the exposure normalized so the relative image brightness is consistent with observation. Following Section 3.3.3, light field compression is achieved for $T < N_h N_v$ mask pairs. Reconstructions with three, six, and nine time-multiplexed mask pairs are shown in the first three columns from the left, respectively. Experimental photographs (fourth column) are compared to predicted images (third column). All images correspond to the central oblique view for the light field in Figure 3-16. While content-adaptive barriers produce some high-frequency artifacts, even with nine mask pairs, they can compress the light field with higher PSNR than conventional barriers (see Figure 3-28). As shown along the bottom row, adaptation also allows the brightness to be increased with minimal degradation in image fidelity (see Figure 3-29).

flicker becomes readily apparent. Thus, supporting simultaneous horizontal and vertical parallax becomes challenging.

Fortunately, content-adaptive parallax barriers allow the light field to be compressed using a set of $T < N_h N_v$ mask pairs. Theoretically, rank-1 light fields occur in a single case: when a textured plane is displayed in the plane of the rear LCD panel (i.e., for a light field without any parallax). Experimentally, rank grows (to the number of views $N_h N_v$) as the plane is translated away from the rear LCD. For example, consider the 2D slice of a captured light field shown in Figure 3-18. As described by Chai et al. [33], the separation of a plane from the rear LCD determines the skew of the 2D light field slice. Thus, distant objects require higher-rank approximations. However, in this example, 17 views were reconstructed with a PSNR greater than 30 dB using three mask pairs. Scenes with limited parallax and depth variation require fewer masks.

Figure 3-28 illustrates compression trends typical with content-adaptive parallax barriers. As before, artifacts are present even when $T = N_h N_v$ mask pairs are used; however, in this case the PSNR exceeds 45 dB. Examples of the predicted and experimentally-measured artifacts are shown in Figure 3-27. We conclude that, as with increasing brightness, content adaptation reveals a novel trade-off between automultiscopic display brightness, refresh rate, and reconstruction error. Additional results, including high-resolution stills, masks, and video sequences are included in the supplementary material and video.

Summary

The analysis of dual-stacked LCDs, as rank-constrained light field displays, points the way along a new direction—one in which the display elements themselves are independently optimized for the target light field. While we show one technique for obtaining content-adaptive parallax barriers that optimize optical transmission and effective refresh rate, it is our hope that future work will reveal a wider range of optimization techniques and classes of adaptive masks. The weight matrix may be used to achieve other effects; weights could be selected to support multiple viewers or a wider field of view. This is a timely development,

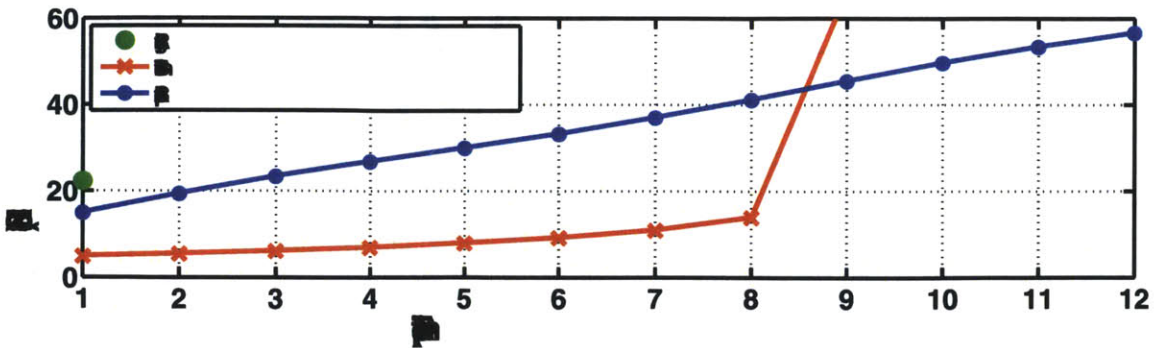


Figure 3-28: Approximation error as a function of decomposition rank. The average PSNR of the reconstruction is plotted for a rank- T decomposition of the light fields in Figures 3-16 and 3-24. For 3×3 views, a theoretical PSNR of infinity is achieved with 9 time-shifted conventional parallax barriers. In comparison, content-adaptive barriers achieve higher PSNR than conventional barriers when fewer frames are used. Experimental and predicted images with varying degrees of compression are shown in Figure 3-27.

as the power and availability of computation has made real-time optimization a reality in many fields. In addition, content-adaptive parallax barriers will benefit from the trend of increasing LCD refresh rates. Refined cost functions, such as those that incorporate human perceptual effects, may provide superior results.

Our optimization is reminiscent of that used with computer generated holograms [173], as well as band moiré images [79]. It may be possible to obtain analytical interpretations of our results, possibly through a frequency-domain analysis; the structure of the masks we obtain appear to mimic local parallax barriers and suggest a broadband nature, reminiscent of the masks used in heterodyne light field cameras [194, 120, 81].

Any commercial implementation must address the current prototype limitations, including: moiré, color-channel crosstalk, and flicker. Our theory only applies to dual-layer displays, yet extensions allowing more layers may reveal additional benefits, including increased fidelity and reduced requirements on the number of mask pairs. Generalizing to arbitrary numbers of spatial light modulators, volumetric occluders could be designed to modify a uniform backlight to reproduce a light field; such occluders may function as the display equivalent of the volumetric occluders used for light field capture with reference structure tomography [25].

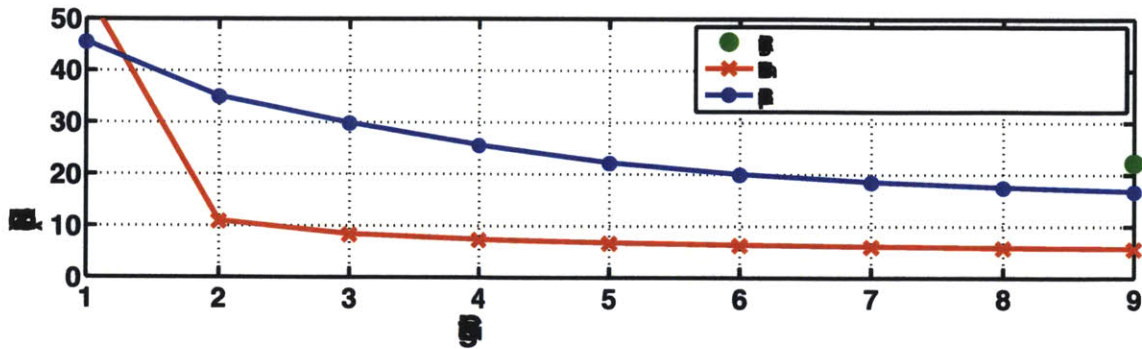


Figure 3-29: Approximation error as a function of gain in brightness. The average PSNR of the reconstruction is plotted for a rank-9 decomposition of the light fields shown in Figures 3-16 and 3-24. For time-shifted parallax barriers, transmission can be increased either by enlarging slits/pinholes or by brightening the rear LCD. The latter is considered here, however simulations of the former also confirm time-shifted parallax barriers cannot achieve a PSNR greater than 15 dB when increasing brightness by a factor greater than two.

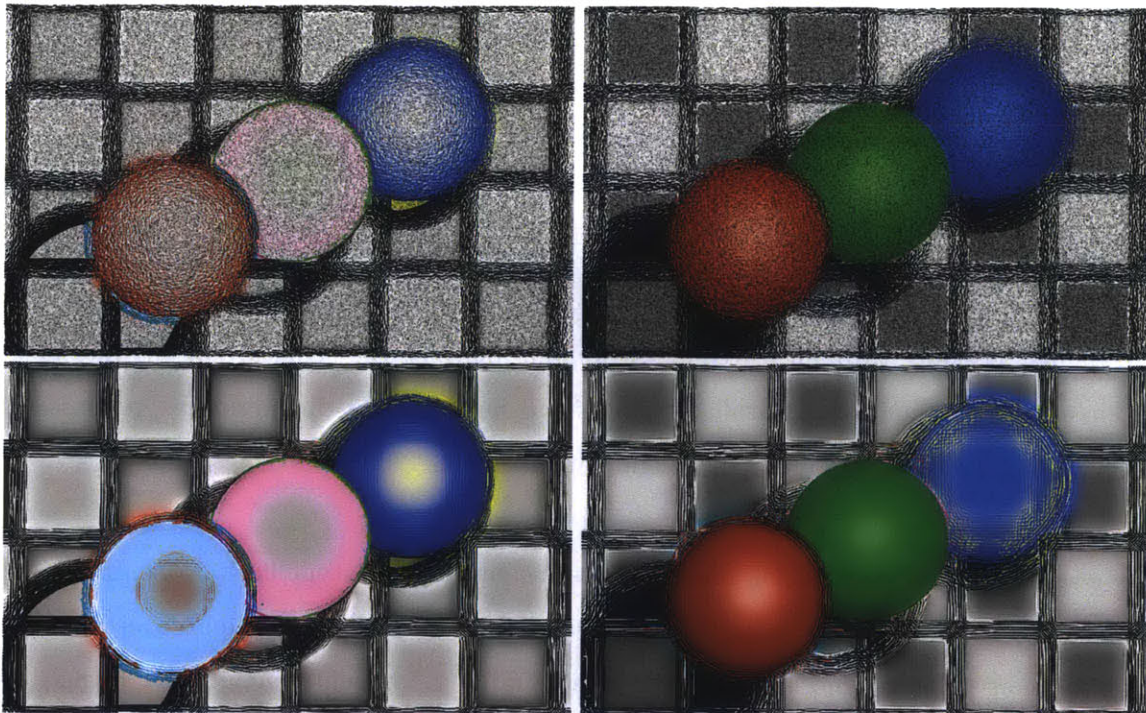


Figure 3-30: Regularized NMF for smooth masks. (Top) The *spheres* light field (see Figure 3-26) is decomposed via Equation 3.33. The masks contain high-frequency patterns, even in uniform regions without parallax. (Bottom) Spatial smoothness is achieved by convolving the masks with a Gaussian filter after every 10 iterations. The filter standard deviation is reduced over time, allowing high-frequencies to appear only in later iterations. Note the close agreement with the local barrier interpretation in Section 3.3.1.

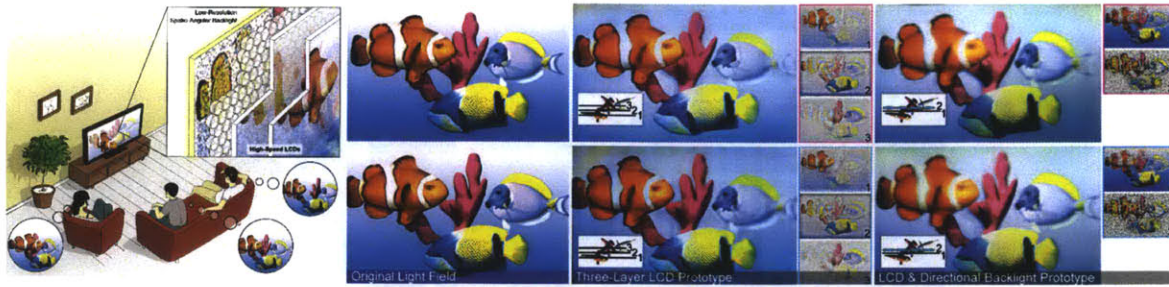


Figure 3-31: Wide field of view glasses-free 3D display using Tensor Displays. (Left) We introduce a new family of light field displays, dubbed Tensor Displays, comprised of stacks of light-attenuating layers (e.g., multilayer LCDs). Rapid temporal modulation of the layers is exploited, in concert with directional backlighting, to allow large separations between viewers. (Right) From left to right: target light field view, photograph of three-layer LCD with uniform backlighting, and photograph of single LCD with directional backlighting. Layers are shown to the right of each photograph. The upper and lower rows depict perspectives seen to the left and right of the display, respectively.

3.4 Tensor Displays: A Compressive Display Framework

3.4.1 Tensor Display Framework

This section presents a unifying framework for depicting arbitrary light fields using Tensor Displays. First, we introduce a tensor representation for multilayer displays illuminated by a uniform backlight. The light field emitted by an N -layer, M -frame display is represented by a sparse set of non-zero elements restricted to a plane within an N^{th} -order, rank- M tensor. Second, we show that this tensor representation allows for optimal decomposition of a light field into time-multiplexed, light-attenuating layers using nonnegative tensor factorization (NTF). Third, we demonstrate that our tensor representation also allows optimization of multilayer displays illuminated by a directional backlight. We conclude by interpreting the structure of Tensor Display decompositions.

Representing Multilayer Displays with Tensors

As shown in Figure 3-33, Tensor Displays consist of a stack of N light-attenuating layers illuminated by either a conventional uniform backlight or a directional backlight. For full generality, we assume that display layers support synchronized, high-speed temporal modulation, such that an observer perceives the time average of an M -frame multilayer mask

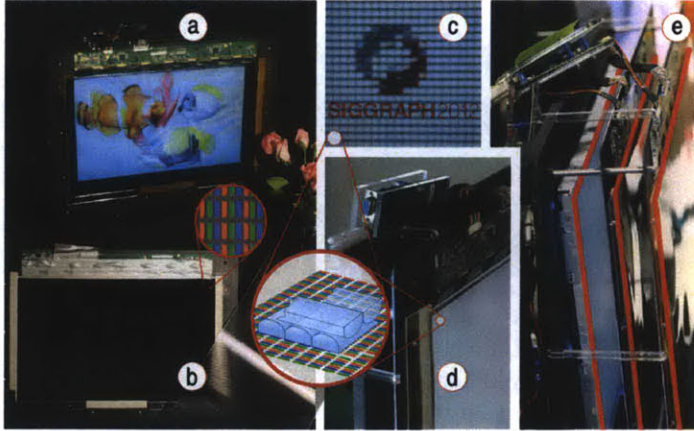


Figure 3-32: The Tensor Display prototype. (a) The prototype configured as a three-layer display, photographed outside the optimized viewing zone so layer patterns are individually visible. (b) An LCD layer mounted on an aluminum frame (left) and a lenticular sheet (right). (c) The directional backlight, consisting of two crossed lenticular sheets on top of the rear LCD (inset). High-resolution text is shown on an LCD layer suspended in front of the directional backlight. (d) The single-layer directional backlight configuration. (e) The three-layer configuration, with layers highlighted in red.

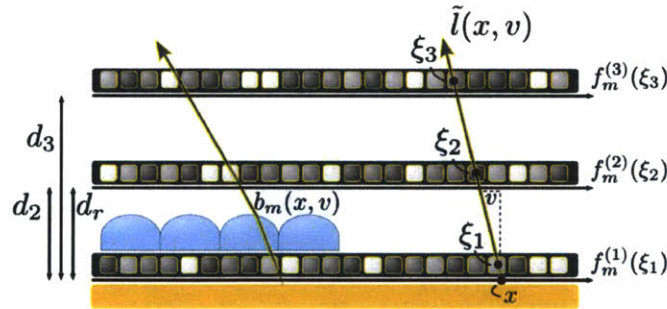


Figure 3-33: Tensor Display coordinates. A stack of N light-attenuating layers is illuminated by a uniform or directional backlight (here depicted as a lenslet array affixed to the rear display).

sequence. We consider 2D light fields and 1D layers in the following analysis, with the extension to 4D light fields and 2D layers covered in Section 3.4.1. A relative two-plane light field parameterization $l(x, v)$ is adopted, shown in Figure 3-33, where v denotes the point of intersection of the ray (x, v) with a plane located a distance d_r from the x -axis, expressed relative to x [33, 51]. In the following analysis we assume familiarity with multilinear algebra, particularly tensor notation; consult Kolda and Bader [114] for a review.

Static Multilayer Displays Consider a fixed stack of N light-attenuating layers (i.e., one that does not support temporal variation of the mask patterns). When illuminated by a uniform backlight with unit radiance, the emitted light field $\tilde{l}(x, v)$ is given by the following expression:

$$\tilde{l}(x, v; N) = \prod_{n=1}^N f^{(n)}(x + (d_n/d_r)v), \quad (3.35)$$

where $f^{(n)}(\xi_n) \in [0, 1]$ is the transmittance at the point ξ_n of layer n , separated a distance d_n from the x -axis. Consider a three-layer configuration, with the transmittances for the rear, middle, and front layers given by $f(\xi_1)$, $g(\xi_2)$, and $h(\xi_3)$, respectively. Equation 3.35 gives the following expression for the emitted light field.

$$\tilde{l}(x, v) = f(\xi_1) g(\xi_2) h(\xi_3), \text{ for } \xi_n = x + (d_n/d_r)v \quad (3.36)$$

We observe that the emitted light field $\tilde{l}(x, v)$ can be represented as the restriction of the function

$$\tilde{t}(\xi_1, \xi_2, \xi_3) = f(\xi_1) g(\xi_2) h(\xi_3), \quad (3.37)$$

defined in the three-dimensional Euclidean space \mathbb{R}^3 spanned by $\{\xi_1, \xi_2, \xi_3\}$, to the two-dimensional subspace defined by the equation $\alpha\xi_1 + \beta\xi_2 + \gamma\xi_3 = 0$, with

$$\alpha = d_3 - d_2, \quad \beta = d_1 - d_3, \quad \gamma = d_2 - d_1. \quad (3.38)$$

Thus, as shown at the top of Figure 3-34, elements of the emitted light field $\tilde{l}(x, v)$ are restricted to the plane corresponding to Equation 3.38.

For the general case with $N > 3$ layers, the emitted light field $\tilde{l}(x, v)$ can also be represented as the restriction of the function $\tilde{t}(\xi_1, \xi_2, \dots, \xi_N) = \prod_{n=1}^N f^{(n)}(\xi_n)$, defined on \mathbb{R}^N , to a plane.

In practice, each layer has discrete pixels with constant transmittances rather than continuously-varying opacities. As a result, we tabulate the transmittance $f_{i_n}^{(n)}$ at each pixel i_n within the vector $\mathbf{f}^{(n)}$. As shown in Figure 3-33, each light field ray (x, v) can be equivalently parameterized by the corresponding points of intersection $\{\xi_1, \xi_2, \dots, \xi_N\}$ with each layer. For a three-layer display with discrete pixels, the intensity of the emitted light field $\tilde{l}(\xi_1, \xi_2, \xi_3)$ is

approximated by the product $f_i g_j h_k$, where $\{i, j, k\}$ denote the pixel indices nearest to the points of intersection $\{\xi_1, \xi_2, \xi_3\}$. With this parameterization we observe that Equation 3.37 can be represented in discrete coordinates as a 3rd-order, rank-1 tensor $\tilde{\mathcal{J}}$, given by

$$\tilde{\mathcal{J}} = \mathbf{f} \circ \mathbf{g} \circ \mathbf{h}, \text{ such that } \tilde{t}_{ijk} = f_i g_j h_k, \quad (3.39)$$

where \circ is the vector outer product. Note that only a subset of tensor elements \tilde{t}_{ijk} correspond to valid light field rays; most tensor elements correspond to “non-physical” rays (i.e., ones that spontaneously change position or direction after passing through a layer). To address this limitation of our tensor representation, we further define a sparse, binary-valued weight tensor \mathcal{W} such that the emitted light field tensor $\tilde{\mathcal{L}}$ is given by the following expression:

$$\tilde{\mathcal{L}} = \mathcal{W} \circledast \tilde{\mathcal{J}}, \text{ for } w_{ijk} = \begin{cases} 1 & \text{if } \{i, j, k\} \text{ gives a light field ray,} \\ 0 & \text{otherwise,} \end{cases} \quad (3.40)$$

where \circledast is the Hadamard (elementwise) product. Following Figure 3-34, non-zero elements of $\tilde{\mathcal{L}}$ are close to the plane defined by Equation 3.38. We conclude that tensors provide sparse, memory-efficient representations for static N -layer displays; as described in Section 3.4.3, only the non-zero elements of $\tilde{\mathcal{L}}$ must be stored.

Time-Multiplexed Multilayer Displays Static multilayer displays have finite degrees of freedom. Artifacts, resulting from limited depths of field and fields of view, persist in the emitted light field, as observed by Gotoda [64, 65] and Wetzstein et al. [205]. Holroyd et al. [87]. The degrees of freedom must be increased to mitigate artifacts, typically observed as blur. We propose exploiting rapid temporal modulation, such that the observer perceives the average of an M -frame sequence. Generalizing Equation 3.35, the emitted light field $\tilde{l}(x, v)$ is given by

$$\tilde{l}(x, v; N, M) = \frac{1}{M} \sum_{m=1}^M \prod_{n=1}^N f_m^{(n)}(x + (d_n/d_r)v), \quad (3.41)$$

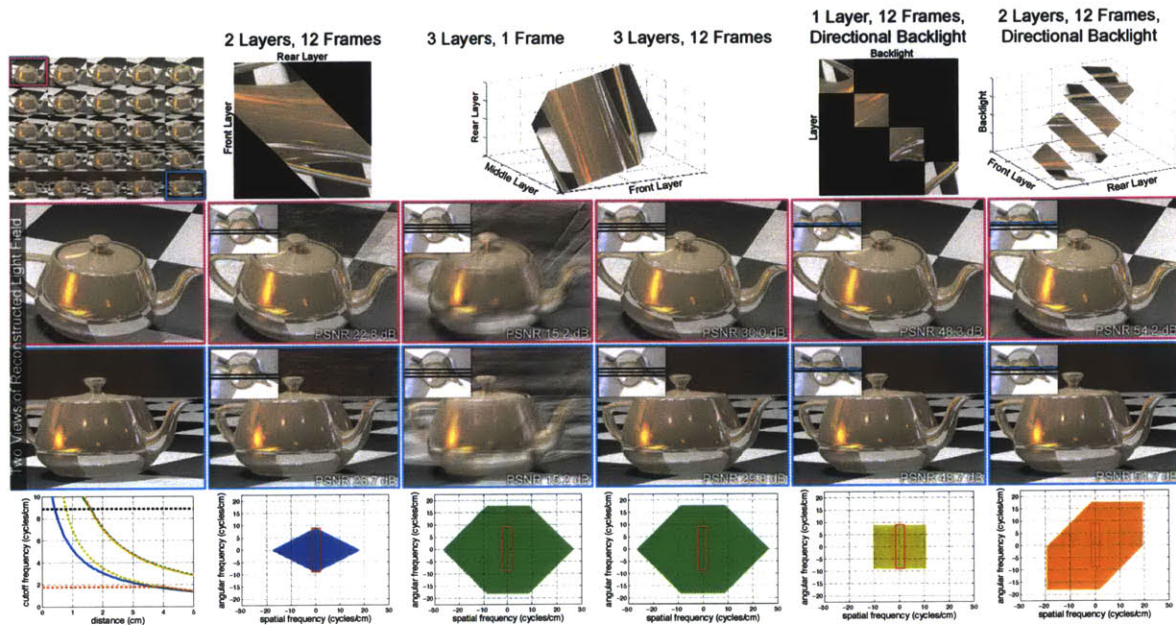


Figure 3-34: Overview of Tensor Displays. (First Row, Left) A target light field for a teapot, rendered as 5×5 views with a 20° field of view. (First Row, Right) Visualizations of the light field, as restricted to the plane within the display tensor \mathcal{T} given by Equation 3.38. Five architectures are compared from left to right: two-layer, 12-frame display, static three-layer display, three-layer, 12-frame tensor display, and single-layer and two-layer Tensor Displays using directional backlights with 12 frames (spatial backlight resolution is a quarter that of each layer). (Second and Third Rows) Two reconstructed views using each display. Note that time-multiplexing, as allowed by Tensor Displays, significantly reduces artifacts observed with the static three-layer configuration. (Fourth Row, Left) Upper bound on depths of field (similar to Figure 3-36). (Fourth Row, Right) Upper bound on the spatio-angular bandwidth for each display, as described in Section 3.4.2. These results demonstrate increased depth of field for Tensor Displays, relative to prior work, as indicated by reduced artifacts for the checkerboard and reflections in the teapot.

where $f_m^{(n)}(\xi_n)$ is the transmittance at the point ξ_n of layer n during frame m . Let columns of the matrix $\mathbf{F}^{(n)} = [\mathbf{f}_1^{(n)} \mathbf{f}_2^{(n)} \dots \mathbf{f}_M^{(n)}]$ define the sequence of M masks displayed on layer n . For a three-layer display, Equation 3.41 can be represented in discrete coordinates as a 3rd-order, rank- M tensor $\tilde{\mathcal{J}}$ given by

$$\tilde{\mathcal{J}} = [\mathbf{F}, \mathbf{G}, \mathbf{H}] \equiv \frac{1}{M} \sum_{m=1}^M \mathbf{f}_m \circ \mathbf{g}_m \circ \mathbf{h}_m, \quad (3.42)$$

where matrices enclosed by double square brackets correspond to the *CP decomposition* of a tensor into a sum of rank-1 tensors [39]. The CP decomposition is equivalent to CANDECOMP (canonical decomposition) and PARAFAC (parallel factors), with elements of the tensor $\tilde{\mathcal{J}}$ given by $\tilde{t}_{ijk} = \frac{1}{M} \sum_{m=1}^M f_{im} g_{jm} h_{km}$ [114]. For the general case with N light-attenuating layers and M time-multiplexed frames, we observe that the emitted light field can be represented as an N^{th} -order, rank- M tensor $\tilde{\mathcal{J}} = [[\mathbf{F}^{(1)}, \mathbf{F}^{(2)}, \dots, \mathbf{F}^{(N)}]]$.

Synthesizing Light Fields

Light field synthesis with time-multiplexed, multilayer displays requires decomposing a target light field $l(x, v)$ into an M -frame sequence of N transmittance functions $f_m^{(n)}(\xi_n)$. This can be formulated as the following constrained nonlinear least squares problem:

$$\arg \min_{f_m^{(n)}(\xi_n)} \int_{\mathcal{V}} \int_{\mathcal{X}} \left(l(x, v) - \tilde{l}(x, v) \right)^2 dx dv, \text{ for } 0 \leq f_m^{(n)}(\xi_n) \leq 1, \quad (3.43)$$

where $\tilde{l}(x, v)$ is the emitted light field, given by Equation 3.41, and \mathcal{X} and \mathcal{V} denote the intervals $[x_{\min}, x_{\max}]$ and $[v_{\min}, v_{\max}]$.

The tensor representation introduced in Section 3.4.1 provides an efficient means for solving Equation 3.43. Using this representation for a three-layer configuration with discrete coordinates, the objective function is expressed as

$$\arg \min_{\mathbf{F}, \mathbf{G}, \mathbf{H}} \|\mathcal{L} - \mathcal{W} \circledast [\mathbf{F}, \mathbf{G}, \mathbf{H}]\|^2, \text{ for } 0 \leq \mathbf{F}, \mathbf{G}, \mathbf{H} \leq 1, \quad (3.44)$$

where \mathcal{L} is the target light field tensor, obtained by assigning the target light field $l(x, v)$ to the plane defined by Equation 3.38, and $\|\mathcal{X}\|^2 = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K x_{ijk}^2$ is the squared tensor norm of \mathcal{X} . We observe that this expression can be solved by applying weighted nonnegative tensor factorization (NTF). Following Cichocki et al. [39], a broad set of procedures have emerged for the solution of NTF problems. In this paper we use multiplicative update rules that extend the weighted nonnegative matrix factorization (NMF) procedure proposed by Blondel et al. [21] to higher-order tensors. For a three-layer display, these update rules have the following forms:

$$\mathbf{F} \leftarrow \mathbf{F} \circledast \left(\frac{(\mathbf{W}_{(1)} \circledast \mathbf{L}_{(1)})(\mathbf{H} \odot \mathbf{G})}{(\mathbf{W}_{(1)} \circledast (\mathbf{F}(\mathbf{H} \odot \mathbf{G})^\top))(\mathbf{H} \odot \mathbf{G})} \right) \quad (3.45)$$

$$\mathbf{G} \leftarrow \mathbf{G} \circledast \left(\frac{(\mathbf{W}_{(2)} \circledast \mathbf{L}_{(2)})(\mathbf{H} \odot \mathbf{F})}{(\mathbf{W}_{(2)} \circledast (\mathbf{G}(\mathbf{H} \odot \mathbf{F})^\top))(\mathbf{H} \odot \mathbf{F})} \right) \quad (3.46)$$

$$\mathbf{H} \leftarrow \mathbf{H} \circledast \left(\frac{(\mathbf{W}_{(3)} \circledast \mathbf{L}_{(3)})(\mathbf{G} \odot \mathbf{F})}{(\mathbf{W}_{(3)} \circledast (\mathbf{H}(\mathbf{G} \odot \mathbf{F})^\top))(\mathbf{G} \odot \mathbf{F})} \right) \quad (3.47)$$

In these expressions \odot is the Khatri-Rao product, defined for a pair of matrices $\mathbf{A} \in \mathbb{R}^{I \times K}$ and $\mathbf{B} \in \mathbb{R}^{J \times K}$, such that

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1 \quad \mathbf{a}_2 \otimes \mathbf{b}_2 \quad \cdots \quad \mathbf{a}_K \otimes \mathbf{b}_K], \quad (3.48)$$

where \otimes is the Kronecker product and \mathbf{a}_i and \mathbf{b}_j denote the i^{th} and j^{th} columns of \mathbf{A} and \mathbf{B} , respectively. These update equations also make use of the tensor matricization (unfolding) operation, defined such that $\mathbf{X}_{(n)}$ arranges the mode- n fibers of \mathcal{X} to be columns of the resulting matrix. We observe, for two layers, these weighted NTF update rules reduce to the weighted NMF update rules used by Blondel et al. [21] and Section 3.3.

For the general case with N light-attenuating layers and M frames, we observe that Equation 3.44 has the following form:

$$\arg \min_{\mathbf{F}^{(n)}} \left\| \mathcal{L} - \mathcal{W} \circledast \tilde{\mathcal{J}} \right\|^2, \text{ for } 0 \leq \mathbf{F}^{(n)} \leq 1, \quad (3.49)$$

where $\tilde{\mathcal{J}} = [[\mathbf{F}^{(1)}, \mathbf{F}^{(2)}, \dots, \mathbf{F}^{(N)}]]$. Similarly, the update rules are generalized such that

$$\mathbf{F}^{(n)} \leftarrow \mathbf{F}^{(n)} \circledast \left(\frac{(\mathbf{W}_{(n)} \circledast \mathbf{L}_{(n)}) \mathbf{F}_{\odot}^n}{(\mathbf{W}_{(n)} \circledast (\mathbf{F}^{(n)} (\mathbf{F}_{\odot}^n)^{\top})) \mathbf{F}_{\odot}^n} \right), \quad (3.50)$$

where \mathbf{F}_{\odot}^n is defined by the following expression:

$$\mathbf{F}_{\odot}^n \equiv \mathbf{F}^{(N)} \odot \dots \odot \mathbf{F}^{(n+1)} \odot \mathbf{F}^{(n-1)} \odot \dots \odot \mathbf{F}^{(1)}. \quad (3.51)$$

4D light fields and 2D layers require vectorizing the 2D layer transmittances, giving a similar set of transmittance vectors $\{\mathbf{f}_m^{(n)}\}$. Following standard practice [39], values are clamped to the feasible range after each iteration of Equation 3.50.

In summary, our tensor representation allows for the decomposition of a target light field into a set of time-multiplexed, light-attenuating layers. As described in Section 3.3.2, the multiplicative update rules allow an efficient, GPU-based implementation that achieves interactive refresh rates with multilayer LCDs.

Incorporating Directional Backlighting

As shown in the fourth column of Figure 3-34, time multiplexing significantly reduces artifacts observed with multilayer displays, as quantified by the peak signal-to-noise ratio (PSNR). Yet, such displays are still restricted to relatively narrow fields of view (i.e., $\lesssim 20^\circ$). Expanding the field of view requires further increasing the refresh rate—a solution that may be precluded by the underlying display hardware. In this section we propose an alternate approach for achieving wider fields of view: replacing conventional uniform backlighting with time-multiplexed directional backlighting.

A directional backlight is equivalent to a low-resolution light field display. In this analysis we assume the directional backlight has significantly lower spatial resolution, but equivalent angular resolution and field of view, as compared to the target light field $l(x, v)$. Thus, our goal is to primarily enhance the spatial resolution by covering a low-resolution light field display with an N -layer stack of light-attenuating layers. Generalizing Equation 3.41, the light field emitted by such a display architecture is given by the following expression:

$$\tilde{l}(x, v) = \frac{1}{M} \sum_{m=1}^M b_m(x, v) \prod_{n=1}^N f_m^{(n)}(x + (d_n/d_r)v), \quad (3.52)$$

where $b_m(x, v)$ denotes the light field emitted by the backlight during frame m . Let \mathbf{B} denote the discrete backlight light field, such that b_{as} corresponds to pixel s of view a . The backlight light field can be equivalently represented as a vector \mathbf{b} , defined as follows.

$$\mathbf{b} = [\mathbf{b}_1^T \ \mathbf{b}_2^T \ \cdots \ \mathbf{b}_S^T]^T, \text{ for } \mathbf{b}_s = [b_{1s} \ b_{2s} \ \cdots \ b_{As}]^T \quad (3.53)$$

Using this parameterization, Equation 3.52 can be represented in discrete coordinates as an $N + 1$ -order, rank- M tensor $\tilde{\mathcal{J}}$, given by

$$\tilde{\mathcal{J}} = \frac{1}{M} \sum_{m=1}^M \mathbf{b}_m \circ \mathbf{f}_m^{(1)} \circ \mathbf{f}_m^{(2)} \circ \cdots \circ \mathbf{f}_m^{(N)}, \quad (3.54)$$

where tensor element $\tilde{l}_{ij_1 j_2 \cdots j_N} = \frac{1}{M} \sum_{m=1}^M b_{im} \prod_{n=1}^N f_{j_n m}^{(n)}$. Since Equations 3.42 and 3.54 are similar, NTF can also be applied to optimize multilayer displays with directional backlighting.

As shown in Figure 3-34, directional backlighting allows multilayer displays to achieve wide fields of view, even with a single high-speed, light-attenuating layer. In summary, our tensor representation for multilayer displays provides a computationally-efficient optimization scheme encompassing a wide variety of display architectures. While providing the first

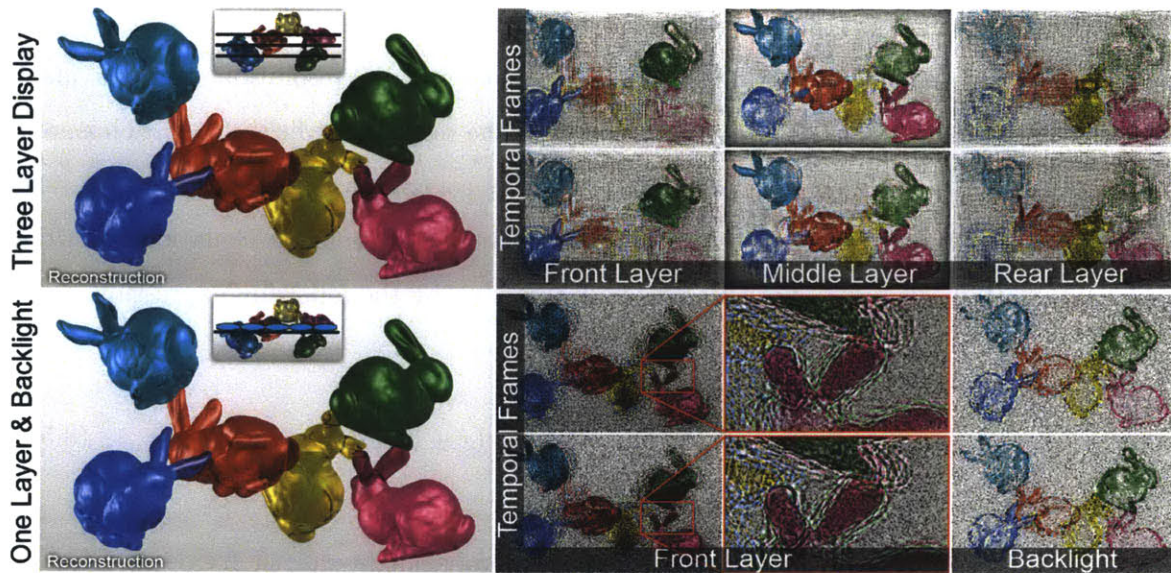


Figure 3-35: Interpreting Tensor Display decompositions. Reconstruction and decomposition results are compared for a three-layer display with uniform backlighting (top) and a single-layer display using a directional backlight (bottom). The structures of the multilayer, multiframe decompositions are discussed in Section 3.4.1.

method for joint multilayer, multiframe decompositions, this framework also naturally extends to emerging directional backlighting. In the following sections we further analyze the theoretical and practical benefits of display architectures supported by the Tensor Display Framework.

Interpreting Tensor Display Decompositions

Tensor Displays exploit the additional degrees of freedom arising from multiple layers and frames to achieve high-fidelity light field reconstructions. The benefits of joint multilayer, multiframe decompositions are demonstrated in Figure 3-34. However, these results do not provide intuition into the underlying structure of the decomposed layers. What spatial and temporal modulation patterns give rise to accurate reconstructions? We examine the decompositions for two architectures: a three-layer display with uniform backlighting and a single-layer display with directional backlighting.

Multilayer decompositions are shown at the top of Figure 3-35. We observe that objects close to the display appear sectioned across layers. The green bunny maps primarily to

the front layer, with residual details assigned to other layers. Similar sectioning behaviors have been observed with multilayer-only decompositions, including those of Gotoda [64] and Wetzstein et al. [205]. Unlike these works, our joint multilayer, multiframe decompositions produce additional time-varying, high-frequency patterns that appear across all layers and resemble content-adaptive parallax barriers shown in Section 3.3.

Decompositions for a single-layer display with directional backlighting are shown at the bottom of Figure 3-35. We observe that the front layer contains the view-independent portions of the scene, with flowing, slit-like patterns appearing around regions with view-dependent features. The directional backlight is primarily comprised of view-dependent features, such as objects extending from the physical display enclosure (e.g., the green bunny).

Tensor Display decompositions exhibit predictable structures, whose arrangement arise from the specific display configuration. A natural direction for future work is to more closely assess these structures for promising architectures, such as the single layer with directional backlighting, in the hope that heuristically-defined methods may achieve similar fidelity with reduced computation.

3.4.2 Analysis

This section analyzes the performance of Tensor Displays, focusing on the quantitative benefits of additional layers, additional frames, and directional backlighting. First, we derive the upper bound on the depth of field for any Tensor Display. This allows comparison of alternative display architectures. The upper bound also provides antialiasing prefilters for each design. Second, we assess the interdependence of display design and decomposition algorithm parameters, documenting their influence on reconstructed image fidelity.

Depth of Field

The performance of an automultiscopic display can be quantified by its depth of field: an expression for the maximum spatial frequency $\omega_{\xi_{\max}}$ that can be depicted in a plane oriented

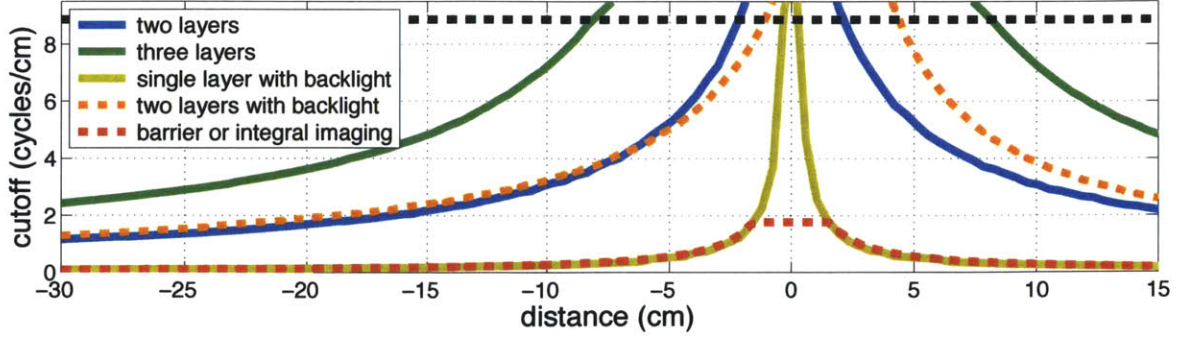


Figure 3-36: Comparison of upper bounds on depth of field for parallax barriers and integral imaging (red), two-layer (blue) and three-layer (green) displays with uniform backlighting, and single-layer (yellow) and two-layer (orange) displays with directional backlighting. The dashed black line denotes the spatial cutoff frequency for each layer. Display parameters correspond to the prototypes described in Section 3.4.4.

parallel to the screen and separated by a distance d_o . As described by Zwicker et al. [219], this expression is derived using a frequency-domain analysis of the emitted light field $\tilde{l}(x, v)$. Taking the 2D Fourier transform of Equation 3.52 yields the following expression for the emitted light field spectrum $\hat{l}(\omega_x, \omega_v)$:

$$\hat{l}(\omega_x, \omega_v) = \frac{1}{M} \sum_{m=1}^M \hat{b}_m(\omega_x, \omega_v) * \left[\sum_{n=1}^N \hat{f}_m^{(n)}(\omega_x) \delta(\omega_v - (d_n/d_r)\omega_x) \right], \quad (3.55)$$

where ω_x and ω_v are the spatial and angular frequencies, $*$ denotes convolution, and the repeated convolution operator is defined as

$$\sum_{n=1}^N \hat{f}_m^{(n)}(\omega_x, \omega_v) \equiv \hat{f}_m^{(1)}(\omega_x, \omega_v) * \dots * \hat{f}_m^{(N)}(\omega_x, \omega_v). \quad (3.56)$$

For uniform backlighting, the backlight spectrum $\hat{b}_m(\omega_x, \omega_v) = \delta(\omega_x, \omega_v)$, the Dirac delta function, reducing Equation 3.55 to the expression derived for multilayer displays by Wetstein et al. [205].

The spectral support of a Tensor Display is the region of non-zero values in the emitted light field spectrum, for all possible layer masks and backlight illumination patterns. Following Chai et al. [33], the spectral support for the light field reflected by a diffuse surface is the

line $\omega_v = (d_o/d_r)\omega_x$. Intersecting this line with the spectral support for a given display provides a geometric construction for the upper bound on the depth of field. For example, the emitted light field spectrum for a parallax barrier or integral imaging display is non-zero only for $|\omega_x| \leq 1/(2\Delta x)$ and $|\omega_v| \leq 1/(2\Delta v)$ (e.g., the red boxes shown in Figure 3-34), where Δx and Δv are the spatial and angular sampling rates, respectively. In practice, the spatial sampling rate Δx is the spacing between barrier slits/pinholes or lenslets. The geometric construction yields the following expression for the depth of field:

$$\omega_{\xi_{\max}}(d_o) = \begin{cases} \frac{1}{2\Delta x} & \text{for } |d_o| \leq d_r \left(\frac{\Delta x}{\Delta v}\right), \\ \frac{d_r}{2|d_o|\Delta v} & \text{otherwise,} \end{cases} \quad (3.57)$$

where $\Delta v = (2d_r/A) \tan(\alpha/2)$ with A views and field of view α .

The geometric construction provides an upper bound on the depth of field for any Tensor Display architecture. Consider a two-layer display with uniform backlighting, with the layers separated by a distance Δd and $\omega_0 = 1/(2p)$ denoting the maximum spatial frequency for each layer with pixel pitch p . Equation 3.55 defines the light field spectrum, where $d_1 = -\Delta d/2$ and $d_2 = \Delta d/2$. As shown in Figure 3-34, a diamond-shaped region bounds the spectral support for any two-layer display. The spatial cutoff frequency $\omega_{\xi_{\max}}$ is again found by intersecting the line $\omega_v = (d_o/d_r)\omega_x$ with the boundary of the spectral support, yielding the following upper bound on the depth of field for any two-layer display.

$$\omega_{\xi_{\max}}(d_o) = \left(\frac{2\Delta d}{\Delta d + 2|d_o|} \right) \omega_0 \quad (3.58)$$

In Section 3.4.3 we compare two Tensor Display architectures: a three-layer display with uniform backlighting vs. a single-layer display with directional backlighting. Using the previously described geometric construction, the depth of field for a three-layer display with uniform backlighting and equally-spaced layers is given by

$$\omega_{\xi_{\max}}(d_o) = \begin{cases} \left(\frac{3\Delta d}{\Delta d + |d_o|}\right) \omega_0 & \text{for } |d_o| \leq 2\Delta d, \\ \left(\frac{2\Delta d}{|d_o|}\right) \omega_0 & \text{otherwise,} \end{cases} \quad (3.59)$$

where Equation 3.55 is again applied to find the spectral support, with $d_1 = -\Delta d$, $d_2 = 0$, and $d_3 = \Delta d$. As shown in the fourth row of Figure 3-34, the spectral support for a three-layer display exceeds that of a similar parallax barrier or integral imaging display, leading to the increased depth of field observed in Figure 3-36.

As described in Section 3.4.1, incorporating directional backlighting can significantly expand the field of view. The depth of field for a single-layer display using directional backlighting is obtained by a similar geometric construction. We assume the directional backlight implements a low-resolution light field display, such that $\hat{b}_m(\omega_x, \omega_v)$ has non-zero support for $|\omega_x| \leq 1/(2\Delta x)$ and $|\omega_v| \leq 1/(2\Delta v)$. This yields the following depth of field expression:

$$\omega_{\xi_{\max}}(d_o) = \begin{cases} \frac{1}{2\Delta x} + \omega_0 & \text{for } |d_o| \leq d_r \left(\frac{\Delta x}{\Delta v + 2\Delta x \Delta v \omega_0}\right), \\ \frac{d_r}{2|d_o|\Delta v} & \text{otherwise,} \end{cases} \quad (3.60)$$

where ω_0 again denotes the spatial cutoff frequency for the layer. As shown in Figure 3-36, the addition of a single light-attenuating layer significantly increases the spatial resolution for a conventional parallax barrier or integral imaging display, particularly near the display surface. However, far from the display, the depth of field is identical to these conventional automultiscopic displays.

Our analysis indicates a promising application for Tensor Displays: increased depth of field can be achieved by covering any low-resolution light field display with time-multiplexed, light-attenuating layers. In this analysis, we assume continuously-varying layer transmittances; a promising research direction is to characterize the upper bound with discrete pixels. However, with our analysis, we observe that static and time-multiplexed Tensor Displays have identical spectral supports (i.e., averaging over an M -frame sequence does not alter the support via Equation 3.55). Yet, as depicted in the second and third rows of

Figure 3-34, time multiplexing significantly reduces artifacts. We attribute this to the additional degrees of freedom allowed with time multiplexing. While the upper bound may be identical, in practice it cannot be achieved with static methods, motivating Tensor Displays for joint multilayer, multiframe decompositions capable of approaching the upper bound.

Design Trade-Offs

One of the main benefits of Tensor Displays is to open a design trade space not accessible to prior automultiscopic displays. Existing multilayer-only or multiframe-only decompositions require many layers or prohibitively high frame rates, limiting their practicality using current LCD technology. However, with joint multilayer, multiframe decompositions, display designers can explore the interdependence of the number of layers, the number of frames, and the image brightness. In this section we demonstrate that Tensor Displays using relatively few layers and frames achieve higher-fidelity reconstructions than prior methods, in a manner supported by current LCD technology. We also show that Tensor Displays achieve wide fields of view, as required for multiviewer scenarios.

We employ PSNR to quantify the difference between reconstructed views and the target light field. We expect perceptual error metrics to better predict subjective assessments; unfortunately, multiview perceptual metrics remain an open research topic. We consider a fixed set of uniformly-spaced viewpoints during optimization. As shown in the supplementary video, providing closely-spaced target views sufficiently constrains the decompositions so minimal artifacts are perceived at intermediate viewpoints.

Interdependence of Layers, Frames, and Brightness Display designers seek to maximize image fidelity (e.g., PSNR) as a function of device complexity (i.e., the number of layers and frames). Consider optimizing multilayer designs with uniform backlighting. The design trade space is shown in Figure 3-37. The teapot scene is decomposed for a field of view $\alpha = 20^\circ \times 20^\circ$, spatial resolution of 160×100 pixels, 3×3 views, and layer separation $\Delta d = 4.0$ cm. Note that these display parameters differ from those for Figure 3-34, where

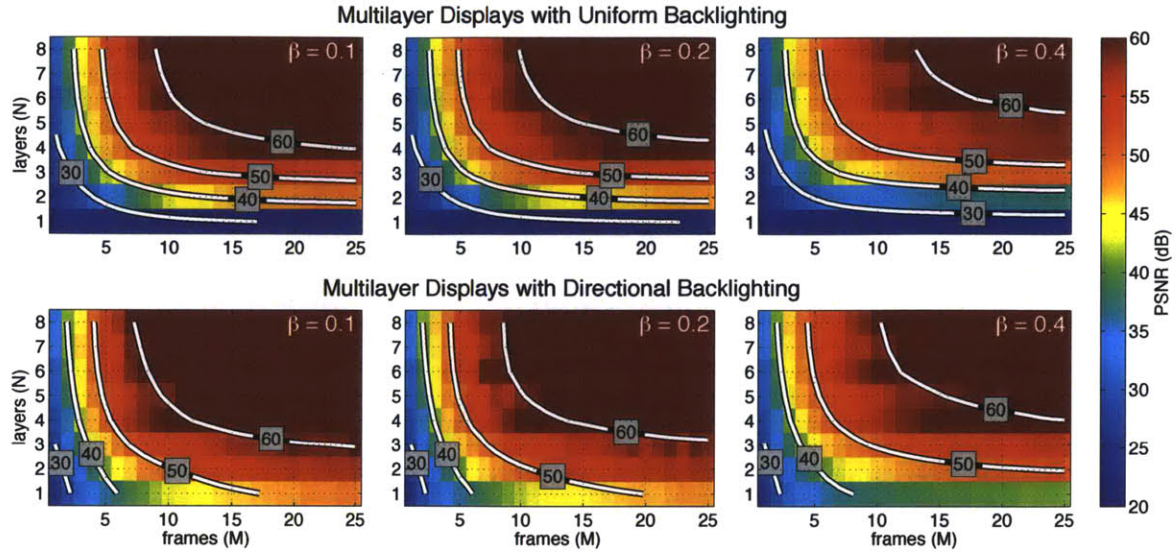


Figure 3-37: Design trade-offs for Tensor Displays. Peak signal-to-noise ratio (PSNR), as a function of the number of frames M , number of layers N , and brightness β , evaluated for the teapot scene and the display parameters in Section 3.4.2. (Top) Results with uniform backlighting. (Bottom) Results with directional backlighting.

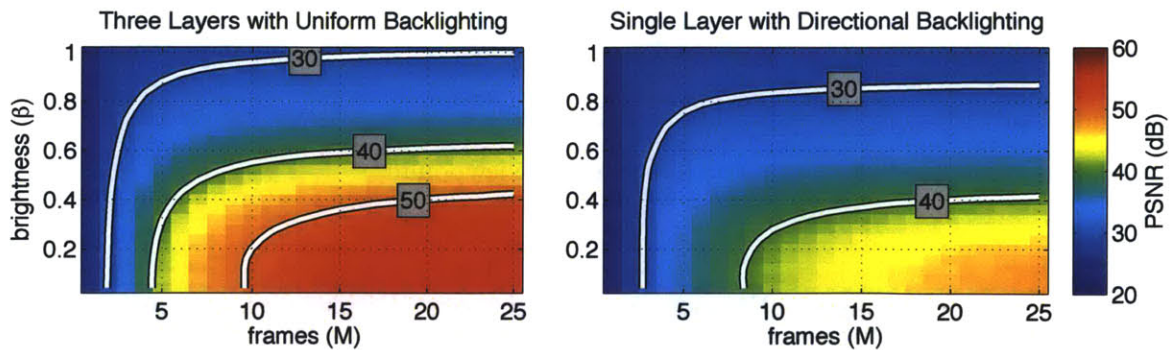


Figure 3-38: Optimizing the Tensor Display prototypes. PSNR is evaluated, as a function of the number of frames M and brightness β , for the teapot scene and the display parameters in Section 3.4.2.

the layers are separated by only 8 mm. These simulations verify a key benefit of Tensor Displays: increasing the number of frames allows the number of layers to be decreased (for a given PSNR). These simulations also reveal the dependence on the brightness scale $\beta \in [0, 1]$ applied to the target light field; specifically, we modify Equation 3.49 to yield the following objective function supporting a trade-off between image brightness and fidelity.

$$\arg \min_{\mathbf{F}^{(n)}} \left\| \beta \mathcal{L} - \mathcal{W} \oplus \tilde{\mathcal{J}} \right\|^2, \text{ for } 0 \leq \mathbf{F}^{(n)} \leq 1 \quad (3.61)$$

We observe that decreasing brightness generally yields higher-fidelity reconstructions for the same number of layers and frames.

The trade space for multilayer displays with brightness $\beta = 0.2$ is shown in the center of the top row of Figure 3-37. We observe that static decompositions (i.e., $M = 1$) cannot exceed 30 dB, even with as many as eight layers. To achieve 40 dB with eight layers, two frames are required. However, note the trade-off between layer complexity and refresh rate along the 40 dB curve. Using six frames, only three layers are required, with more frames providing marginal benefits. Thus, with Tensor Displays, designers can exploit high-speed displays to reduce device complexity, minimizing the number of layers to achieve a certain image fidelity.

Adding a directional backlight alters the design trade space, as shown at the bottom of Figure 3-37 for a directional backlight with 47×29 lenslets. We observe that two frames are still required to reach 40 dB using eight layers. However, only a single layer is now required using eight frames. For this example, the directional backlight effectively reduces the number of required layers by one. This underscores the practical benefits of the tensor display framework, which is the first to combine the benefits of multilayer decompositions, time-multiplexing, and directional backlighting.

Tensor Displays encompass a broad set of architectures. In Section 3.4.4, we configure the prototype to demonstrate two designs: three layers with uniform backlighting and a single layer with directional backlighting. The design trade spaces are shown in Figure 3-38. For

three layers, four frames are required to achieve 40 dB. With additional frames, brightness can be significantly increased (up to $\beta \approx 0.6$). To our knowledge, this is the first automultiscopic display demonstrating such trade-offs between display refresh rate and brightness, providing additional motivation for developing high-speed spatial light modulators. Similarly, with directional backlighting, a minimum of eight frames are required to achieve 40 dB. We confirm predicted PSNR trends in Section 3.4.4.

Increasing Field of View Conventional automultiscopic displays, including parallax barriers and integral imaging, exhibit a set of periodically-repeating viewing zones. In contrast, recent computationally-optimized multilayer and multiframe displays generally exhibit a set of non-repeating viewing zones; while yielding extended depths of field, greater resolution, and increased brightness, viewers are typically limited to a field of view of $\alpha \lesssim 20^\circ$. As shown in Figure 3-39, Tensor Displays support wider fields of view, while retaining the benefits of computational optimization. A field of view of $\alpha = 50^\circ \times 20^\circ$ is achieved, for a light field with 9×3 views, using either five layers and uniform backlighting or a single layer and directional backlighting. We observe that prior multilayer-only and multiframe-only decompositions lack sufficient degrees of freedom to achieve high-PSNR reconstructions for this scenario. Differences between predicted and observed depths of field and PSNR, as shown in Figures 3-36–3-38 and Figure 3-39, are due to differing fields of view in these experiments.

3.4.3 Implementation

This section describes a Tensor Display prototype reference design, and assesses its performance. We first review the prototype hardware and software implementation. Afterwards, we evaluate the performance for two prototype configurations: a three-layer LCD with uniform backlighting and a single LCD with directional backlighting.

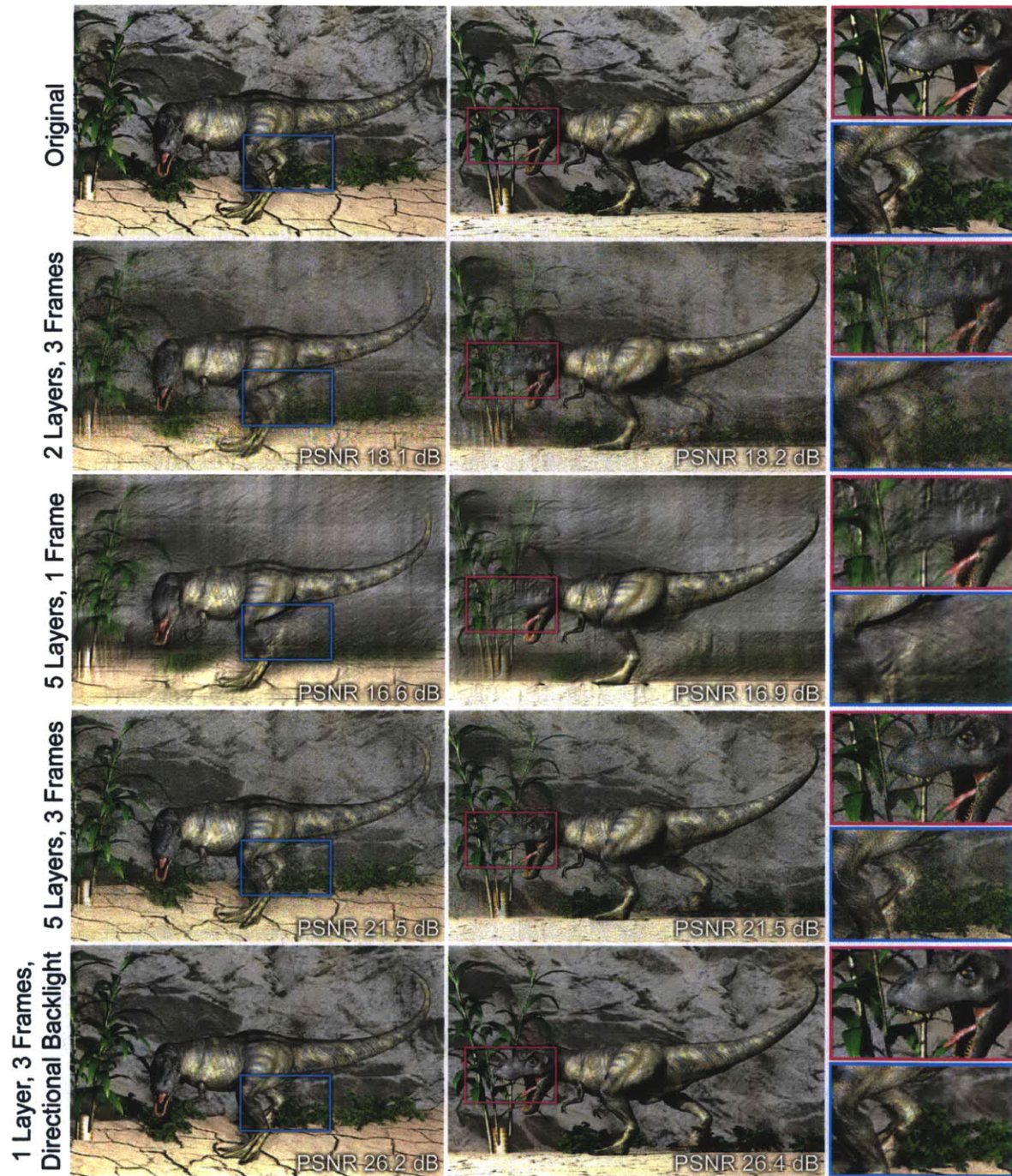


Figure 3-39: Tensor Displays achieve wider fields of view than prior multilayer displays. This example assumes a field of view of $\alpha = 50^\circ \times 20^\circ$ and three frames. We observe that Tensor Displays, using five layers with uniform backlighting (fourth row) or a single layer and directional backlighting (fifth row), minimize artifacts compared to multiframe-only (second row) and multilayer-only (third row) decompositions.

Software

Target light fields are rendered using POV-Ray or, for interactive applications, using OpenGL. Rendered light fields have a spatial resolution of 840×525 pixels (i.e., half the resolution of LCDs used in the prototype) and an angular resolution of 5×5 views.

We implemented nonnegative tensor factorization (NTF) using the multiplicative update rules from Section 3.4.1. An offline, Matlab-based solver is used for simulations. Decomposing a target light field into a six-frame sequence for three layers takes approximately 30 minutes using 50 updates. Color channels are processed independently. An online, GPU-accelerated solver is implemented in OpenGL and Cg. Our update rules can be cast as additive combinations of the logarithms of the layer transmittances. Using this representation, the update rules are mapped to standard operations of the graphics pipeline, including projective texture mapping, accumulation buffers, floating point framebuffers, and perspective rendering. These operations are not only computationally efficient, but also memory-efficient, as only the non-zero tensor elements need to be stored and processed. For interactive applications we exploit temporal coherence between decompositions, seeding each frame with the prior result, as shown in the supplementary video.

Separate threads are used to decouple the decomposition from the display routines. Decompositions are evaluated in an asynchronous thread, updating layer patterns as they become available. This ensures that all display layers can be continuously refreshed at 120 Hz, without waiting for updated decompositions. Using the prototype hardware, we achieve up to 10 multiplicative updates per second for as many as 12 frames. Light fields with reduced spatial or angular resolution can be decomposed and displayed at interactive refresh rates, as shown in the supplementary video. All experiments using the prototype display employ the GPU-accelerated solver.

Hardware

We built a reconfigurable Tensor Display prototype capable of implementing two-layer and three-layer architectures with uniform or directional backlighting (see Figure 3-32). The

layers are constructed using three modified Viewsonic VX2268wm 120 Hz LCD panels. The front and rear polarizing films are removed from the front two LCDs, and the stack is interleaved with alternating crossed linear polarizers. Aluminum brackets added to the rear panel allow lenslet arrays to be affixed for operation as a directional backlight. A rectangular lenslet array is approximated using two crossed lenticular sheets, purchased from Micro Lens Technology, Inc. The corrugated surfaces of the sheets are held in direct contact, minimizing astigmatic aberrations [11]. The directional backlight supports varying spatio-angular resolution trade-offs using 10, 15, and 20 lenses per inch (LPI) lenticular sheets. We observe that the sheets are birefringent due to stresses introduced during manufacturing. In directional backlighting modes, an additional polarizing film is placed after the lenslet arrays, restoring the linear polarization state before rays impinge on the next LCD in the stack.

We implemented offline and online solvers based on Equation 3.50. Computation is divided between CPUs, for the offline solver, and GPUs for the online solver. The offline solver is run on an Intel Core i5 workstation with 10 GB of RAM. The online solver is run on an Intel Core i7 workstation with 6 GB of RAM and an external Nvidia QuadroPlex 7000 graphics unit containing two Quadro GPUs and a G-Sync card. This provides four frame-synchronous DVI outputs capable of driving the LCDs at 120 Hz.

Display Calibration Strategies

Moiré fringes are observed when two patterns of different spatial frequency are multiplied. The effect is often observed in digital photography and display when patterns in a scene approach the spatial frequency of the underlying pixel grid of the image capture or display device. From a signal processing perspective, Moiré can be understood as a beat frequency between signals of similar spatial frequencies, or equivalently, spatial frequency aliasing. In this section, we demonstrate how to use moiré fringes to accurately align layered lens array and LCD systems.

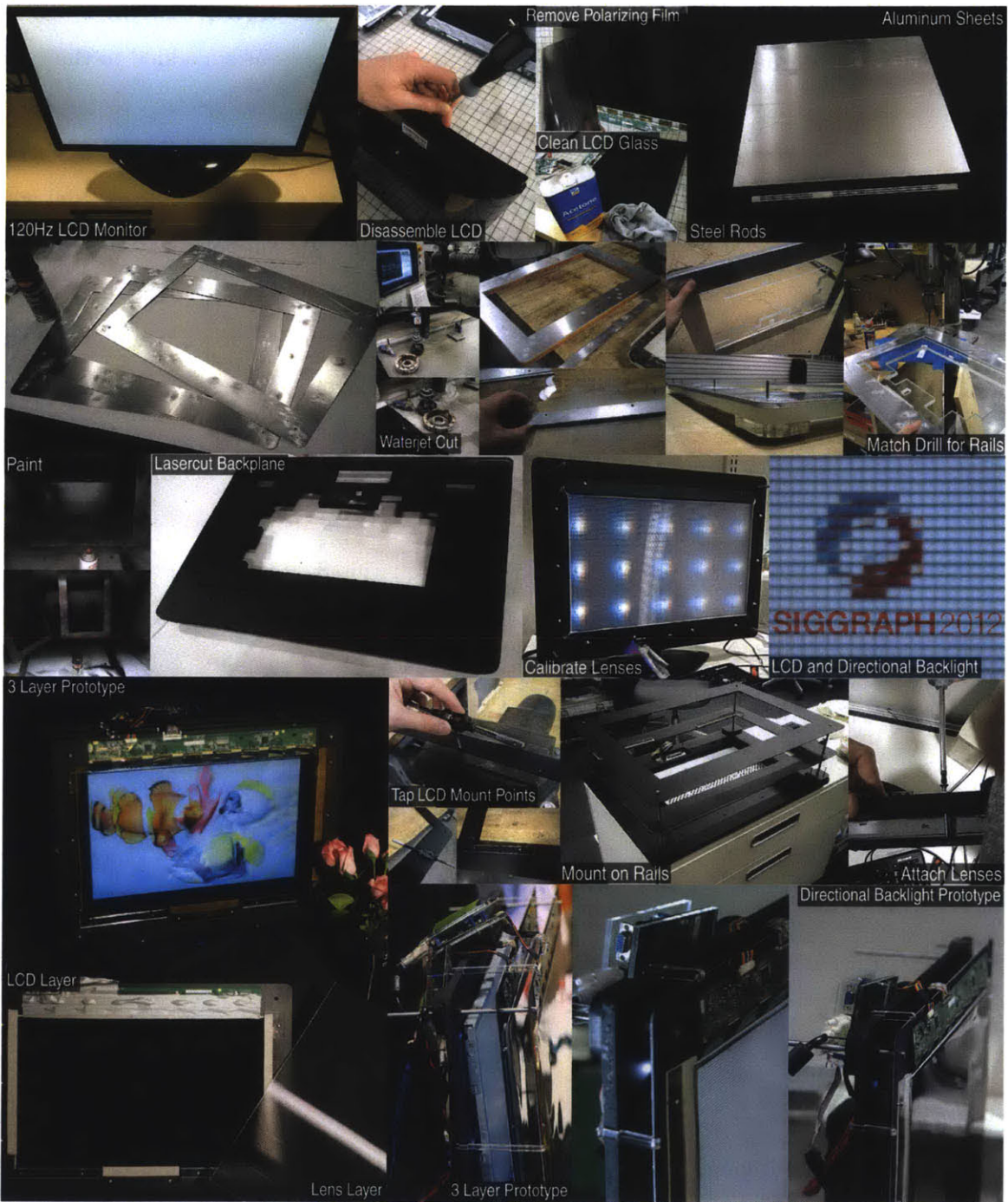


Figure 3-40: Prototype construction. Three LCD panels were modified to implement two-layer and three-layer Tensor Displays. Custom waterjet-cut and laser-cut parts ensured accurate alignment of the display components.

Lenticular Alignment In many light field displays, including our directional-backlight Tensor Display (Section 3.3.2), it is necessary to align a lenticular sheet or lens array to an underlying pixel grid. Here we described a simple technique to perform rotational alignment using the moiré effect. In the case of lens sheets, this effect has been succinctly described as the *moiré magnifier* [92]. Following Hutley et. al., an expression can be obtained for the relative rotation and magnification of the image of the pixel grid of the LCD panel as viewed through the lens array. For convenience, we reproduce Hutley et. al.’s magnification, m and rotation, ϕ , here, with one minor modification: we simplify the denominator using the Pythagorean trigonometric identity.

$$m = \frac{a}{\sqrt{a^2 + b^2 - 2ab \cos(\theta)}} \quad (3.62)$$

$$\sin(\phi) = \frac{-b \sin(\theta)}{\sqrt{a^2 + b^2 - 2ab \cos(\theta)}} \quad (3.63)$$

As we show numerically below, for practical values of LCD and lens pitch, small rotations of the lens array will be magnified in the moiré pattern. The calibration task reduces to leveling the perceived moiré fringes by eye. If the lens pitch is nearly an integer multiple of the LCD pitch, as is the desired case, then m will be nearly infinite when $\theta = 0$. Moiré bands will not be visible under these conditions. However this calibration technique applies equally to patterns displayed on the LCD as the pixel structure of the screen itself. It is often desirable to display a pattern on the LCD to improve the contrast of the observed moiré pattern. Once the pattern has been leveled by rotating the lens sheet, the pitch of the lens sheet can be calibrated by adjusting the pitch of repeating pattern displayed on the LCD until m is infinite, or equivalently, no fringe patterns are visible. In the case of a research prototype using an imperfectly matched lens array and LCD panel, the displayed pattern may be interpolated to achieve sub-pixel alignment, with a small angular cross-talk penalty in the resulting light field display.

To determine the expected accuracy of the above method, we consider the physical values from our Tensor Display prototype. The lens pitch is $a = 2.54mm$, and LCD pixel pitch

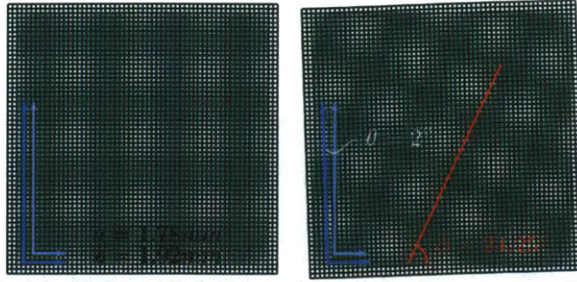


Figure 3-41: Moiré interference patterns caused by scaled and rotated grids. Best viewed at full resolution. The grids differ in pitch by 7%. On the right, the larger grid is rotated by $\theta = 2^\circ$, causing an apparent rotation of the moiré fringes by $\phi = 24.25^\circ$.

is $b = 282\mu m$. We found that the rotation of the moiré fringes could be aligned to within $\phi = 0.5^\circ$. Substituting into Equations 3.63 and 3.62, and solving for θ and m , respectively, we get $\theta = 4^\circ$ and $m = \text{undefined}$. This is a result of choosing a and b as nearly integer multiples ($a/b = 9.007$), and indicates that alignment by eye will not be very accurate. To improve accuracy, we can display a linearly interpolated pattern on the LCD with a pitch of $b = 2.1mm$. Now, for $\phi = 0.5^\circ$, $m = 5.77$ and $\theta = 0.105^\circ$, allowing nearly $5\times$ improvement in accuracy over alignment by eye.

LCD Alignment Though it is possible to use the scale of moiré fringes to perform alignment in depth, we find it is much simpler in practice to use CNC machines to cut spacer clips, which can fasten to multiple layers of optical elements and space them accurately to the tolerance of the CNC machine – $0.25mm$ or less. In this section we concentrate primarily on rotational alignment of LCD panels, to which moiré fringes are more sensitive.

Though the analysis for lenticular sheets was derived from the *moiré magnifier* effect of lens arrays, we observe that Equations 3.62 and 3.63 apply equally to lens arrays and grid patterns. Oster et. al. [151] use an analysis based on indicial representations of curves to derive Equations 6 and 7 in their paper for moiré fringe pitch and rotation, which match our Equations 3.62 and 3.63, save for a sign difference. We show in Figure 3-41 that the analysis holds for a printed grid pattern.

When aligning LCD screens spaced by a distance d_s , the difference in pixel size Δp , observed

by a viewer at distance d_o is due to perspective projection. By similar triangles,

$$\Delta p = \frac{pf}{d_o} - \frac{pf}{d_o + d_s}. \quad (3.64)$$

where f is the focal length of the human eye, accepted to be approximately 22mm . For the physical dimensions of our three-layer Tensor Display prototype (considering the front two layers) $p = 282\mu\text{m}$, $d_o = 1\text{m}$, $d_s = 4\text{cm}$, we calculate that $\Delta p = 0.241\mu\text{m}$. Substituting the two apparent LCD pitches into Equation 3.63, we find that a pattern rotation of $\phi = 0.5^\circ$ yields a screen rotation of just $\theta = 7.5 \times 10^{-6}$ degrees, indicating that aligning the LCD layers by straightening the visible moiré fringes will achieve very accurate alignment. It is useful to note that, with such a small difference in pitch, the magnification, m , will be large, making it more difficult to achieve accurate visual rotation alignment. Thus, $\phi = 0.5^\circ$ may be an overly generous estimate.

Moiré Mitigation While moiré is beneficial for accurate calibration, it is an unpleasant visual nuisance when observing a light field. In order to eliminate moiré, one need only prevent the multiplication of similar spatial frequency signals. We find that there are two approaches that can mitigate moiré:

- Achieve a small magnification factor, m , such that aliased copies of the signal are small relative to image features
- Implement a spatial low-pass or notch filter to remove the offending frequencies

In the case of LCD panels, the first of the above strategies implies separating the panels by a large distance. Larger separation increases Δp from Equation 3.64. However, a large separation distance is not always practical, and does not apply to lenticular sheets and lens arrays.

The second approach can be achieved in two-layer and lenticular devices by placing an appropriately chosen diffuser on the rear LCD layer. An appropriate diffuser choice will

impose a spatial frequency cut-off such that any moiré observed will have a small magnitude or small magnification. We find that a light weight diffuser such as Grafix Matte Acetate 0.005 works well for LCD panels with a pixel pitch in the $\frac{1}{4}mm$ range.

3.4.4 Assessment

Three-Layer LCD with Uniform Backlighting

As shown in Figure 3-32, the prototype was configured as a three-layer LCD with uniform backlighting. Acrylic spacers separated each panel by $\Delta d = 4.0$ cm. The target light field was rendered with a field of view of $\alpha = 20^\circ \times 20^\circ$ and brightness $\beta = 0.2$ (see Section 3.4.2). Photographs of the central view, seen directly in front of the prototype, are shown along the center column of Figure 3-42. Each light field was decomposed using twelve frames. The camera exposure was set to 100 ms, simulating a 720 Hz display for a human observer (i.e., for a 60 Hz flicker fusion threshold). We observe that fine details are preserved (e.g., the fish scales and specular highlights on the teapot) and occlusion cues are correctly rendered (e.g., between the bunnies). See the supplementary video for demonstrations of smooth horizontal and vertical motion parallax.

Experiments with the prototype provide insights into practical engineering issues. Foremost, we found that accurate mechanical alignment is crucial. As shown in Figure 3-35, decomposed layers exhibit high-frequency patterns that must be properly aligned. Accurate alignment was ensured by displaying perspective images of a crosshair array on each layer. A camera was placed at the desired viewer position (e.g., directly in front of the display at a distance of 2 m) and the patterns were shifted until alignment was obtained. We also found that radiometric calibration is necessary, including measuring the black levels and gamma values. The former are incorporated as constraints in the update rules, while the latter are addressed by applying gamma correction at runtime. We attribute remaining variations in color and intensity to differences in the LCD color gamut, color filter cross-talk, moiré due to stacking multiple layers, and angular color variation common to high-speed LCDs.

Single LCD with Directional Backlighting

As shown in Figure 3-32, the prototype was also configured as a single LCD with a directional backlight. The backlight was fashioned using crossed 10 LPI lenticular sheets, yielding a field of view of $\alpha = 48^\circ \times 48^\circ$ and backlight resolution of 187×117 lenslets. The front LCD was separated by $\Delta d = 8.5$ mm from the middle of the lenticular sheets. Remaining system parameters were identical to the three-layer prototype. Photographs of the central view are shown along the right column of Figure 3-42. We observe the crossed lenticular sheets produce strong absorption along lens boundaries. In a commercial implementation, lenslet arrays could be manufactured with minimal absorption. Alternatively, edge-lit directional backlighting could eliminate this artifact. As demonstrated in Figure 3-43, adding an LCD in front of a low-resolution directional backlight increases the spatial resolution for virtual objects appearing on the display surface (e.g., the logo) and for objects extending in depth (e.g., the fish tail on the right). While resolution can be enhanced at the display surface without time multiplexing, enhancement for extended scenes can only be achieved with time multiplexing, as facilitated by our tensor framework.

3.4.5 Understanding Tensor Displays

Tensor Displays comprise a family of display architectures that includes many possible implementations. The characteristic feature shared by all Tensor Display incarnations is that multiple light-attenuating optical elements are combined in a way such that each ray in a target light field intersects each optical element at most once. Light-attenuating elements are usually arranged in layers which can be composed of any of the following: angularly-invariant spatial light modulators, purely directional modulators, and spatio-angular modulators. A low-resolution light field backlight, for instance, implemented by a lenslet array on top of an LCD, is one type of spatio-angular modulator. In this section we use a series of examples to provide an intuition for how nonnegative matrix factorization and nonnegative tensor factorization decompose a given light field for a specific Tensor Display implementation. In Section 3.4.6, we illustrate the tensor space spanned by different display types

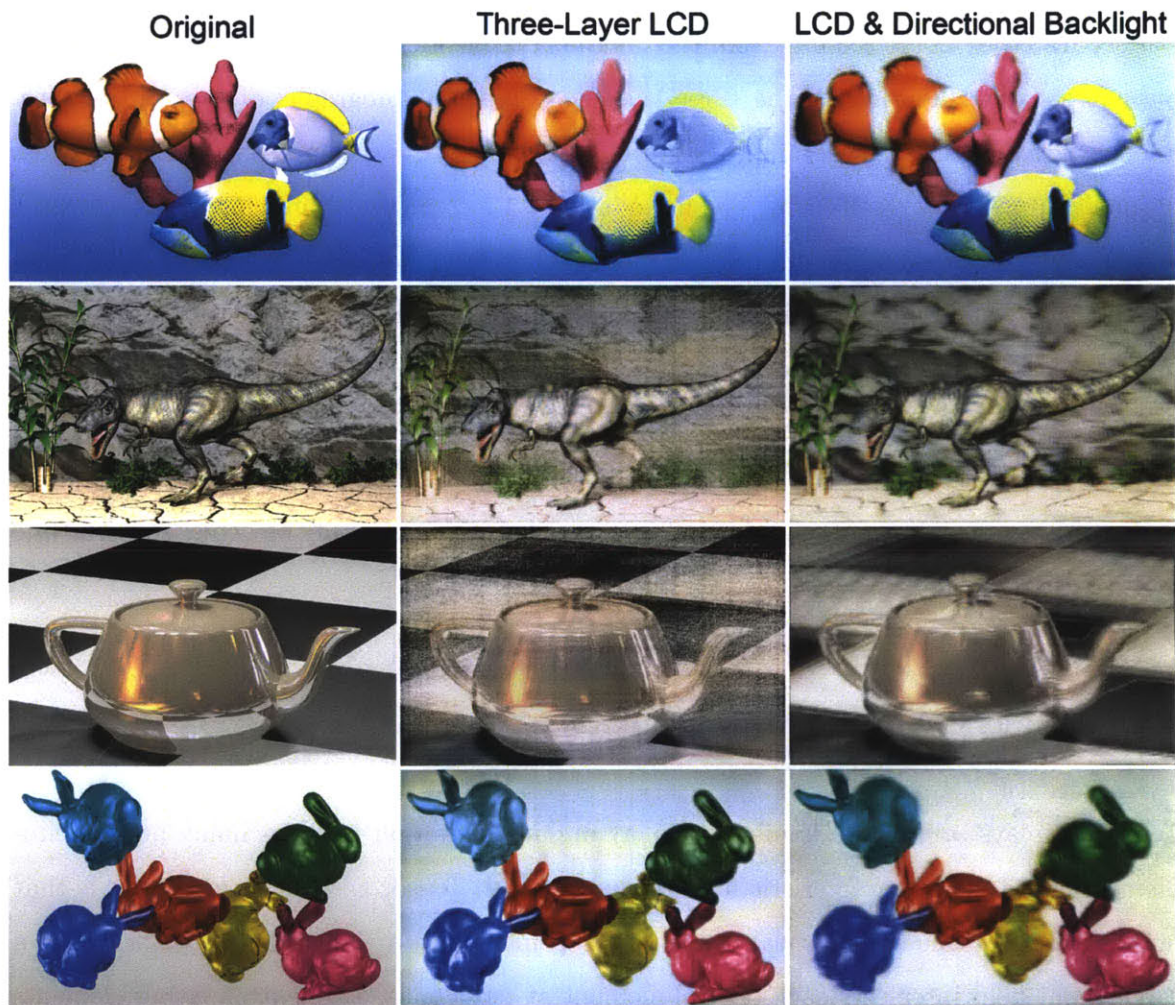


Figure 3-42: Experimental results using the Tensor Display prototype. Central views of four scenes are shown for the input light fields (left column), photographs of the three-layer LCD (center column), and the single LCD with directional backlighting (right column).

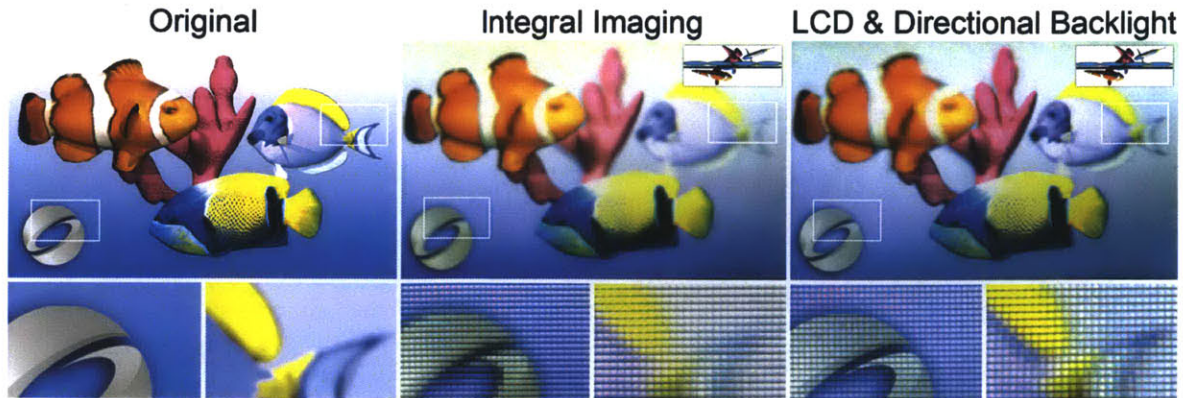


Figure 3-43: Enhancing integral imaging with Tensor Displays. While integral imaging, here implemented with a lenslet array affixed to an LCD, achieves a convincing 3D effect, spatial resolution is significantly reduced (center). Adding an LCD in front of the low-resolution backlight and exploiting temporal multiplexing using our tensor framework increases the spatial resolution, not only on the physical layers, but also outside the hardware enclosure (right).

in detail, whereas Section 3.4.7 demonstrates NTF decompositions for a variety of display implementations and compares them to decompositions computed with alternative methods proposed in the literature.

3.4.6 Light Field Tensors

The tensor space spanned by a Tensor Display with N optical elements, such as layers, is of dimension N . As observed in Figure 3-44, the light field only occupies a low-dimensional manifold within the tensor space. The shape of the manifold depends on a particular tensor display configuration and is shown for a three-layer display as well as for a dual-layer configuration with an additional directional backlight. A weighted nonnegative tensor decomposition has non-zero values only on the low-dimensional manifold created by the light field in tensor space. These visualizations illustrate the tensor space for different displays in an intuitive manner.

3.4.7 Light Field Tensor Factorization

The following subsections show nonnegative tensor factorizations for a variety of Tensor Display implementations.

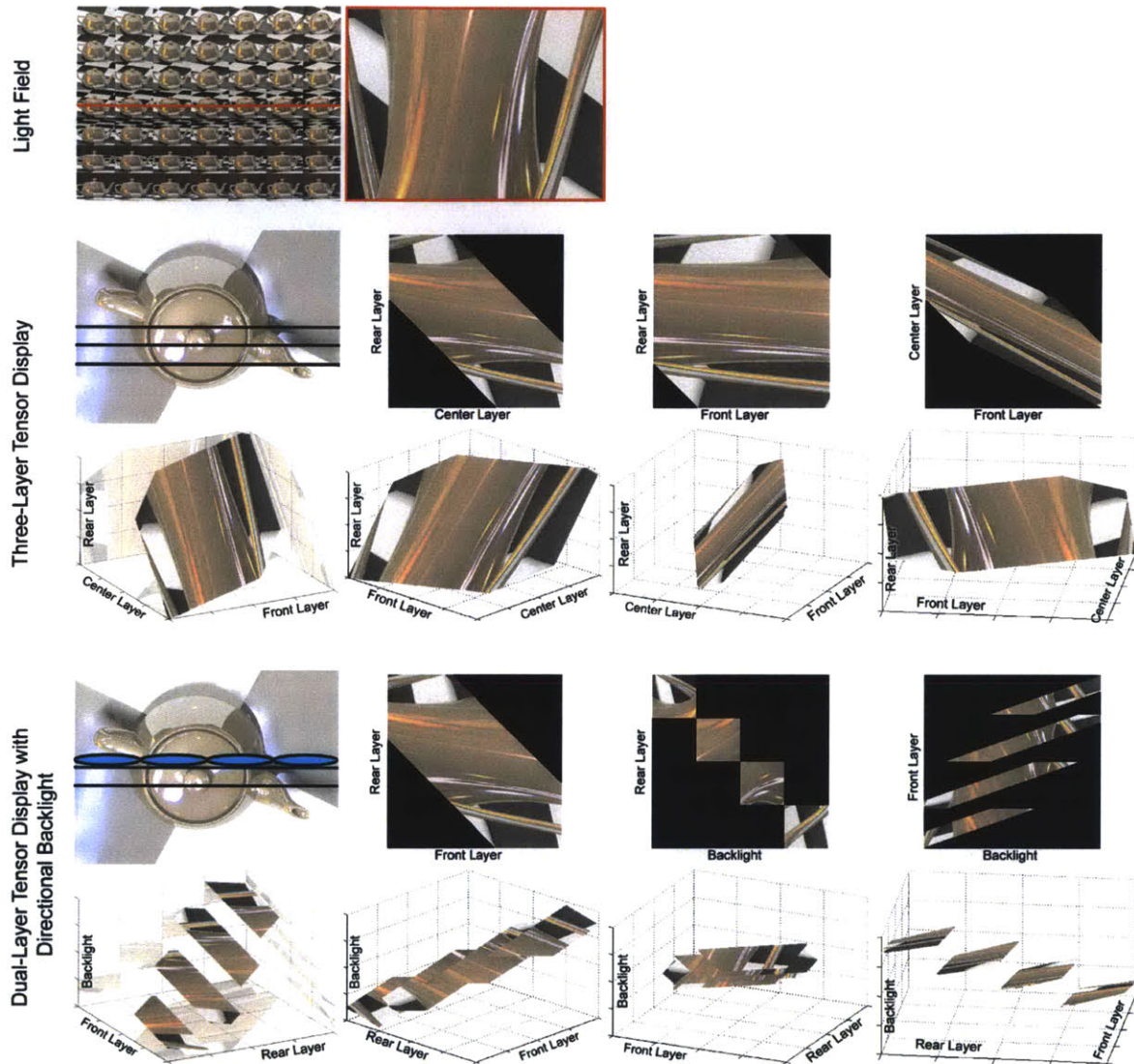


Figure 3-44: Tensor Display visualization for a three-layer implementation and a dual-layer display with a directional backlight. For illustrative purposes, the light field is a 1D slice (upper right) of a full 4D light field (upper left). While the pairwise layer parameterizations in the three-layer case span individual matrices (row 2), the tensor itself spans a higher-dimensional space with the light field embedded in a two-dimensional manifold within that tensor space (row 3). A similar effect can be observed for the dual-layer and backlight configuration (row 5); the lower-dimensional manifold within the tensor space differs from the three-layer case. Elementwise parameterizations are shown in row 4. A layer held against the directional backlight creates blockwise-independent matrix components (row 4, center right), whereas a gap between the two optical elements creates a shear in the light field (row 4, right).

Single Layer and Purely Directional Backlight

Figure 3-45 illustrates an intuitive case: a single, high-resolution layer, for instance an LCD, is combined with a purely angular light source. This could be a large lens directly behind the layer. The angular resolution of the backlight is assumed to be the same as the target light field (Fig. 3-45, top row), in this case 3×3 . Given the target light field and the physical setup, a nonnegative tensor factorization can then be performed for any desired or feasible number of temporally-multiplexed frames. The naïve solution would be to use nine frames and illuminate a single backlight direction at a time, showing the corresponding light field view on the front layer. We demonstrate in Figure 3-45 (rows 6 and 7) that NTF converges to the naïve solution if nine time frames are available. Doing so would, however, require a synchronized LCD and backlight to run at a minimum of 540 Hz, assuming a flicker fusion rate of the human visual system of 60 Hz. A lower frame rate may be required by the available hardware, which does not have an obvious heuristic solution. Nonnegative tensor factorization handles these cases naturally and provides the optimal decompositions, in a least-squared error sense (Fig. 3-45, rows 2–5).

Single Layer and Low-Resolution Light Field Backlight

Figure 3-46 evaluates the performance of a single light-attenuating layer combined with a low-resolution backlight. The backlight is simulated with four spatial resolutions, all lower than the layer resolution. PSNRs of the reconstructions are given in the insets. As shown, a low-resolution directional backlight combined with a high-resolution layer, such as an LCD, can achieve high image quality by temporally multiplexing only a few frames.

Dual Layer Factorization

Dual-layer automultiscopic display architectures have been driven using nonnegative matrix factorization (NMF) in Section 3.3. Nonnegative tensor factorization (NTF) mathematically reduces to NMF for the special case of dual-layer displays, because the spanned tensor is just

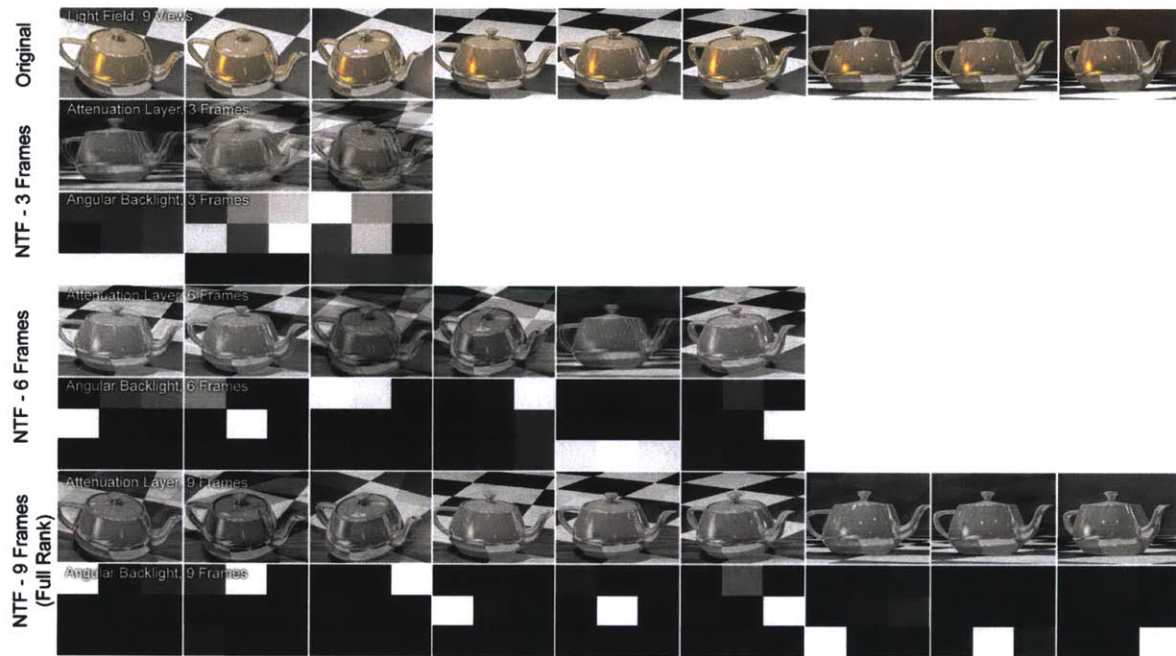


Figure 3-45: Original light field with 3×3 views (top row) and decompositions for a high-resolution layer directly on top of a purely directional backlight. This kind of backlight corresponds to a single large lens directly behind an LCD with another spatial light modulator (SLM) mounted at the focal length of the lens; the secondary SLM has a resolution of 3×3 , corresponding to the angular resolution of the light field. Decompositions for both high-resolution LCD and low resolution angular backlight are shown for three time-multiplexed frames (rows 2 and 3), six frames (rows 4 and 5), and nine frames (rows 6 and 7). The brightness for all decompositions is scaled by the inverse of the number of frames. As seen in the lower two rows, NTF converges toward the obvious solution: turning on each direction of the backlight sequentially over time with the LCD showing the corresponding view of the light field. NTF, however, generalizes the factorization problem to an arbitrary number of frames and different brightness tradeoffs. For the case of rank-deficient decompositions (rows 2–5), the views and corresponding backlight directions are automatically grouped into the set of structurally similar views that result in the optimal image quality.

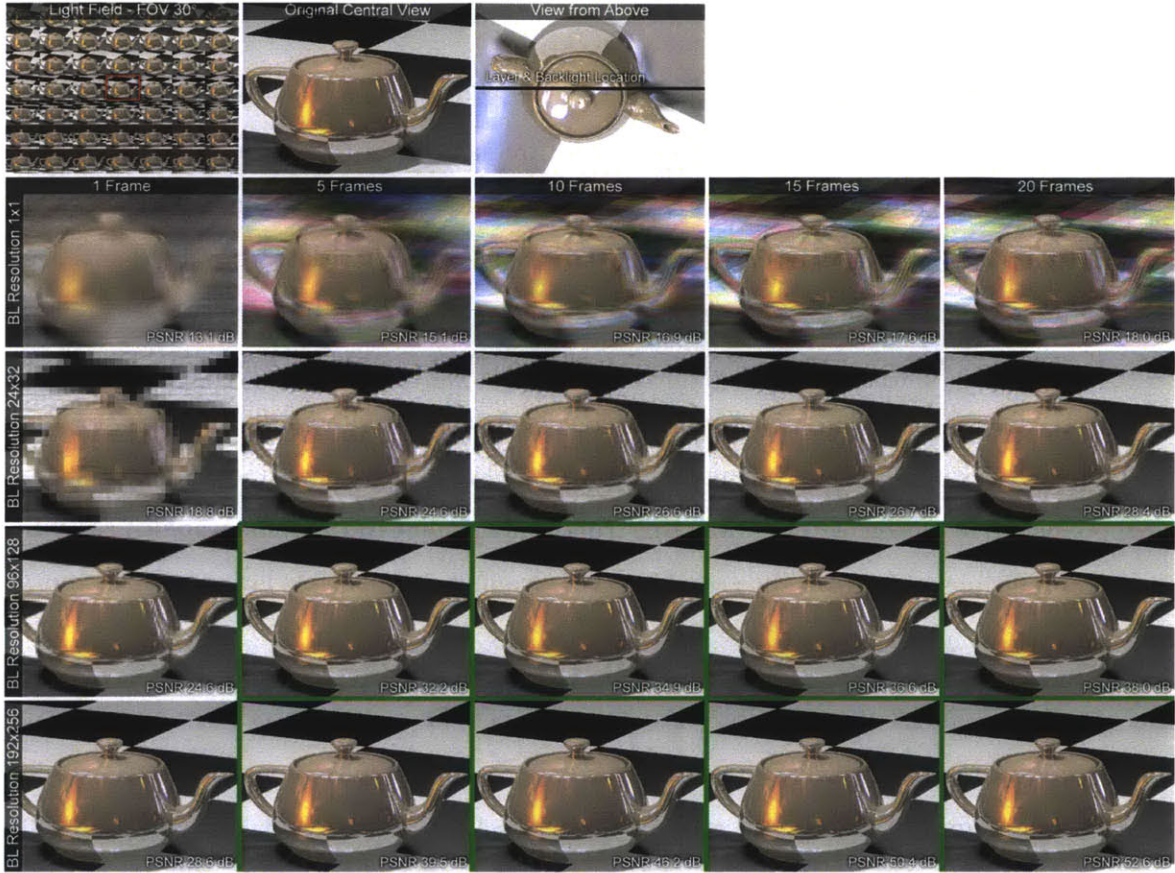


Figure 3-46: Simulated reconstructions of the central view for a light field covering a field of view of 30° with a varying number of frames and different spatial resolutions of the backlight. The spatial resolution of the layer is 512×384 ; the backlight is simulated to have (rows, from bottom) a spatial resolution of a factor of 2, 4, and 16 times lower than the layer resolution, as well as no spatial resolution at all (row two). As illustrated by the green boxes, a backlight with a spatial resolution of 1/4-1/8 of that of the layer can achieve high-quality reconstructions for only a few temporally-multiplexed frames.

a matrix. Therefore, NTF produces identical layer decompositions as NMF for this special display configuration. The NTF framework, however, generalizes to multilayer architectures as well as combined multilayer and directional backlight configurations.

Multilayer Tensor Factorization

With Figure 3-47, we want to build an intuition for NTF-based multilayer decompositions. As illustrated in these examples, the low spatial frequencies in the decomposed layers are comparable to the tomographic solution. This acts similarly to a 3D geometry slicing operator for Lambertian objects on the layers. Multiframe decompositions computed with our tensor framework additionally contain high-frequency variations in image regions exhibiting motion parallax. Although these high-frequencies could be perceived as noise, they actually contain the information that increases the 3D image quality for temporally-multiplexed Tensor Displays. With these experiments, we confirm that multilayer decompositions computed with nonnegative tensor factorization are structurally similar to the tomographic case if no temporal multiplexing is used, but combine the advantages of multiple layers with temporal multiplexing for all other cases.

In Figures 3-48 and 3-49 we analyze the behavior of NTF with respect to the number of update iterations and the rank; these results compare photographs of our three-layer prototype. A minimum of 50 iterations is generally necessary to ensure high image fidelity, but about 6–12 time-multiplexed frames achieve a high image quality even for the challenging teapot scene exhibiting a large depth of field. Figure 3-50 demonstrates how light fields with uncorrelated views, such as Arabic numerals, can be successfully synthesized using the proposed low-rank tensor factorization.

Multilayer and Purely Directional Backlight

In addition to the multilayer-only decompositions, we show decompositions for a dual-layer display with an additional, purely directional backlight in Figure 3-51. This setup resembles dual-layer configurations explored in Section 3.3, but generalizes to include an

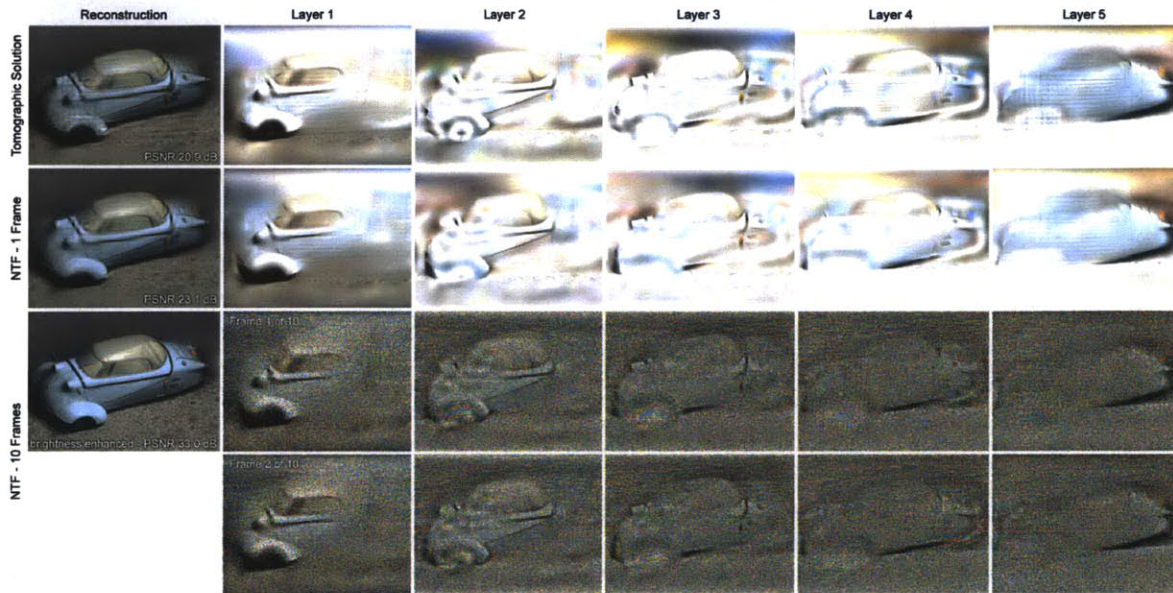


Figure 3-47: Multilayer decompositions. A tomographic five-layer decomposition (upper row) is intuitive, because it acts similar to a 3D geometry slicing operator for diffuse objects inside the physical display enclosure. For global illumination effects and objects outside the layers, however, the decompositions are more complicated. A nonnegative tensor factorization for the same optical configuration, without any time multiplexing, is shown in row two. The decompositions show a close similarity to the tomographic solution. The difference between the two is that the tomographic solution is computed in log-space, resulting in a linear problem which can be solved efficiently, but with biased errors. As seen in column one, specular highlights are slightly blurred and artifacts resulting from the parallax between different viewpoints are more pronounced. By adding temporal multiplexing, as shown in the lower two rows, the achieved quality can be significantly improved. The decompositions themselves still resemble a slicing operator in the lower frequencies, but what is perceived as temporally-varying high-frequency noise (lower two rows, columns two to six) actually contains the information necessary to improve the resulting 3D image quality. Note that any multi-frame decompositions computed with NTF represent a tradeoff between PSNR and brightness; the latter is enhanced for the simulated reconstruction in row three.

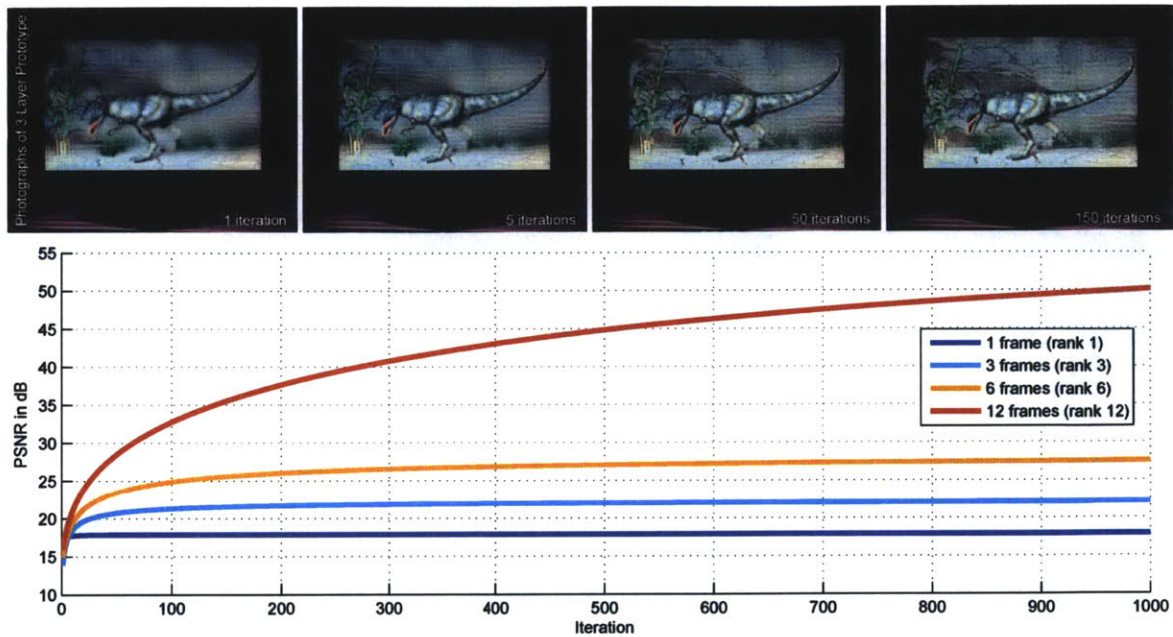


Figure 3-48: Convergence rate of multiplicative update rules. Top: four photographs of our three layer prototype showing an increasing number of NTF iterations for a rank 6 light field. At least 50 iterations (center right) are necessary to produce high-quality 3D images. Using only a few iterations (top left two photographs) result in blurred reconstructions, whereas a larger number of iterations (top right) do not significantly improve quality and represent an increased overhead in processing times. Bottom: plots showing convergence rates of the multiplicative update rules simulating the above experiments using 1, 3, 6, and 12 frames, respectively. With an increasing number of unknowns (i.e., larger numbers of frames), more iterations are required to converge.

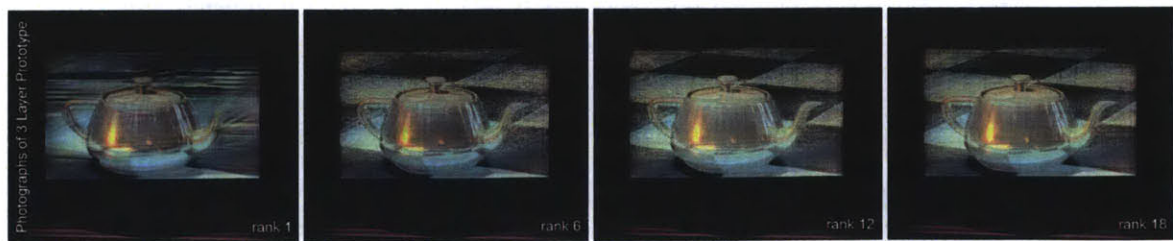


Figure 3-49: Rank analysis. Four photographs of our three layer prototype showing an increasing rank of the light field tensor. Without any time multiplexing (left photograph), low image quality is achieved for this scene due to the large depth of field. Low-rank approximations using 6 (center left) and 12 (center right) time-multiplexed frames create a visually appealing approximation of the light field. Higher-rank factorizations (right photograph) do not improve image quality significantly, demonstrating that light field tensors are inherently of low rank.



Figure 3-50: Synthesis of uncorrelated views. Five photographs of the three-layer prototype showing the performance of a rank 12 factorization for a light field comprising uncorrelated views (i.e., Arabic numerals in this example).



Figure 3-51: Decomposition for dual-layer display containing a purely angular backlight behind the rear layer. The original light field has 4×4 views within a field of view of 30° . Two of the original views are shown on the upper left with corresponding reconstructions next to them. This data set represents a rank-16 light field, which is decomposed using 10 frames. The two layers are separated by a distance that corresponds to the separation distance for an equivalent parallax barrier display. Each frame is shown for the front layer (row 3), for the rear layer (row 4), and for the angular backlight (bottom row). The layer decompositions resemble what NMF produces for dual-layer setups (see Section 3.3), but adds an angular backlight (see Section 3.4.7) for improved depth of field and field of view.

additional directional backlight. The layer decompositions exhibit high spatial frequencies, as analyzed in Section 3.3.1, whereas the directional backlight adds more degrees of freedom that increase the field-of-view and depth-of-field of the Tensor Display as compared to a dual-layer configuration (see Section 3.4.2).

3.4.8 Limitations

All stacked LCDs contend with a similar set of challenges, including moiré and color-channel crosstalk. The proposed method also confronts the further challenges of display flicker and the current limitations of NMF algorithms. We discuss solutions for each of these issues in

the remainder of this section.

Moiré: Viewing one LCD through another causes visible fringes (moiré) to appear. Commercial dual-stacked LCDs have eliminated moiré by increasing the blur introduced by the front diffuser on the rear LCD [16]. Ideally, the diffuser should only blur neighboring color subpixels—minimizing moiré while preserving spatial resolution. We use this solution in our prototype; a thin paper vellum sheet is placed against the rear LCD. Experimentally, the diffuser eliminates moiré, however the image resolution is reduced from 1680×1050 to 840×525 ; a custom diffuser, with a properly-selected point spread function, would prevent this reduction.

Color-channel Crosstalk: Each LCD color filter transmits a range of wavelengths. The relative transmission, as a function of wavelength, is known as the color filter transmission spectrum. The transmission spectra exhibit some overlap in commercial panels. Since the panels we use are not optimized for dual-stacked configurations, the overlapping transmission spectra cause visible color-channel crosstalk (see the supplementary material for experimental measurements). This crosstalk is ignored in our optimization; while allowing independent decompositions for each channel, this simplification results in visual artifacts. In a commercial implementation, the transmission spectra could be designed with minimal overlap. However, to minimize crosstalk for grayscale regions, we initialize each color channel with the same random set of values. As shown in Figure 3-16, the deterministic optimization algorithm leads to grayscale masks that minimize crosstalk in these regions.

Flicker: Humans perceive an intermittent light source as steady when it varies between 16–60 Hz, depending on illumination conditions. For dim stimuli in darkened rooms, 16 Hz is a commonly-accepted lower bound. Our prototype can multiplex up to eight mask pairs at 15 Hz. Multiplexing five mask pairs achieves a 24 Hz refresh, equivalent to cinematic projection. However, 240 Hz LCDs are commercially available and allow doubling the decomposition rank without altering the refresh rate. Thus, our method will benefit from the trend of LCDs with increased refresh rates. However, as observed by Woods and Sehic [210], such high-speed panels may require further optimization for autostereoscopic applications, rather than their current focus on reducing motion blur.

Non-negative Matrix Factorization: As described by Lee and Seung [122], multiplicative update rules (e.g., Equation 3.33) are easy to code, but are not as efficient as other algorithms [37]. Our algorithm scales linearly with the number of light field elements and the decomposition rank. The prototype requires at least 50 iterations to converge, resulting in an average run-time of around eight minutes per frame, preventing interactive content. However, as shown in the supplementary video, masks can be precomputed to allow dynamic content. Figure 3-16, 3-21, and 3-22 show our optimization produces high-frequency patterns, even in uniform regions. Regularized NMF algorithms [216] remain a promising direction of future work; preliminary results with smoothed masks are included in the supplementary material and Figure 3-30.

Chapter 4

Compressive Methods for Visual Capture

This chapter addresses the challenge of recording the spatial and angular variation of a 4D light field. The problem of compressive capture faces challenges analogous to that of compressive display—electro-optical technology available for sampling irradiance distributions is grossly under-sampled in a Dirac sampling sense. A summary of existing techniques to solve the light field capture problem is presented in Section 2.4. With the exception of recent examples that have begun to consider compressive, model-driven, frameworks [135], existing methods suffer from severe limitations in spatioangular resolution, capture speed, and optical efficiency.

We are not the first to consider compressive light field capture. The contribution of this chapter is an analysis of a compressive capture framework that, coupled with optical systems that carefully satisfy the requirements of Section 4.1 with complementary measurements, allow the application of multiple complementary reconstruction algorithms. For example, in Section 4.4 we detail a prototype *switchable* light field camera system that can take a single measurement amenable to reconstruction with fast linear methods and slower, but higher-quality non-linear methods.

4.1 Requirements for a Compressive Capture System

The framework developed over the course of this chapter represents a flexible and general method for driving advanced camera systems capable of capturing light intensity over a region of space and angle. However, in order to apply the framework to a given scene captured by a given camera system it is necessary to satisfy the following basic requirements.

It is not a coincidence that the requirements in this section are similar to those listed in Section 3.1, as light transport is reciprocal and the insights driving compressive capture and compressive display are related. However, differences arise between the two due the order of operations relative to general purpose computation, light transport, and measurement or emission.

Structured Data The light field imagery to be captured must be compressible—for the purposes of the compressive light field camera framework it must be sparse in some basis.

Suitable Measurement Basis The measurement basis of the optical system implementing the light field camera must be incoherent with the basis in which the light field is sparse.

4.2 Optically Efficient Methods

4.2.1 Angle Sensitive Pixels

Whereas light field cameras typically rely on modern algorithms applied to data captured with off-the-shelf opto-electronic systems, recent advances in complementary metal-oxide-semiconductor (CMOS) processes have created opportunities for more specialized sensors. In particular, angle sensitive pixels (ASPs) have recently been proposed to capture spatio-angular image information [197]. These pixel architectures use a pair of near-wavelength gratings in each pixel to tune the angular response of each sensor element using the Talbot



Figure 4-1: Prototype angle sensitive pixel camera (left). The data recorded by the camera prototype can be processed to recover a high-resolution 4D light field (center). As seen in the close-ups on the right, parallax is recovered from a single camera image.

effect. Creating a sensor of tiled ASPs with pre-selected responses enables range imaging, focal stacks [198], and lensless imaging [61]. Optically optimized devices, created with phase gratings and multiple interdigitated diodes can achieve quantum efficiency comparable to standard CMOS imagers [172].

ASPs represent a promising sensor topology, as they are capable of reconstructing both sensor-resolution conventional 2D images and space/angle information from a single shot (see Section 4.4.1). However, general light field reconstruction techniques have not previously been described with this hardware. We analyze ASPs in the context of high-resolution, compressive light field reconstruction and explore flexible image modalities for an emerging class of cameras based on ASP sensors.

4.3 Sparse Reconstruction

In this thesis we are not the first to propose the application of sparsity constrained reconstruction techniques to the problem of capturing light fields. A brief overview of relevant literature is given in Section 2.6. Here, we propose the use of emerging pixel structures, described in Section 4.2.1, which leads to a camera architecture that is well-suited for compressive reconstructions — for instance with dictionaries of light field atoms [135]. In addition, our flexible approach allows for high-quality 2D image and lower-resolution light

field reconstruction from the *same measured data* without numerical optimization.

This section summarizes the techniques of dictionary learning and patch-by-patch sparsity constrained reconstruction employed in the light field camera described in Section 4.4. These techniques were adapted in large part from the work of Marwah et al. [135].

4.3.1 Dictionary Learning

Following Candès et al. [31] and Marwah et al. [135] it is possible to learn the fundamental building blocks of natural light fields—light field atoms—in overcomplete dictionaries.

In this section we closely follow the formulation of Marwah et al. [135]. They consider 4D spatio-angular light field patches of size $n = p_x \times p_x \times p_\nu \times p_\nu$. Given a large set of such patches, randomly chosen from a collection of training light fields, a dictionary $\mathcal{D} \in \mathbb{R}^{n \times d}$ can be learned as

$$\arg \min_{\mathcal{D}, \mathcal{A}} \|\mathbf{L} - \mathcal{D}\mathcal{A}\|_F \text{ subject to } \forall j, \|\alpha_j\|_0 \leq k \quad (4.1)$$

where $\mathbf{L} \in \mathbb{R}^{n \times q}$ is a training set comprised of q light field patches and $\mathcal{A} = [\alpha_1, \dots, \alpha_q] \in \mathbb{R}^{d \times q}$ is a set of k -sparse coefficient vectors. The Frobenius matrix norm is $\|\mathbf{X}\|_F^2 = \sum_{ij} x_{ij}^2$, the ℓ_0 pseudo-norm counts the number of nonzero elements in a vector, and k ($k \ll d$) is the sparsity level we wish to enforce.

As noted by Marwah et al., training sets for the dictionary learning process are extremely large and often contain redundancy. Solving Equation 4.1, however, is computationally expensive. *Coresets* are a cheap means to reduce large dictionary training sets to manageable sizes. We follow Feigin et al. [55] in choosing a subset of training samples \mathbf{L} that have high variance.

4.3.2 Reconstruction

We choose to follow Marwah et al. [135] and apply nonlinear sparse coding techniques to recover a high-resolution 4D light field from the same measurements. This is done by representing the light field using an overcomplete dictionary as $\mathbf{l} = \mathbf{D}\boldsymbol{\chi}$, where $\mathbf{D} \in \mathbb{R}^{n \times d}$ is a dictionary of light field atoms and $\boldsymbol{\chi} \in \mathbb{R}^d$ are the corresponding coefficients. Natural light fields have been shown to be sparse in such dictionaries [135], i.e. the light field can be represented as a weighted sum of a few light field atoms (columns of the dictionary). For robust reconstruction, a basis pursuit denoise problem (BPDN) is solved

$$\begin{aligned} \arg \min_{\boldsymbol{\chi}} \quad & \|\boldsymbol{\chi}\|_1 \\ \text{subject to} \quad & \|\mathbf{l} - \Phi\mathbf{D}\boldsymbol{\chi}\|_2 \leq \epsilon, \end{aligned} \tag{4.2}$$

where ϵ is the sensor noise level. Whereas this approach offers significantly increased light field resolution, it comes at an increased computational cost. Note that Equation 4.2 is applied to a small, sliding window of the recorded data, each time recovering a small 4D light field patch rather than the entire 4D light field at once.

4.4 A Switchable Light Field Camera

This section introduces the image formation model for ASP devices. In developing the mathematical foundation for these camera systems, we entertain two goals: to place the camera in a framework that facilitates comparison to existing light field cameras, and to understand the plenoptic sampling mechanism of the proposed camera.

4.4.1 Light Field Acquisition with ASPs

The Talbot effect created by periodic gratings induces a sinusoidal angular response from ASPs [172]. For a one-dimensional ASP, this can be described as

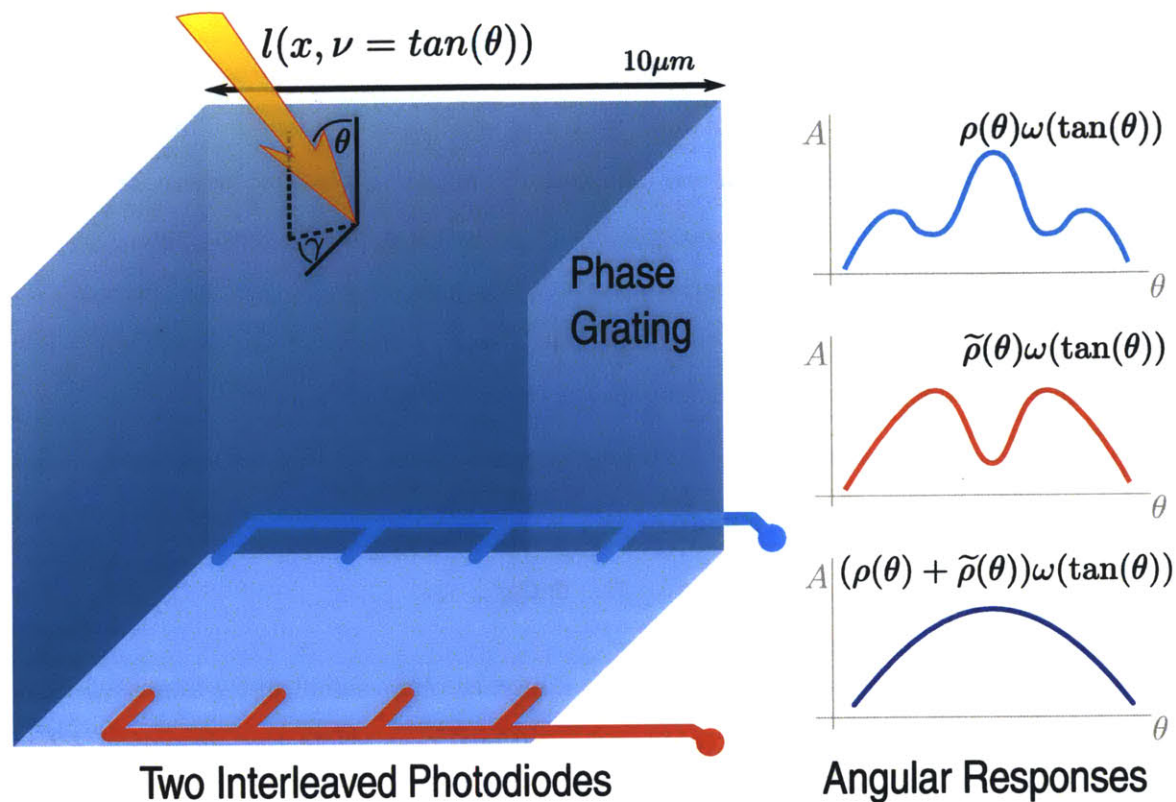


Figure 4-2: Schematic of a single angle sensitive pixel. Two interleaved photodiodes capture a projection of the light field incident on the sensor (left). The angular responses of these diodes are complementary: a conventional 2D image can be synthesized by summing their measurements digitally (right).

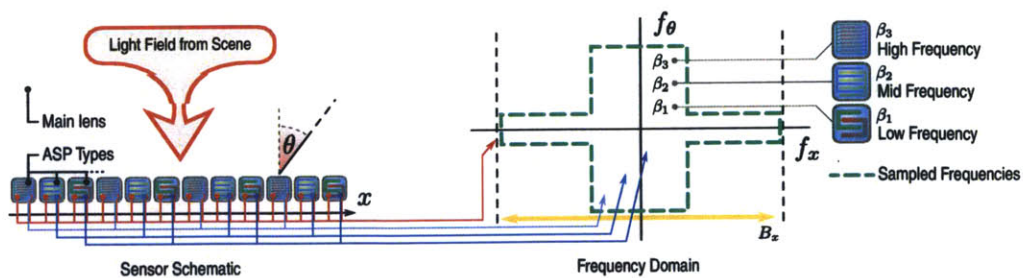


Figure 4-3: Illustration of ASP sensor layout (left) and sampled spatio-angular frequencies (right). The pictured sensor interleaves three different types of ASPs. Together, they sample all frequencies contained in the dashed green box (right). A variety of light field reconstruction algorithms can be applied to these measurements, as described in the text.

$$\rho^{(\alpha,\beta)}(\theta) = 1/2 + m/2 \cos(\beta\theta + \alpha). \quad (4.3)$$

Here, α and β are phase and frequency, respectively, m is the modulation efficiency, and θ is the angle of incident light. Specific values of these parameters used in our experimental setup can be found in Section 4.4.4. Both α and β can be tuned in the sensor fabrication process [198]. Common implementations choose ASP types with $\alpha \in 0, \pi/2, \pi, 3\pi/4$. We note that prior publications describe the ASP response without the normalization constant of $1/2$ introduced here. Normalizing Equations 4.3 and 4.4 simplifies the discussion of 2D image recovery using ASPs.

Similarly, 2D ASP implementations exhibit the resulting angular responses for incident angles θ_x and θ_y :

$$\rho^{(\alpha,\beta,\gamma)}(\boldsymbol{\theta}) = 1/2 + m/2 \cos(\beta(\cos(\gamma)\theta_x + \sin(\gamma)\theta_y) + \alpha), \quad (4.4)$$

where α is phase, β frequency, and γ grating orientation.

The captured sensor image i is then a projection of the incident light field l weighted by the angular responses of a mosaic of ASPs:

$$i(\mathbf{x}) = \int_{\mathcal{V}} l(\mathbf{x}, \boldsymbol{\nu}) \rho(\mathbf{x}, \tan^{-1}(\boldsymbol{\nu})) \omega(\boldsymbol{\nu}) d\boldsymbol{\nu}. \quad (4.5)$$

In this formulation, $l(\mathbf{x}, \boldsymbol{\nu})$ is the light field inside the camera behind the main lens. We describe the light field using a relative two-plane parameterization [51], where $\boldsymbol{\nu} = \tan(\boldsymbol{\theta})$. The integral in Equation 4.5 contains angle-dependent vignetting factors $\omega(\boldsymbol{\nu})$ and the aperture area \mathcal{V} restricts the integration domain. Sensor noise is discounted in this idealized representation, though it is addressed during discretization below. Finally, the spatial coordinates $\mathbf{x} = \{x, y\}$ are defined on the sensor pixel-level; the geometrical microstructure of ASP gratings and photodiodes is not observable at the considered scale.

In practice, the spatially-varying pixel response function $\rho(\mathbf{x}, \boldsymbol{\theta})$ is a periodic mosaic of a few different ASP types. A common example of such a layout for color imaging is the Bayer filter array that interleaves red, green, and blue subpixels. ASPs with different parameters (α, β, γ) can be fabricated following this scheme. Mathematically, this type of spatial multiplexing is formulated as

$$\rho(\mathbf{x}, \boldsymbol{\theta}) = \sum_{k=1}^N \left(\mathbb{I}\mathbb{I}^{(k)}(\mathbf{x}) * \rho^{(\zeta(k))}(\boldsymbol{\theta}) \right), \quad (4.6)$$

where $*$ is the convolution operator and $\mathbb{I}\mathbb{I}^{(k)}(\mathbf{x})$ is a sampling operator consisting of a set of Dirac impulses describing the spatial layout of one type of ASP. A total set of N types is distributed in a regular grid over the sensor. The parameters of each are given by the mapping function $\zeta(k) : \mathbb{N} \rightarrow \mathbb{R}^3$ that assigns a set of ASP parameters (α, β, γ) to each index k .

Whereas initial ASP sensor designs use two layered, attenuating diffraction gratings and conventional photodiodes underneath [197, 198, 61], more recent versions enhance the quantum efficiency of the design by using a single phase grating and an interleaved pair of photodiodes [172]. For the proposed switchable light field camera, we illustrate the latter design with the layout of a single pixel in Figure 4-2.

In this sensor design, each pixel generates two measurements: one that has an angular response described by Equation 4.4 and another one that has a complementary angular response $\tilde{\rho} = \rho^{(\alpha+\pi, \beta, \gamma)}$ whose phase is shifted by π .

The discretized version of the two captured images can be written as a simple matrix-vector product:

$$\mathbf{i} = \mathbf{\Phi}\mathbf{l} + \boldsymbol{\epsilon}, \quad (4.7)$$

where $\mathbf{i} \in \mathbb{R}^{2p}$ is a vector containing both images $i(\mathbf{x})$ and $\tilde{i}(\mathbf{x})$, each with a resolution of p pixels, and $\mathbf{\Phi} \in \mathbb{R}^{2p \times \mathbb{R}^n}$ is the projection matrix that describes how the discrete, vectorized

light field $\mathbf{l} \in \mathbb{R}^n$ is sensed by the individual photodiodes. In Equation 4.7, sensor noise is modeled as Gaussian, i.i.d., and represented by ϵ .

4.4.2 Synthesis

In this section, we propose three alternative ways to process the data recorded with an ASP sensor.

2D Image Synthesis

As illustrated in Figure 4-2, the angular responses of the complementary diodes in each pixel can simply be summed to generate a conventional 2D image, i.e. $\rho^{(\alpha,\beta,\gamma)} + \tilde{\rho}^{(\alpha,\beta,\gamma)}$ is a constant. Hence, Equation 4.5 reduces to the conventional photography equation:

$$i(\mathbf{x}) + \tilde{i}(\mathbf{x}) = \int_{\mathcal{V}} l(\mathbf{x}, \boldsymbol{\nu}) \omega(\boldsymbol{\nu}) d\boldsymbol{\nu}, \quad (4.8)$$

which can be implemented in the camera electronics. Equation 4.8 shows that a conventional 2D image can easily be generated from an ASP sensor. While this may seem trivial, existing light field camera architectures using microlenses or coded masks cannot easily synthesize a conventional 2D image for in-focus and out-of-focus objects.

Linear Light Field Synthesis

Using a linear reconstruction framework, the same data can alternatively be used to recover a low-resolution 4D light field. We model light field capture by an ASP sensor as Equation 4.7 where the rows of Φ correspond to vectorized 2D angular responses of different ASPs. These angular responses are either sampled uniformly from Equation 4.4 or fit empirically from measured impulses responses. The approximate orthonormality of the angular wavelets (see Section 3.4.3) implies $\Phi^T \Phi \approx \mathbf{I}$. Consequently $\Sigma = \text{diag}(\Phi^T \Phi)$ is used as a preconditioner for inverting the capture equation: $\mathbf{l} = \Sigma^{-1} \Phi^T \mathbf{i}$.

The main benefit of a linear reconstruction is its computational performance. However, the spatial resolution of the resulting light field will be approximately k -times lower than that of the sensor ($k = n/p$) since the different ASPs are grouped into tiles on the sensor. Similarly to demosaicing from color filter arrays, different angular measurements from the ASP sensor can be demosaiced using interpolation and demultiplexing [204] to improve visual appearance. In addition, recent work on light field super-resolution has demonstrated that resolution loss can be slightly mitigated for the particular applications of image refocus [196] and volume reconstruction [28].

Sparse Coding for High-resolution Light Fields

To achieve high resolution reconstructions, albeit at an increased computational cost, we employ the sliding window technique described in Section 4.3.2. In particular, window blocks with typical sizes of 9×9 pixels are processed in parallel to yield light field patches with $9 \times 9 \times 5 \times 5$ rays each. See Section 4.4.4 for implementation details.

4.4.3 Analysis

In this section, we analyze the proposed methods and compare them to alternative light field sensing approaches.

Frequency Analysis

As discussed in the previous section, Angle Sensitive Pixels sample a light field such that a variety of different reconstruction algorithms can be applied to the same measurements. To understand the information contained in the measurements, we can turn to a frequency analysis. Figure 4-3 (left) illustrates a one-dimensional ASP sensor with three interleaved types of ASPs sampling low, mid, and high angular frequencies, respectively. As discussed in Section 4.4.2, the two measurements from the two interdigitated diodes in each pixel can

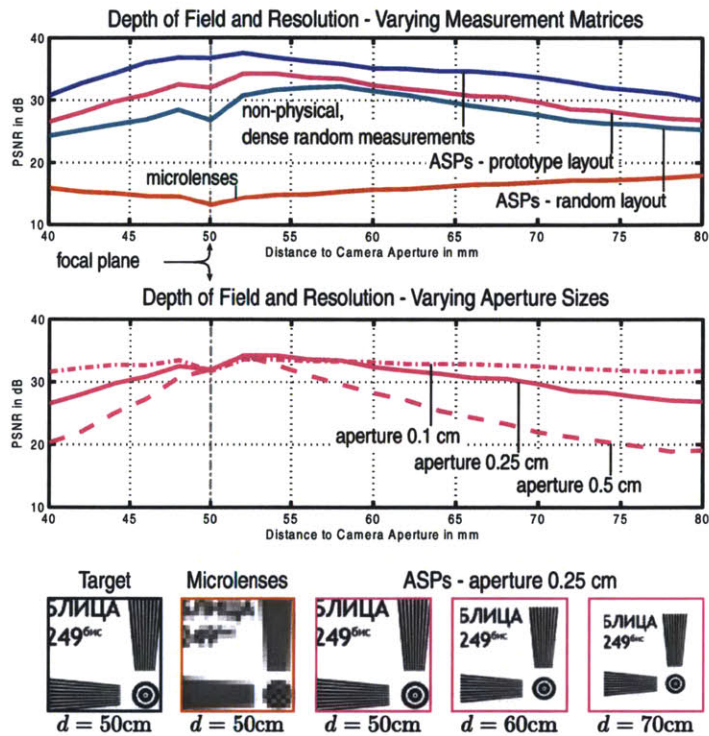


Figure 4-4: Evaluating depth of field. Comparing the reconstruction quality of several different optical setups shows that the ASP layout in the prototype camera is well-suited for sparsity-constrained reconstructions using overcomplete dictionaries (top). The dictionaries perform best when the parallax in the photographed scene is smaller or equal to that of the training light fields (center). Central views of reconstructed light fields are shown in the bottom.

be combined to synthesize a conventional 2D image. This image has no angular information but samples the entire spatial bandwidth B_x of the sensor (Figure 4-3 right, red box).

The measurements of the individual photodiodes contain higher angular frequency bands, but only for lower spatial frequencies due to the interleaved sampling pattern (Figure 4-3 right, solid blue boxes). A linear reconstruction (Section 4.4.2) would require an optical anti-aliasing filter to be mounted on top of the sensor, as is commonly found in commercial sensors. In the absence of an optical anti-aliasing filter, aliasing is observed. For the proposed application, aliasing results in downmixing of high spatio-angular frequencies (Figure 4-3 right, hatched blue boxes) into lower spatial frequency bins. As spatial frequencies are sampled by an ASP sensor while angular frequencies are measured continuously, aliasing occurs only among spatial frequencies. The region of the spatio-angular frequency plane sampled by the ASP sensor in Figure 4-3 is highlighted by the dashed green box. Although aliasing makes it difficult to achieve high-quality reconstructions with simple linear demosaicing, it is crucial in preserving information for nonlinear, high-resolution reconstructions based on sparsity-constrained optimization (Section 4.4.2).

Depth of Field

To evaluate the depth of field that can be achieved with the proposed sparsity-constrained reconstruction methods, we simulate a two-dimensional resolution chart at multiple different distances to the camera's focal plane. The results of our simulations are documented in Figure 4-4. The camera is focused at 50 cm, where no parallax is observed in the light field. At distances closer to the camera or farther away the parallax increases—we expect the reconstruction algorithms to achieve a lower peak signal-to-noise ratio (PSNR). The PSNR is measured between the depth-varying target 4D light field and the reconstructed light field.

Figure 4-4 (top) compares sparsity-constrained reconstructions using different measurement matrices and also a direct sampling of the low-resolution light field using microlenses (red plot). Slight PSNR variations in the latter are due to the varying size of the resolution

chart in the depth-dependent light fields, which is due to the perspective of the camera (cf. bottom images). Within the considered depth range, microlenses always perform poorly.

The different optical setups tested for the sparsity-constrained reconstructions include the ASP layout of our prototype (magenta plot, described in Section 4.4.4), ASPs with completely random angular responses that are also randomized over the sensor (green plot), and also a dense random mixing of all light rays in each of the light field patches (blue plot). A dense random mixing across a light field patch requires that each measurement within the patch is a random mixture of all spatial and angular samples that fall within the patch. Though such a mixture is not physically realizable, it does yield an intuition of the approximate achievable upper performance bounds. Unsurprisingly, such a dense, random measurement matrix Φ performs best. What is surprising, however, is that random ASPs are worse than the choice of regularly-sampled angular wavelet coefficients in our prototype (see Section 4.4.4). For compressive sensing applications, the rows of the measurement matrix Φ should be as incoherent (or orthogonal) as possible to the columns of the dictionary \mathcal{D} . For the particular dictionary used in these experiments, random ASPs seem to be more coherent with the dictionary. These findings are supported by Figure 4-5. We note that the PSNR plots are content-dependent and also dependent on the employed dictionary.

The choice of dictionary is critical. The one used in Figure 4-4 is learned from 4D light fields showing 2D planes with random text within the same depth range as the resolution chart. If the aperture size of the simulated camera matches that used in the training set (0.25 cm), we observe high reconstruction quality (solid line, center plots). Smaller aperture sizes will result in less parallax and can easily be recovered as well, but resolution charts rendered at larger aperture sizes also contain a larger amount of parallax than any of the training data. The reconstruction quality in this case drops rapidly with increasing distance to the focal plane (Figure 4-4, center plots).

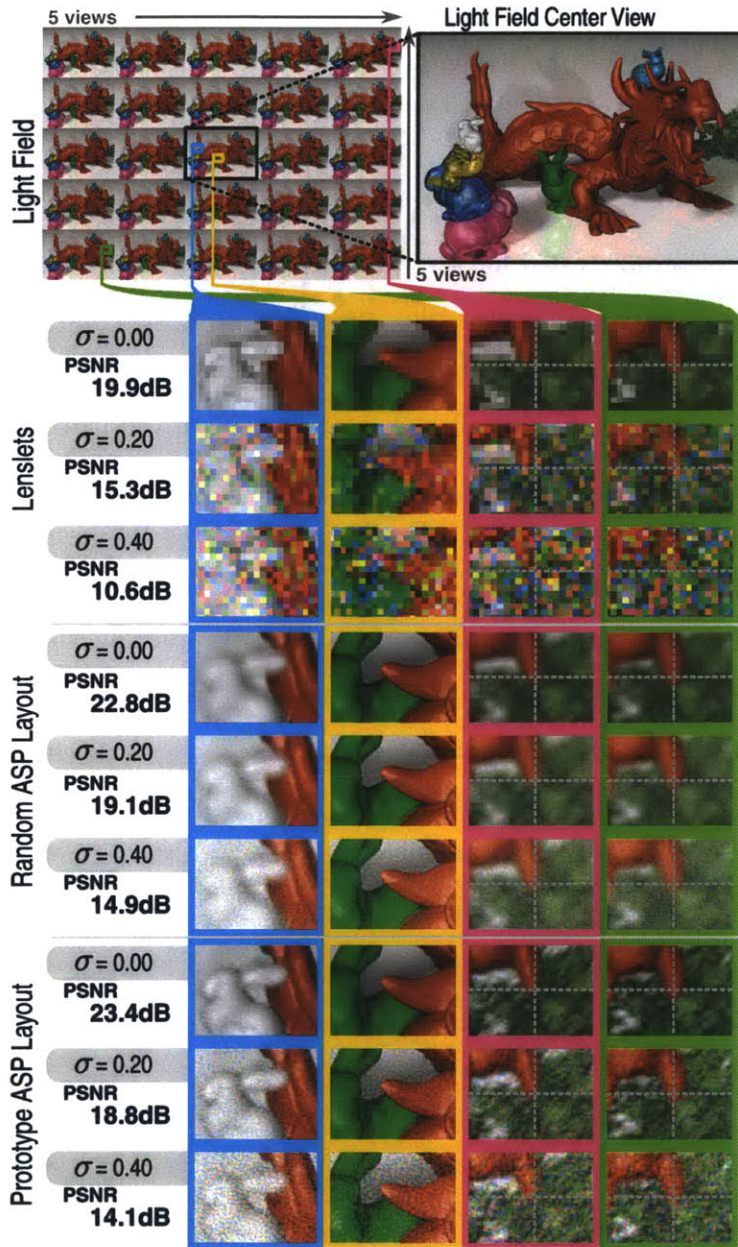


Figure 4-5: Simulated light field reconstructions from a single coded sensor image for different levels of noise and three different optical sampling schemes. For the ASP layout in the prototype camera (bottom), high levels of noise result in noisy reconstructions—parallax is faithfully recovered (dragon’s teeth, lower right, fiducials added). A physically-realizable random ASP layout (center) does not measure adequate samples for a sparse reconstruction to recover a high-quality light field from a single sensor image; the reconstructions look more blurry and parallax between the views is poorly recovered (center, right). A standard lenslet-based reconstruction (top) subsamples spatial information. Noise is more apparent in the lenslet case as BPDN attenuates noise in the other cases. In all cases, the peak sensor measurement magnitude is normalized on $[0 \ 1]$ prior to adding Gaussian noise.

Resilience to Noise

Finally, we evaluate the sparse reconstruction algorithm proposed in Section 4.4.2 w.r.t. noise and compare three different optical sampling schemes. Figure 4-5 shows a synthetic light field with 5×5 different views. We simulate sensor images with zero-mean i.i.d. Gaussian noise and three different standard deviations $\sigma = \{0.0, 0.2, 0.4\}$. In addition, we compare the ASP layout of the prototype (see Section 4.4.4) with a random layout of ASPs that each also have a completely random angular response. Confirming the depth of field plots in Figure 4-4, a random ASP layout achieves a lower reconstruction quality than sampling wavelet-type angular basis functions on a regular grid. Again, this result may be counter-intuitive because most compressive sensing algorithms perform best when random measurement matrices are used. However, these usually assume a dense random matrix Φ (simulated in Figure 4-4), which is not physically realizable in an ASP sensor. One may believe that a randomization of the available degrees of freedom of the measurement system may be a good approximation of the fully random matrix, but this is clearly not the case. We have not experimented with optical layouts that are optimized for a particular dictionary [135], but expect such codes to further increase reconstruction quality.

4.4.4 Implementation

Angle Sensitive Pixel Hardware

A prototype ASP light field camera was built using an angle sensitive pixel array sensor [199]. The sensor consists of 24 different ASP types, each of which has a unique response to incident angle described by Equation 4.4. Since a single pixel generates a pair of outputs, a total of 48 distinct angular measurements are read out from the array. Recall from Section 4.4.1 that ASP responses are characterized by the parameters α , β , γ , and m which define the phase, two dimensional angular frequency, and modulation efficiency of the ASP. The design includes three groups of ASPs that cover low, medium, and high frequencies with β values of 12, 18 and 24, respectively. The low and high frequency groups of ASPs have orientations (γ

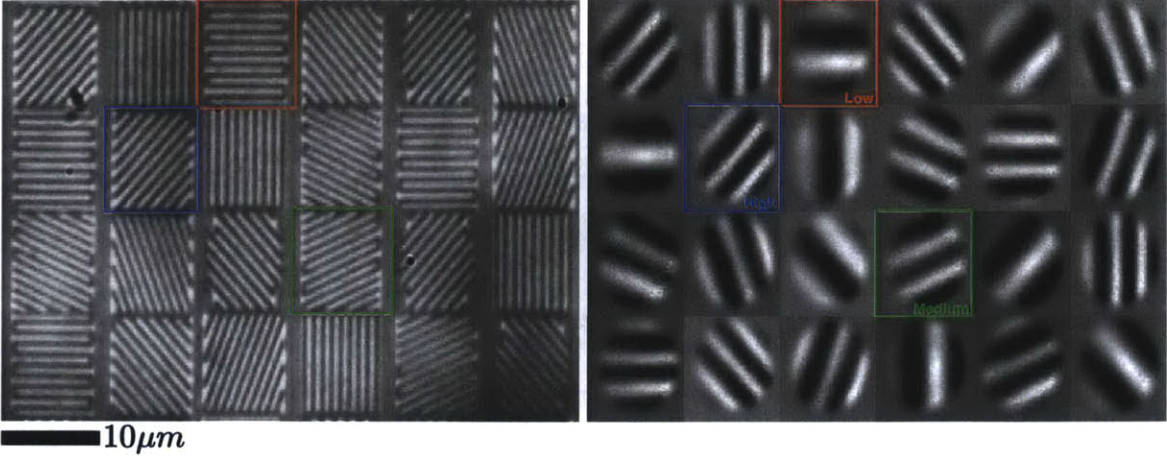


Figure 4-6: Microscopic image of a single 6×4 pixel tile of the ASP sensor (left). We also show captured angular point spread functions (PSFs) of each ASP pixel type (right).

in degrees) of 0° , 90° and $\pm 45^\circ$ whereas the mid frequency group is staggered in frequency space with respect to the other two and has γ values of $\pm 22.5^\circ$ and $\pm 67.5^\circ$. Individual ASPs are organized into a rectangular unit cell that is repeated to form the array. Within each tile, the various pixel types are distributed randomly so that any patch of pixels has a uniform mix of orientations and frequencies as illustrated in Figure 4-6. The modulation efficiency, m , is a process parameter and typical values are measured to be near 0.5 with some dependence on wavelength [197]. The die size is 5×5 mm which accommodates a 96×64 grid of tiles, or 384×384 pixels.

In addition to the sensor chip, the only optical component in the camera is the focusing lens. We used a commercial 50 mm Nikon manual focus lens at an aperture setting of $f/1.2$. The setup, consisting of the data acquisition boards that host the imager chip, and the lens, can be seen in Figure 4-1. The target imaging area was staged at a distance of 1m from the sensor which provided a 10:1 magnification. Calibration of the sensor response was performed by imaging a 2mm diameter, back-illuminated hole positioned far away from the focal plane. Figure 4-6 shows the captured angular point spread function for all 24 ASP types. These responses were empirically fitted and resampled to form the rows of the projection matrix Φ for both the linear and nonlinear reconstructions on captured data.

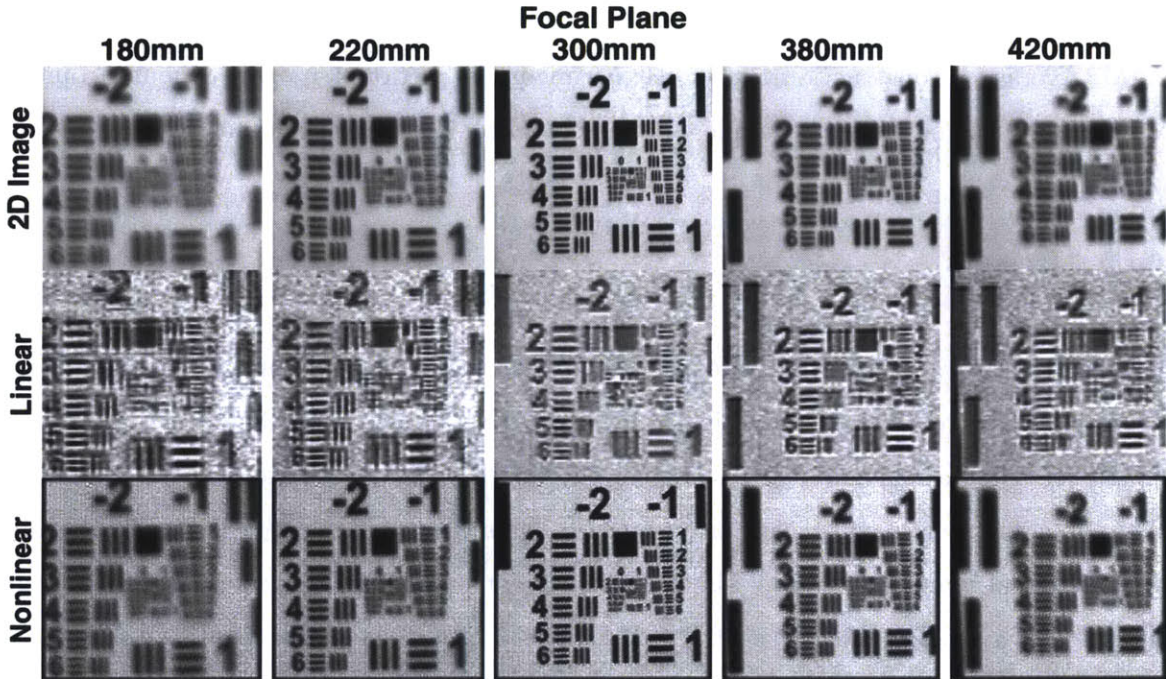


Figure 4-7: Evaluation of prototype resolution. We capture images of a resolution target at different depths and compare the 2D image (top), center view of the linearly reconstructed light field (center), and center view of the nonlinearly reconstructed light field (bottom).

Software

The compressive part of our software pipeline closely follows that of Marwah et al. [135]. Conceptually, nonlinear reconstructions depend on an offline dictionary learning phase, followed by an online reconstruction over captured data. To avoid the challenges of large-scale data collection with our prototype hardware, we used the dictionaries provided by Marwah et al. to reconstruct light fields from the prototype hardware. Dictionaries used to evaluate depth of field in Figure 4-4 were learned using KSVD [4].

Online reconstruction was implemented by the Alternating Direction Method of Multipliers (ADMM) [24] with parameters $\lambda = 10^{-5}$, $\rho = 1$, and $\alpha = 1$, to solve the ℓ_1 -regularized regression (BPDN) of Equation 4.2. ASP sensor images were subdivided into sliding, 9×9 pixel windows; small 4D light field patches were reconstructed for each window, each with 5×5 angles. The sliding reconstruction window was translated in one pixel increments over the full 384×384 pixel sensor image and the results were integrated with an average

filter. Reconstructions were computed on an 8-core Intel Xeon workstation with 16GB of RAM. Average reconstruction time for experiments in Section 4.4.5 was 8 hours. Linear reconstruction algorithms are significantly faster, taking less than one minute for each result.

4.4.5 Results

This section shows an overview of experiments with the prototype camera. In Figure 4-7, we evaluate the resolution of the device for all three proposed reconstruction algorithms. As expected for a conventional 2D image, the depth of field is limited by the f-number of the imaging lens, resulting in out-of-focus blur for a resolution chart that moves away from the focal plane (top row). The proposed linear reconstruction recovers the 4D light field at a low resolution (center row). Due to the lack of an optical anti-aliasing filter in the camera, aliasing is observed in the reconstructions. The anti-aliasing filter would remove these artifacts but also decrease image resolution. The resolution of the light field recovered using the sparsity-constrained nonlinear methods has a resolution comparable to the in-focus 2D image. Slight artifacts in the recovered resolution charts correspond to those observed in noise-free simulations (cf. Figure 4-5). We believe these artifacts are due to the large compression ratio—25 light field views are recovered from a single sensor image via sparsity-constrained optimization.

We show additional comparisons of the three reconstruction methods for a more complex scene in Figure 4-8. Though an analytic comparison of resolution improvement by our nonlinear method is not currently possible, referring to Figure 4-4 (top) at the focal plane depth yields a numerical comparison for a simulated resolution chart.

Figure 4-9 shows several scenes that we captured in addition to those already shown in Figures 4-1 and 4-8. Animations of the recovered light fields for all scenes can be found in the ICCP 2014 video. We deliberately include a variety of effects in these scenes that are not easily captured in alternatives to light field imaging (e.g., focal stacks or range imaging), including occlusion, refraction, and translucency. Specular highlights, as for instance seen on the glass piglet in the two scenes on the right, often lead to sensor saturation, which

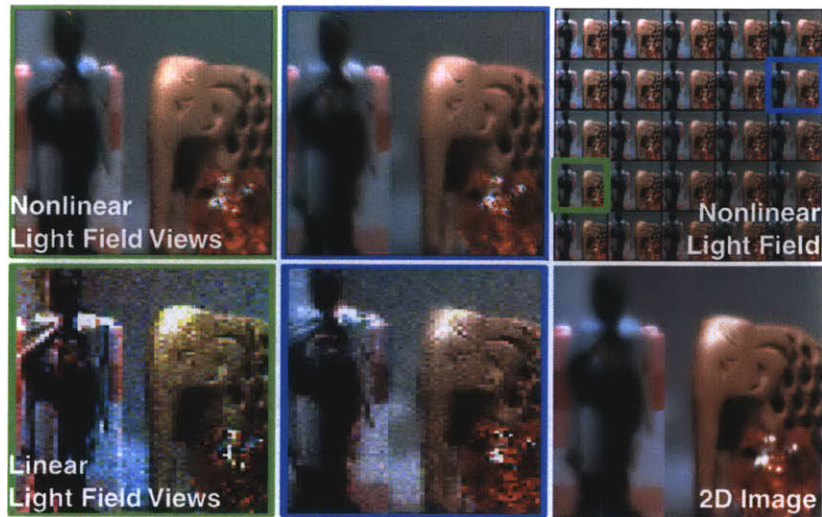


Figure 4-8: Comparison of different reconstruction techniques for *the same captured data*. We show reconstruction of a 2D image (bottom right), a low-resolution light field via linear reconstruction (bottom left and center), and a high-resolution light field via sparsity-constrained optimization with overcomplete dictionaries (top). Whereas linear reconstruction trades angular for spatial resolution—thereby decreasing image fidelity—nonlinear reconstructions can achieve an image quality that is comparable to a conventional, in-focus 2D image for each of 25 recovered views.

causes artifacts in the reconstructions. This is a limitation of the proposed reconstruction algorithms.

Finally, we show in Figure 4-10 that the recovered light fields contain enough parallax to allow for post-capture image refocus. Chromatic aberrations in the recorded sensor image and a limited depth of field of each recovered light field view place an upper limit on the resolvable resolution of the knight (right).



Figure 4-9: Overview of captured scenes showing mosaics of light fields reconstructed via sparsity-constrained optimization (top), a single view of these light fields (center), and corresponding 2D images (bottom). These scenes exhibit a variety of effects, including occlusion, refraction, specularly, and translucency. The resolution of each of the 25 light field views is similar to that of the conventional 2D images.



Figure 4-10: Refocus of the “Knight & Crane” scene.

Chapter 5

Compressive 8D Display

5.1 About This Chapter

In the previous sections of this thesis, we have presented separate theory for compressive light field display and compressive light field capture. Each of the frameworks developed has been backed up by prototypical implementations. As articulated in Chapter 1, the goal of this thesis is to develop a road map towards implementing a practical 8D display—one capable of satisfying the demands of the human visual system. This chapter takes a different tack: developing separately the motivation, in the form of prototyped applications on a classical device, and proposing theoretical hardware devices to support future compressive 8D displays.

There are practical and philosophical reasons for adopting this approach in the thesis. Sections 3.1 and 4.1 lay out the requirements for compressive display and capture, respectively. One of the key assumptions in both of these cases is that the information to be displayed or captured is highly redundant, meaning that there are strong correlations within the data. Experimentally, this holds true for natural light field scenes in both capture and display applications. One approach to implementing a compressive 8D display is to physically or optically combine independent compressive display and capture systems. Just as compressive displays and compressive capture systems exploit correlations within emitted or captured

information, another choice for constructing a compressive 8D display would be to attempt to exploit correlations between the ensemble input and output data. However, for arbitrary applications of an 8D display, there is little reason to expect strong correlation between the content on the display and the environmental lighting.

A second approach, and the one adopted in this chapter, is to consider an 8D display where the input and output channels function independently. In the case of both approaches, large technical hurdles will need to be overcome in order to create functional prototypes on the scale of the display and capture prototypes presented in Chapters 3 and 4. This is because the application of collocated 8D display demands hardware that is not yet available in the commercial (as in the case of display) or research (as in the case of capture) pipelines.

Therefore, while the second approach outlined above is reasonable given the structure of the data to be transmitted and received, there is little justification within the bounds of this thesis for undertaking the technically challenging project of implementing a compressive 8D display. In Section 5.2 we describe a prototype classical device to demonstrate interaction scenarios possible with an 8D display, and in Section 5.3 we suggest a thin, efficient hardware implementation based on the work of Chapters 3 and 4.

5.2 A Classical Method

In order to prototype a subset of the interactive applications made possible by 8D displays, we have constructed a prototype 8D display using classical, Dirac-sampled methods. This section details the implementation of the classical 8D display, and the applications developed on top of the implemented prototype.

5.2.1 Implementation

Hardware

Optics We propose to implement an 8D display by placing an array of microlenses on a SIP LCD screen. Due to the pixel pitch limits of existing SIP hardware, we implement a projector-camera system to substitute for the SIP display. The configuration of our system is shown in Figure 5-3. We place a $150mm \times 150mm$ hexagonal lens array (Fresnel Tech. sheet #360, appx. $0.5mm$ lens pitch) on top of a Grafix acetate diffuser. We then image and project onto the diffuser. This optically simulates the orthographic light field produced by the SIP display.

We use a grayscale Point Grey Gazelle 2048×2048 pixel, $120fps$ camera with a $50mm$ Schneider Xenoplan lens as the sensor in our prototype. The display element is comprised of a Sanyo PLV-Z800 1920×1280 projector with a modified lens. The projector optics are modified by shifting the lens forward by $4mm$, allowing the projector to create a focused $325dpi$ image that matches the horizontal dimension of the hexagonal lens sheet. The projector and camera share an optical path through a 40/60 beamsplitter. We prevent cross-talk between the camera and projector by multiplexing through crossed linear polarizers.

Computation The computation necessary for the 8D display is implemented in DirectX 11 HLSL running on an NVIDIA GTX 470 GPU. The GPU is hosted by an 8-core Intel i5 CPU with 8GB of RAM.

Real-time GPU Pipeline

The light field rendering and decoding necessary for the 8D display is implemented in DirectX 11 HLSL. For each output light field view, the camera matrix is updated to an appropriate shifted off-axis projection [118], and the scene geometry is rendered. At each rendering stage, each view of the captured and resampled incident light field is used for illumination. This step is accomplished through a projective texture map.

5.2.2 Assessment

Sensor-In-Pixel Displays

Much of the potential impact of this project is predicated on the existence of Sensor-In-Pixel (SIP) Liquid Crystal Displays (LCDs). In recent years LCD manufacturers have begun to introduce a variety of semiconductor technologies that combine light sensitive elements into the driver matrix for typical liquid crystal displays [27]. While our real-time prototype was certainly enabled by high-end computer graphics hardware, the optical configuration of the presented camera and projector implementation is somewhat unremarkable. However, in combination with collocated, thin, optical capture and display elements, such as those provided by a SIP LCD, this work suggests a straightforward route to achieving a thin, low-cost, commercially realizable, real-time, 8D display.

Calibration

The use of a camera and projector system necessitates a calibration step to align the sample grids of the camera and projector with the real-world coordinates of the lens sheet.

To calibrate the camera, a Fresnel lens is placed on top of the lens sheet. An acrylic guide was cut to facilitate placing a point light source at the focal point of the Fresnel lens (see Figure 5-1). The resulting collimated beam creates a lit point beneath the center of each hexagonal lens, where it can be photographed with the camera in the prototype. An offline MATLAB script is used to find the camera coordinates of each lens center point. This procedure automatically accounts for global distortion in the camera lens and beamsplitter/mirror system. A grid of 3rd order polynomial lines are fit to the grid of detected lens centers to reduce the contribution of local intensity variation caused by non-uniformity in the diffuser sheet.

The projector is calibrated using the moir'e magnifier [92] effect. Though this method can only account for scale and rotation variation, we found in practice that lens distortion in the

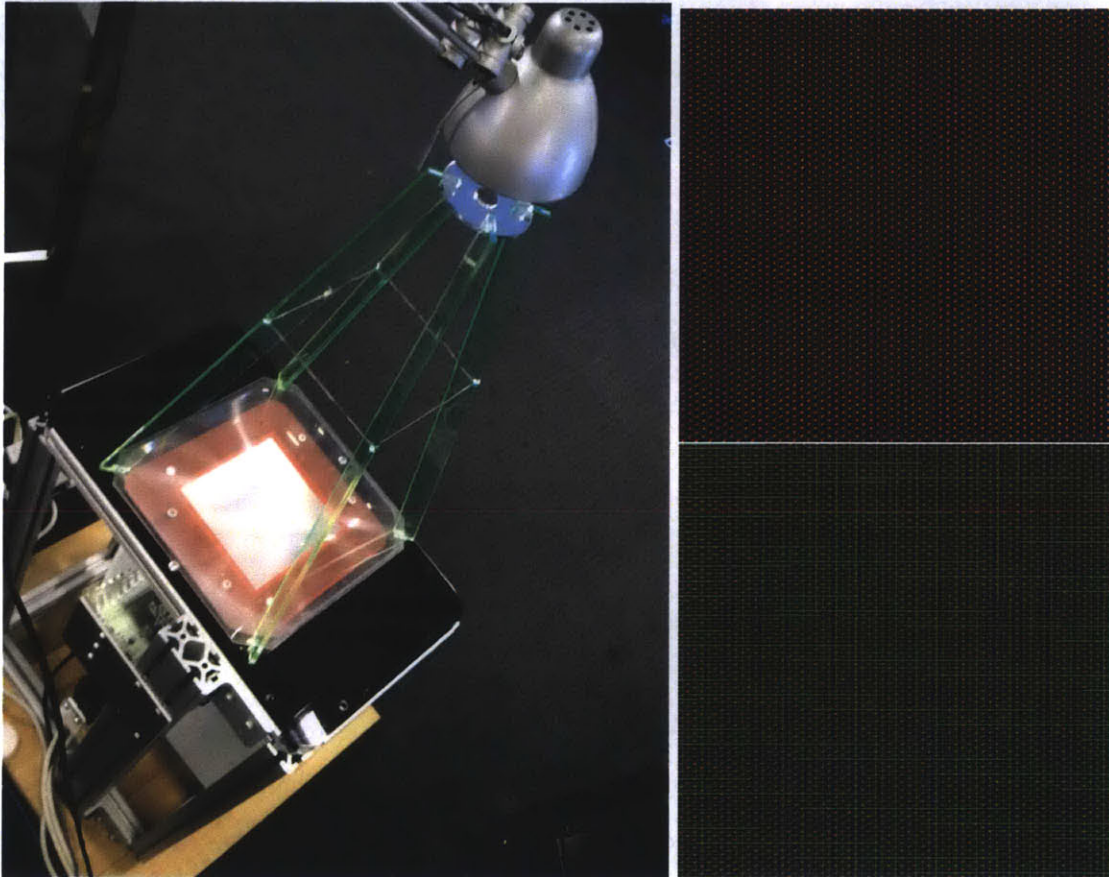


Figure 5-1: (Left) A Fresnel lens is placed on top of the 8D display prototype to facilitate calibration of the input side. Placing a point light source at the annular mirror disk ensures a collimated beam is emitted from the Fresnel lens. (Top, Right) The camera view of the calibration image. (Bottom, Right) A grid of polylines fit to the lens-center grid.

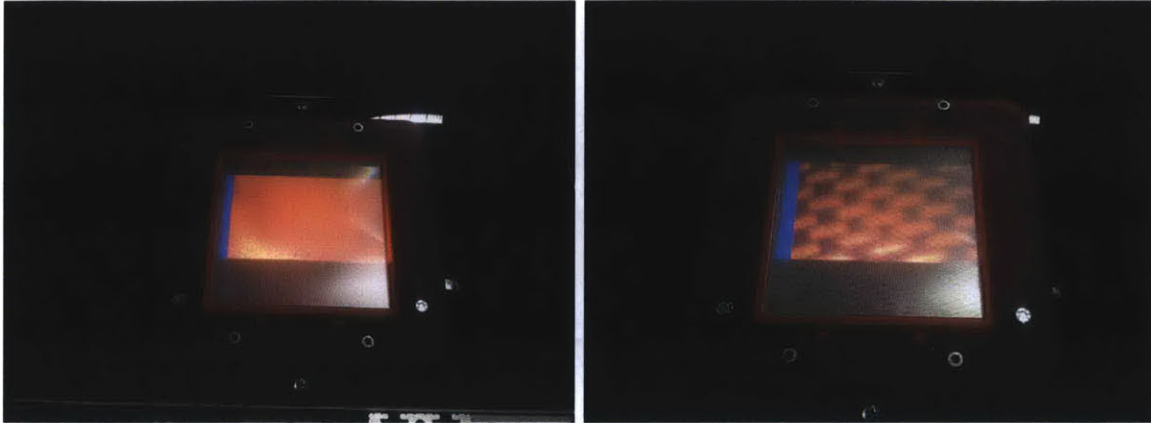


Figure 5-2: (Top) A correctly calibrated projector image. (Bottom) An incorrectly scaled and rotated projector image.

projector was negligible. A hexagonal grid of red dots on a black background is projected at the expected lens center locations. The user places her eye above the display and adjusts the scale and rotation until the central view above the lens sheet is solid red. See Figure 5-2 for reference.

Performance

Our prototype supports 7×7 views optically. However, due to limitations of our GPU pipeline, we are able to support only 5×5 views in real time. Though our output images are resampled onto a hexagonal grid in our GPU pipeline in order to accommodate the hexagonal lens array, the approximate equivalent rectilinear resolution of our display is 274×154 per view. With a $3mm$ focal length, the lens array offers a 19° field-of-view. We have characterized the depth of field of the system empirically, and can obtain satisfactory results for objects extending up to $3cm$ from the display surface.

5.2.3 Prototyped Applications

The importance of lighting in perception has long been recognized in photography, and computer graphics. It has been studied in detail with respect to the human visual system

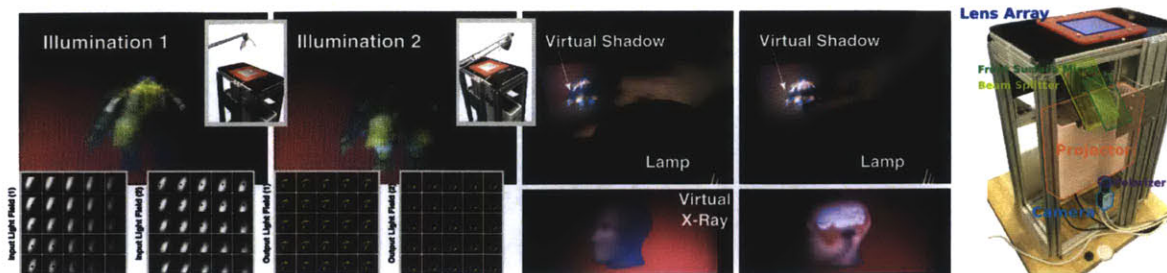


Figure 5-3: The 8D prototype allows for glasses-free display of 3D content, whilst simultaneously capturing any incident light for re-illumination and interaction. From left to right: A user shines a light source (lamp) at the 3D display, and the rendered 3D model is correctly re-illuminated. Virtual shadows can be cast by placing a finger between the display and light source. This allows any light source to act as an input controller, for example to allow intuitive interaction with a medical data set. 8D works by simultaneously capturing and displaying a 4D light field (as shown bottom left inset).

[1, 56], and plays a central role in our understanding of the world. Our goal in creating an 8D Display is to take a step towards displays that can produce the convincing illusion of physical reality. A key aspect of this goal will be the ability of these displays to react to incident environmental lighting in a realistic and believable way. Going one step further, beyond the reproduction of physical reality, it is possible to take advantage of the computational nature of an 8D display to break physical laws to render non-physical scenes that fulfill interface or interaction goals. We mock up two interaction scenarios to demonstrate physical and non-physical rendering on our prototype 8D display.

Relighting

In this intuitive interaction a user moves a real light-emitting widget, such as a lamp or a flashlight, over the 8D Display prototype. An object on the display appears to be 3D, will full parallax, and responds to the incident light as the user would expect from a real object (Figure 5-4). The object, in this case a troll figure, can be set to rotate, demonstrating the real-time rendering capabilities provided by our GPU implementation.

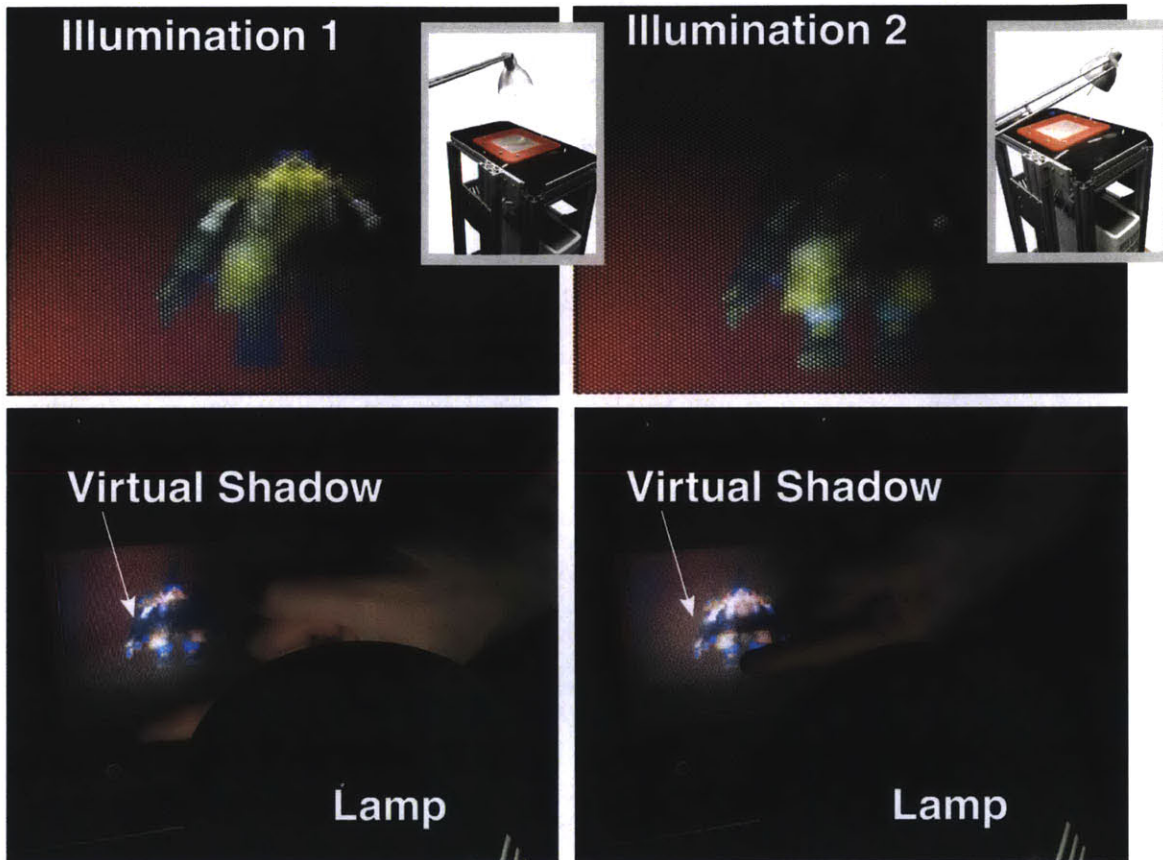


Figure 5-4: The 8D Display prototype demonstrates realistic relighting capabilities. Note how the lighting changes as the lamp is moved from one perspective (Left) to another (Right), and how the sharpness of the shadow cast changes as a function of object distance (Left vs. Right).

Hand-held X-Ray

In this interaction scenario, a hand-held light widget with adjustable intensity is used to cut through segmented structures in real MRI data. More intense light acts like a virtual x-ray beam, revealing inner structure. In the case of this demo, the skin of the patient is visible under lower light conditions, allowing a clinician to visually identify the patient. As the virtual x-ray is turned up in power (the output intensity of the real, hand-held light widget is increased) the skin layer becomes transparent, revealing the segmented brain imagery (Figure 5-3).

5.3 Architecture

This thesis is not, in particular, *about* low-level hardware development. In Figure 5-5, we propose a number of hardware approaches to create computational 8D displays. They are only sketches. The unavailability of near term hardware that simultaneously satisfies these design criteria laid out in Chapters 3 and 4 makes meaningful simulation of the designs difficult.

Below, we address two strategies available when considering the problem of 8D display.

Simple system A simple 8D display system is one in which the input and output components function independently. Optically and computationally the input and output are isolated, and the algorithms developed in Chapters 3 and 4 can be directly applied to the measurements.

Joint System The implementation of a joint compressive 8D display system assumes correlation between the light field incident on the display and the light field emitted by the display. As shown in Figure 5-6, the input and output light fields from the prototype 8D display described in Section 5.2 have strong internal correlation, as much of the energy of the distribution is encoded in just a few singular values. The singular values of the

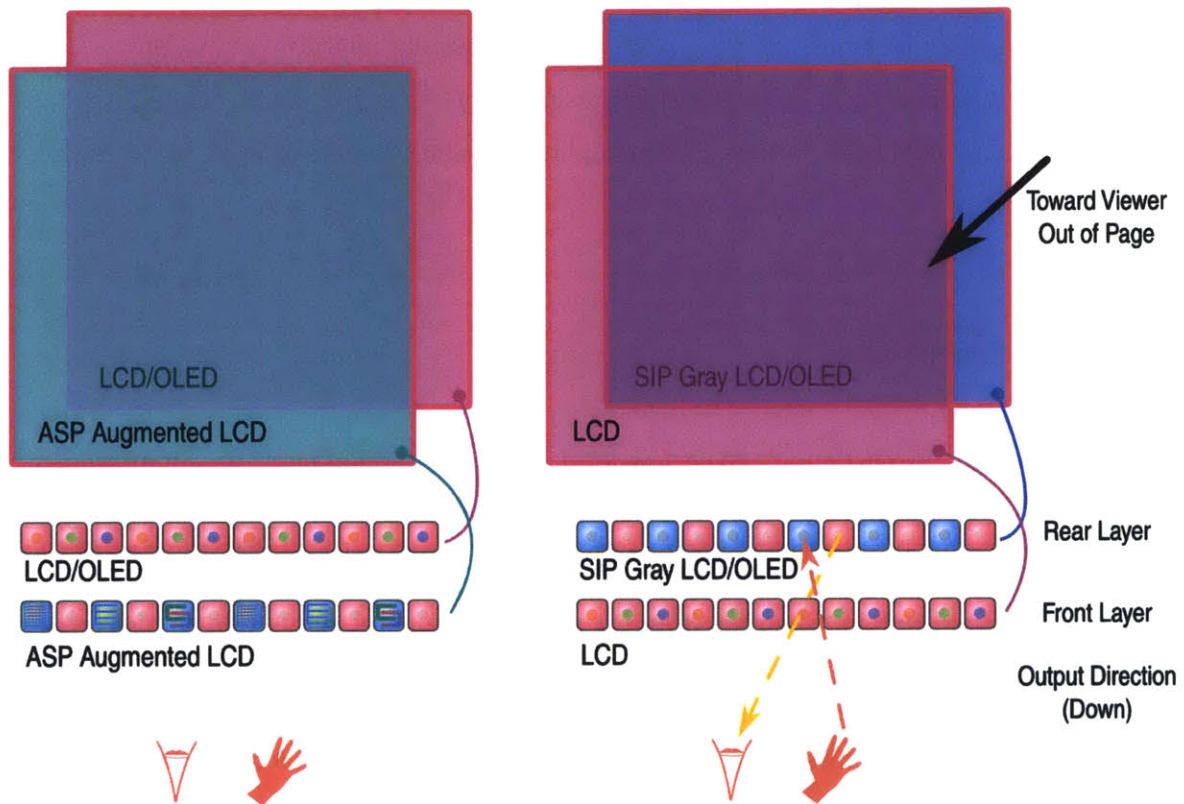


Figure 5-5: Two conceptual designs for computational 8D displays. (Left) Because the etching method used to create diffraction gratings for angle sensitive pixels is compatible with LCD glass, it may be possible to embed ASPs into the front layer of a layered light field display, creating a system in which compressive capture and compressive display function independently. (Right) Sensor-in-pixel or SIP LCDs that incorporate optical sensors into the transistor matrix that drives the LCD panel are already appearing in commercial products. Incorporating a SIP LCD into a dual layer display could create a device where the rank-1 terms of a compressive light field factorization function as pseudo-random masks for compressive light field capture. This arrangement would require new mathematics. It is possible in both cases to employ organic light emitting diodes (OLEDs) in the rear layer of each device, rather than using LCD technology.

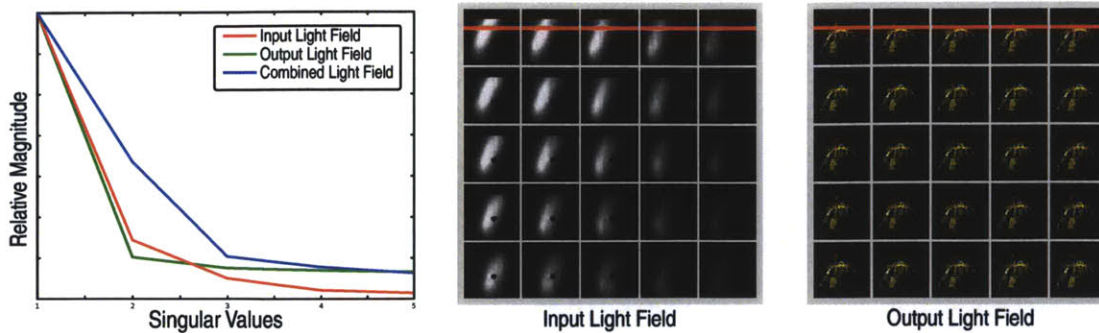


Figure 5-6: The relative magnitudes of the singular values of input (red), output (green) and combined (blue) light field slices from real data from the prototype 8D display described in Section 5.2. Note that the singular values of the combined light field slices fall off more slowly, meaning that the energy distribution among rays in the input and output light fields are less well correlated than the distributions within the input or output light field individually. The slice of each light field is depicted by the red line shown atop the input and output light fields.

combined input and output distribution fall off more slowly. This indicates that there are fewer correlations to be exploited by a computational display system that considers the ensemble of input and output rays.

Additionally, the theories presented in Chapters 3 and 4 for handling compressive display and compressive capture, while conceptually similar, are disparate at the lowest level. The difference in the structure of the two frameworks reflects the importance of the ordering of the linear optical processing and non-linear computation in the two frameworks. As such, a unifying theory, capable of singly driving an 8D display to take advantage of correlation between input and output light fields remains for future discovery.

Chapter 6

Applications and Extensions

In this chapter we present applications of compressive display to various challenges in display. In Section 6.1 we apply the Tensor Display Framework introduced in Chapter 3 to the problem of providing accommodation depth cues (Section 2.1.2) in TV sized displays. In Section 6.2 we show how the same Tensor Display Framework applies to the problem of large scale projection, suitable for theaters.

Also in this chapter, we present a conceptual extension of 8D display to the domain of audio. Just as in the optical domain, where we seek to provide a general purpose light transducer, an 8D display in the audio regime should provide a single interface capable of rendering aural phenomena into space, and accurately receiving aural phenomena from space. In Section 6.3 we present a wave-based, classical approach to creating such an 8D audio display.

6.1 Focus 3D: Accomodation

Most available 3D displays share a common limitation: lack of the focus depth cues, accommodation and retinal blur. The challenge of producing focus depth cues is that of producing an extremely high angular resolution light field. Creating such a light field over a large spatial extent remains a challenging problem in the optical and computational domains. In

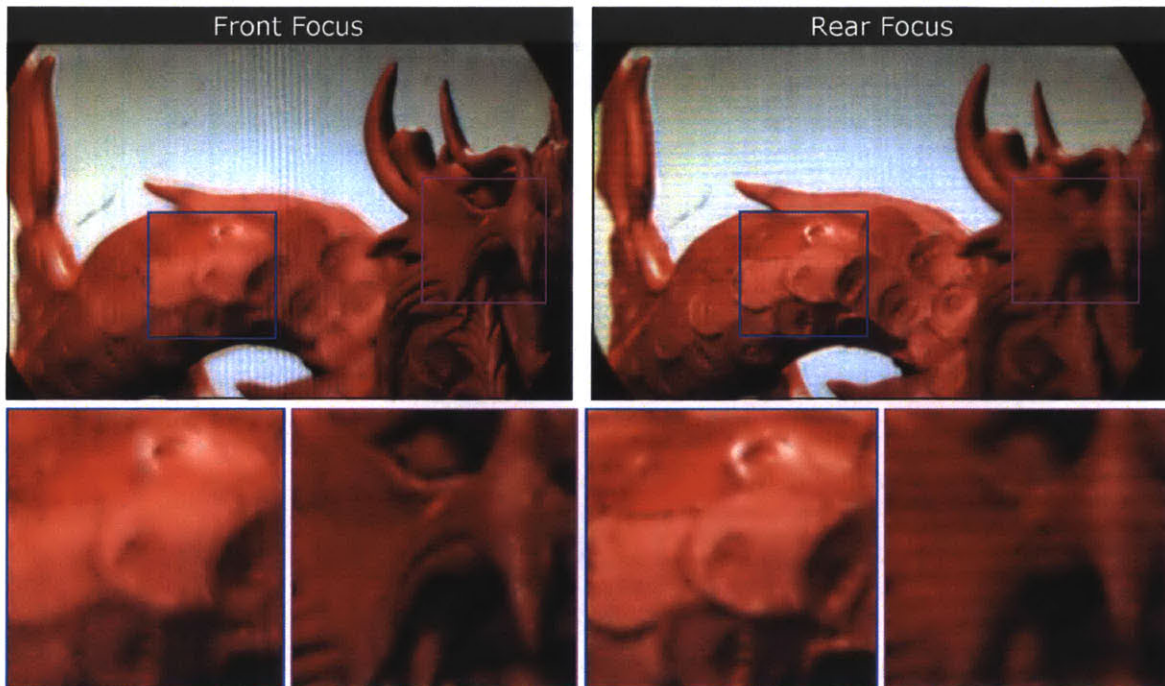


Figure 6-1: Photograph of prototype display focused at two different depths. Bottom row shows magnifications of inset regions. The prototype shown was configured with a single LCD layer placed directly in front of a high angular resolution backlight (HARB) and was photographed with a large aperture at a distance of 127 cm.

this section we provide an example application of compressive display, in combination with head tracking, to create a display capable of delivering focus depth cues.

Accommodation is an important depth cue driven by the focal state of the lens in a human eye; the ciliary muscles contract and relax to change the shape of the lens, causing a change in focus. Takaki [180] experimentally verified that projecting as few as two different perspectives in one pupil stimulates accommodative responses in a human observer.

Retinal blur is a complementary depth cue stimulated by the sensed magnitude of focal blur on the retina; inclusion of this cue has been shown to improve the performance of certain visual tasks [84]. When these focus cues are correct or nearly correct (i.e., they closely match the depths of the displayed scene), as in a natural environment, the performance of the visual system is enhanced; however, displays lacking these cues cause significant viewer fatigue, due to a conflict with other cues [86]. Since retinal blur is preserved by most displays that support accommodation, we concentrate on accommodation in the majority of this paper while also discussing retinal blur in Section 6.1.1.

With the exception of ultra-high resolution displays, such as holograms, small volumetric displays, and multi-focal devices requiring specialized eye-worn equipment, no existing 3D display simultaneously supports correct accommodation, binocular disparity, and motion parallax over a wide field of view. We propose a new computational display design, dubbed Focus 3D, that has the potential to synthesize light fields with sufficient angular resolution to allow near correct viewer accommodation and retinal blur in addition to smooth motion parallax and binocular disparity. The key innovation is a combination of display optics and compressive light field synthesis through nonnegative light field tensor factorization. Following the approach in Section 3.4, we extend multilayer display architectures with directional backlighting; however, instead of synthesizing a low angular resolution light field with a predefined field of view, we introduce high angular resolution (HAR) backlighting that allows high-resolution *view cones* to be steered into an observer's eyes. Due to the novel architecture, each view cone has a significantly larger depth of field than previously proposed solutions, offering the potential for the visual system to focus the eyes. We demonstrate the viability of this design through the construction of a prototype display that allows a

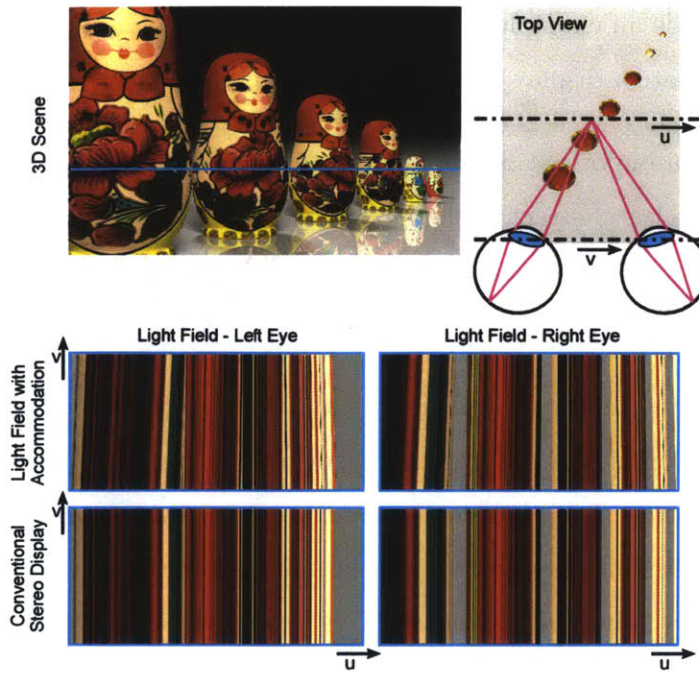


Figure 6-2: Natural light fields (*top*) observed by the human visual system (HVS) exhibit stereoscopic disparity (i.e., different left and right eye perspectives) and parallax within the area of each pupil (*center*). The subtle, but important, variation of the light field over the pupils allows the HVS to accommodate on different depths within the scene. Conventional 3D displays do not provide enough angular resolution to support this important depth cue (*bottom*).

camera to focus at multiple depths about the display (see Figure 6-1).

We explore a computational approach to synthesizing light fields as a set of narrow, but ultra high-resolution view cones that are steered only where required: into the viewer’s eyes.

Benefits We describe a new optical display architecture, consisting of stacked display layers and a high resolution directional backlight, that provides a significantly increased depth of field over a small set of view cones steered into the eyes of the viewer, offering the potential to provide binocular disparity, motion parallax, and near correct accommodation. As a compressive light field display, Focus 3D increases the display brightness and field of view while reducing the required number of time-multiplexed frames as compared to conventional displays. Previous displays providing correct accommodation cues require either additional eyewear or a significantly higher optical and computational complexity. To the authors’ best knowledge, Focus 3D is the first practical display that has the potential

to support near correct accommodative depth cues while allowing the viewer to move around the device from a wide range of viewpoints – including multiple distances from the screen.

Limitations As with other multilayer displays, stacking multiple display elements increases moiré and color-channel crosstalk, decreases the overall display brightness, and presents an alignment challenge. Obtaining good performance in the proposed multilayer framework also requires display panels which exceed currently available refresh rates, although upcoming display technologies have been demonstrated with much higher rates [69]. While our current prototype is about 50 cm thick, future generations of the proposed display may benefit from optical folding techniques such as wedge optics [189]. We employ an efficient GPU-based implementation of nonnegative tensor factorization to compute content-adaptive light field decompositions. While this approach adds to the computational complexity of the system, no heuristics are known to drive multilayer displays with the proposed type of directional backlighting.

Our prototype display is suitable for testing with a camera; several engineering enhancements would allow the display to be tested with human viewers. Constructing a display with sufficient angular resolution to support multiple depths of focus over a human-sized pupil diameter requires high quality optics. Although we provide simulations with such an aperture, our prototype display is limited to focus over a larger 2 cm camera aperture due to the performance of the inexpensive integrated Fresnel lens, which exhibited poor focus, especially off-axis. Our approach also requires high speed eye tracking; although in this paper we assume the eye positions of the observers are known, we note high speed (≥ 500 Hz) encumbrance-free commercial trackers are available from such vendors as SensoMotoric Instruments. Finally, the brightness of the display’s backlight must be improved to permit human viewing.

6.1.1 Focus 3D Architecture

The goal of the Focus 3D architecture is to efficiently provide accommodation, stereo, and motion parallax by steering a set of narrow high resolution light cones directly into

the viewer's eyes. Our approach is a hybrid of the view sequential Cambridge display design [187] and the multilayer display architecture of Tensor Displays (Section 3.4).

In one variation of the Cambridge display design, an LCD layer is placed against a lens and illuminated by a backlight (refer to Figure 6-6). If the backlight and viewer are placed at conjugate distances with respect to the lens, a point light source from the backlight will illuminate the LCD layer and rays will subsequently converge to a point at the viewing plane. Thus an image displayed on the LCD layer will be visible only to an observer in the viewing position corresponding to the illuminated region of the backlight. To create a time-multiplexed multiview display, a set of views are displayed in rapid sequence on the LCD layer, each while the corresponding region of the backlight is illuminated. We observe that it is straightforward to extend this design to support accommodation by incorporating a *high angular resolution (HAR) backlight*; with sufficient backlight resolution, multiple viewpoints can be created within the area of the pupil, providing the focus cues to the eyes. However, such a design would require display rates that far exceed currently technology; for example, a set of 5×5 views over each eye with a 60 Hz refresh rate would require a 3000 Hz display. The result would also be very dim, as each of the M views would be illuminated only a fraction $1/M$ of the time.

Exploiting the correlation within a large set of 4D light field views using the compressive Tensor Display Framework developed in Section 3.4 enables eye accommodation with brighter imagery using the refresh rates of current and upcoming displays. In this embodiment, a compressed set of correlated view patterns is displayed in sequence on the LCD layer, each while multiple regions of the backlight (and thus the eyes) are illuminated simultaneously. Furthermore, we can replace the single LCD layer in front of the lens with an N layer stack of LCDs, increasing the spatial and angular resolution of the display as well as compression performance.

In the remainder of this section, we describe the details of this approach and analyze performance and limitations. Section 6.1.1 establishes how to emit a light field to support correct accommodation using an N -layer, M -frame multilayer display illuminated by a high angular resolution (HAR) backlight. We show that such a display can be optimized

using the Tensor Display Framework, albeit with a modified backlight illumination model. Section 6.1.1 assesses the structure of the backlight illumination and layer patterns produced by the decomposition; this analysis reveals the source of enhanced brightness achieved with Focus 3D over prior methods utilizing direct time-multiplexed backlight illumination schemes. Section 6.1.1 derives upper bounds on the accommodation range for both existing display architectures and Focus 3D. Section 6.1.1 examines how the design is affected by diffraction, and Section 6.1.1 concludes by showing the influence of diffraction and light field compression on retinal blur quality.

Displays with HAR Backlighting

As described above and shown in Figs. 6-3 and 6-6, Focus 3D consists of an N -layer stack of light-attenuating panels illuminated by a *high angular resolution (HAR) backlight* capable of synthesizing multiple uniform light sources that converge along a closely-spaced set of points spanning the viewer’s pupils. Similar to Travis [187], such a backlight can be fashioned by placing a large lens (e.g., a Fresnel lens or folded waveguide) against the rear layer. If another display is placed at a distance d_b behind the lens, then a virtual layer will be created at a distance $d_v = (f d_b)/(d_b - f)$ in front of the lens. A HAR backlight is obtained when d_b is selected such that d_v equals the distance d_e from the lens to the viewer’s pupil.

Representing Emitted Light Fields As shown in Figure 6-3, we propose Focus 3D as a generalization of prior displays capable of supporting near correct accommodation through high angular resolution backlighting. Rather than using a single layer placed directly in front of the lens, we propose placing a stack of light-attenuating layers. For greater generality, we further assume that these layers support a higher refresh rate than the human eye, such that the viewer perceives the time average of an M -frame sequence. This follows the N -layer, M -frame display archetype developed in Section 3.4. As shown there the emitted light field $\tilde{l}(x, v)$ can be modeled following Equation 3.52.

The Tensor Display Framework considers two cases: uniform backlighting, such that $b_m(x, v) = 1$, and directional backlighting, such that $b_m(x, v)$ is a low-resolution light field produced

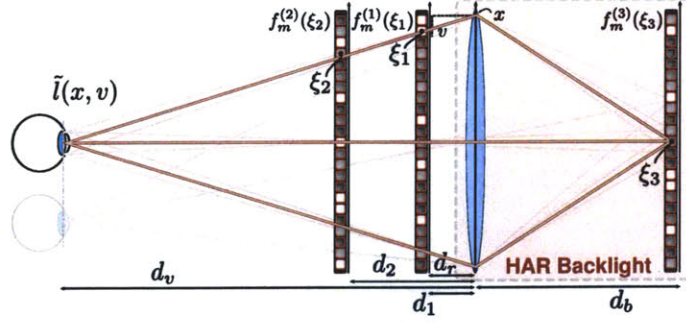


Figure 6-3: Focus 3D coordinate system. An N -layer stack of light-attenuating panels is illuminated by a high angular resolution backlight. Each pixel on the backlight layer illuminates a small region of the viewer’s pupil. We show a specific embodiment of a HAR backlight, comprising a large lens and a backlight display separated a distance d_b behind the lens, following the design of Travis [187]. A generalized system is shown in Figure 6-6.

by an auxiliary system (e.g., a lenticular display). We observe that Equation 3.52 can be modified to support high angular resolution backlighting, as depicted in Figure 6-3, such that

$$\tilde{l}(x, v) = \frac{1}{M} \sum_{m=1}^M f_m^{(N+1)}(\phi(x, v)) \prod_{n=1}^N f_m^{(n)}(x + (d_n/d_r)v), \quad (6.1)$$

where $\phi(x, v)$ defines the point of intersection ξ_{N+1} of ray (x, v) with the backlight layer, $f_m^{(N+1)}(\xi_{N+1})$ denotes the emitted irradiance of the backlight layer during frame m , and $\{f_m^{(n)}(\xi_n)\}$, for $n \in [1, N]$, remain the transparencies of the N layers in front of the lens. We observe that the point of intersection is found by tracing the ray (x, v) backwards through the lens, with focal length f , and propagating a distance d_b to the backlight layer. Using ray transfer matrix analysis [73] with paraxial ray and thin lens approximations, these operations are given by:

$$\begin{pmatrix} \phi(x, v) \\ -\eta/d_r \end{pmatrix} = \begin{pmatrix} 1 & d_b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix} \begin{pmatrix} x \\ -v/d_r \end{pmatrix}, \quad (6.2)$$

where η/d_r denotes the slope of the ray leaving the backlight layer. Thus, $\phi(x, v)$ is given



Figure 6-4: Performance of computational displays vs. display complexity. We simulate the ability to refocus the light field emitted from various displays, with and without HAR backlighting, following the design proposed in Section 6.1.1. Up to three layers were placed in front of a Fresnel lens, with focal length $f = 30$ cm, each separated by 0.5 cm. We decompose the target light field to emit 5×5 views spanning each viewer pupil, separated by a distance $d_e = 100$ cm from the lens. *Left:* The light field corresponding to a dragon model is provided as input to the decomposition algorithm. *Right:* The first four columns show the received images for the left and right eye, when focused in front of and behind the lens. The remaining two columns show inset regions centered on the dragon’s eye. Five system architectures are compared from top to bottom, with varying numbers of layers and frames. The first three rows evaluate Tensor Display designs using a uniform backlight ($b_m(x, v) = 1$). The last two rows illustrate the benefits of HAR backlighting, demonstrating that its inclusion enables clear focus cues; note that the dragon’s eye can be brought into sharp focus, in contrast to cases without HAR backlighting. Quantitative assessment of focus is provided by the peak signal-to-noise ratio (PSNR) with reference to the original refocused light field, confirming that increasing layers and frames reduces artifacts.

by the following expression.

$$\phi(x, v) = \left(1 - \frac{d_b}{f}\right) x - \frac{d_b}{d_r} v \quad (6.3)$$

Decomposing Light Fields Using Weighted NTF Following the Section 3.4, the light field emitted by a N layer display can be decomposed into a set of M time-multiplexed layer patterns using *nonnegative tensor factorization (NTF)*. Substituting Equation 6.3 into Equation 6.1 provides a closed-form expression for the light field emitted by such a display, $\tilde{l}(x, v)$, in terms of the time-multiplexed layer patterns, $\{f_m^{(n)}(\xi_n)\}$. In practice, the decomposition of a target light field, $l(x, v)$, into the layer patterns requires solving the

following nonlinear least squares problem:

$$\arg \min_{\{f_m^{(n)}(\xi_n)\}} \int_{\mathcal{V}} \int_{\mathcal{X}} \left(l(x, v) - \tilde{l}(x, v) \right)^2 dx dv, \text{ for } 0 \leq f_m^{(n)}(\xi_n) \leq 1 \quad (6.4)$$

The solution to this optimization problem follows from the Tensor Display Framework detailed in Section 3.4.

Figure 6-4 evaluates the performance of the weighted NTF decomposition for varying display architectures. From these simulations, we conclude that the addition of high angular resolution (HAR) backlighting to the prior Tensor Display Framework is a viable approach to eliminate accommodation-convergence conflicts using current generation and upcoming display technologies. Assuming the viewer’s position is known, such a design has the potential to deliver all five “missing” perceptual depth cues for a single user: binocular disparity, convergence, accommodation, retinal blur, and motion parallax.

Focus 3D Decompositions

While Figure 6-4 confirms that the Focus 3D design can successfully synthesize accommodation cues with a sufficient number of layers and frames in simulation, it does not provide intuition into the decomposed patterns. In this section we briefly examine decomposed layer and backlight illumination patterns to understand the expected benefits of our decomposition algorithm over prior direct time-multiplexed backlight illumination schemes. As shown at the top of Figure 6-5, direct time-multiplexing requires a single layer placed in contact with the lens and a secondary layer placed behind the lens, conjugate to the viewer’s pupil. In this mode of operation, each pixel on the backlight that maps to a region of the pupil is sequentially illuminated; simultaneously, the front layer displays the perspective corresponding to a center of projection located in the center of the pupil region. As shown in the refocused images, the depicted light field preserves accommodation cues, but suffers from severe attenuation since each backlight pixel only illuminates the eye for a brief period.

As shown at the bottom of Figure 6-5, the Tensor decomposition algorithm used with

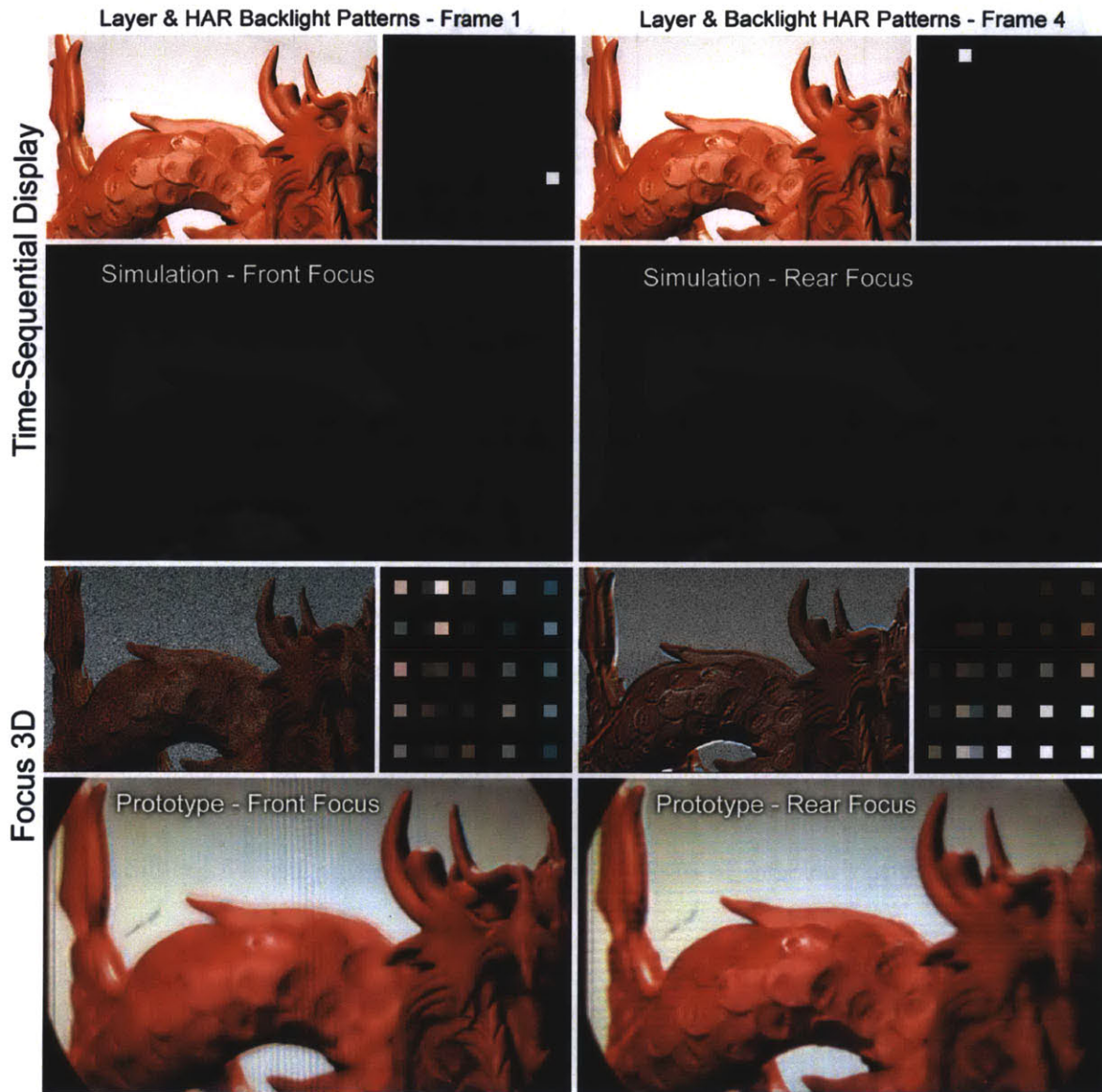


Figure 6-5: Comparing direct vs. compressive display modes. A single-layer Focus 3D prototype is considered, comprising one layer in front of a large lens and a backlight conjugate to the viewer's pupil. These examples evaluate a light field with 5×5 views spanning a single pupil located along the optical axis. *Top:* Using direct time-multiplexing, following the approach of Travis [187], only a single backlight pixel is active in each frame, resulting in a dim image. *Bottom:* Focus 3D exploits correlations between views to illuminate each pupil region for a longer duration, increasing image brightness, as shown in photographs of prototype.

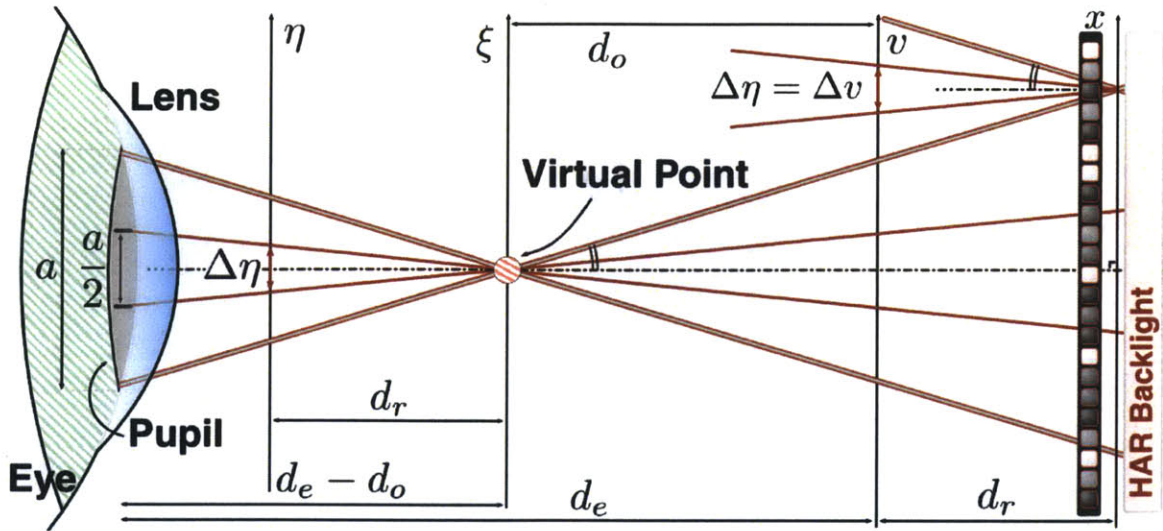


Figure 6-6: A virtual point source is created, with accommodation cues. A geometric argument is made for the angular resolution requirements for accommodation in Section 6.1.1. Here, we observe that the angular sampling frequency at the eye, $\Delta\eta$, equals Δv , the angular sampling frequency of the HAR backlight, enforcing a lower bound such that at least two rays enter the pupil of the eye.

Focus 3D exploits correlations between views to enable each backlight pixel to illuminate the pupil for a longer duration, yielding a brighter image. As is the case when driving a display using the Tensor Display Framework (Section 3.4) reconstruction artifacts result from the compression process. In summary, Focus 3D extends the new design trade-space between brightness, reconstruction fidelity, and effective frame rate to one that may enable near-term display technologies to resolve the accommodation-convergence conflict.

Upper Bound on Accommodation Range

In this section we formally assess the benefits of high angular resolution (HAR) backlighting for extending the range over which accommodation cues can be achieved. We adapt the prior frequency-domain analysis of light field displays developed by Zwicker et al. [219] and Wetzstein et al. [205, 206], and Section 3.4.2. While these works derive an upper bound on the depth of field, we perform a similar analysis to reveal an upper bound on the accommodation range for multilayer displays, including those with HAR backlighting.

Accommodation Threshold Consider the arrangement depicted in Figure 6-6, in which a virtual point light source is located a distance $d_e - d_o$ in front of the viewer's pupil, where d_e and d_o are the distance from the eye to the display and from the virtual point to the display, respectively. Following Takaki et al. [180, 181], we assume that a minimum of two rays must enter the pupil from this point to support correct accommodation. Let each ray (ξ, η) passing through the virtual point be defined using a two-plane parameterization, where the ξ -axis is coincident with the point and the η -axis is located a distance d_r in front. Under this parameterization, the maximum angular sampling rate $\Delta\eta_{\max}(d_o)$ supporting accommodation is:

$$\frac{\Delta\eta_{\max}(d_o)}{d_r} = \frac{a}{2(d_e - d_o)}. \quad (6.5)$$

As proven below, the angular sampling rate for a light field display is invariant to the depth of a virtual point. In other words, the maximum angular sampling rate $\Delta v_{\max}(d_o)$ equals $\Delta\eta_{\max}(d_o)$, as defined in the two-plane parameterization of the emitted light field (see Figure 6-3). As a result, the angular sampling rate required for accommodation (ω_v) must satisfy the following expression:

$$\omega_v(d_o) \geq \frac{1}{2\Delta v_{\max}(d_o)} = \frac{d_e - d_o}{d_r a} \quad (6.6)$$

As shown at the bottom of Figure 6-7, the supported accommodation range for a given light field display can be estimated by determining the point of intersection of the maximum angular frequency, $\omega_{v_{\max}}$, supported by the display architecture with the *accommodation threshold* given by Equation 6.6. Points closer to the eye than this point of intersection (i.e., $0 < d_e - d_o \leq d_r a \omega_{v_{\max}}$) will emit a minimum of two rays into the viewer's pupil, whereas points further away will not.

Maximum Angular Frequency for a Multilayer Display To estimate the accommodation range, the maximum angular frequency $\omega_{v_{\max}}$ is required for a given light field display. A direct analysis for conventional architectures, including parallax barriers and integral imaging displays, is possible. Yet, for multilayer displays, it is not clear how to estimate the maximum angular frequency. We propose an upper bound on the maximum

angular frequency, based on frequency-domain analyses previously applied to characterize the depth of field of such displays.

Equation 6.1 can be transformed into the following simplified form:

$$\tilde{l}(x, v) = \frac{1}{M} \sum_{m=1}^M \left[\prod_{n=1}^{N+1} f_m^{(n)}(x + (d_n/d_r)v) \right], \quad (6.7)$$

where $f_m^{(N+1)}(\xi_{N+1})$ now denotes the effective transparency of the virtual layer corresponding to the image of the backlight layer formed by the lens. In this interpretation, the virtual layer is located a distance $d_{N+1} = d_v = (fd_b)/(d_b - f)$ in front of the lens. Taking the two-dimensional Fourier transform of this expression yields an estimate of the emitted light field spectrum in terms of angular frequency ω_v and spatial frequency ω_x for a display with HAR backlighting:

$$\hat{l}(\omega_x, \omega_v) = \frac{1}{M} \sum_{m=1}^M \left[\overset{\frown}{n=1} N + 1 \Sigma \hat{f}_m^{(n)}(\omega_x) \delta(\omega_v - (d_n/d_r)\omega_x) \right], \quad (6.8)$$

where $*$ denotes convolution and the repeated convolution operator is defined such that

$$\overset{\frown}{n=1} N + 1 \Sigma \hat{f}_m^{(n)}(\omega_x, \omega_v) \equiv \hat{f}_m^{(1)}(\omega_x, \omega_v) * \dots * \hat{f}_m^{(N+1)}(\omega_x, \omega_v). \quad (6.9)$$

Following the procedure outlined by Zwicker et al. [219], Wetzstein et al. [205, 206], and Section 3.4.2 the spatio-angular bandwidth of a multilayer display is determined by the region of non-zero support in the emitted light field spectrum $\hat{l}(\omega_x, \omega_v)$. Intersecting the line $\omega_v = (d_o/d_r)\omega_x$ with the spectral support provides a geometric construction for the upper bound on the depth of field. Figure 6-7 compares the upper bound on the depth for field for two competing display architectures: a two-layer display with uniform backlighting and a single-layer display with HAR backlighting. However, this upper bound does not account for limitations of our decomposition algorithm; in practice, the number of ray constraints (i.e. non-zero values of tensor \mathcal{W} in Equation 3.44) and the compressibility of the input light field determine actual performance. Section 6.1.3 provides a performance evaluation in simulation and on a prototype device.

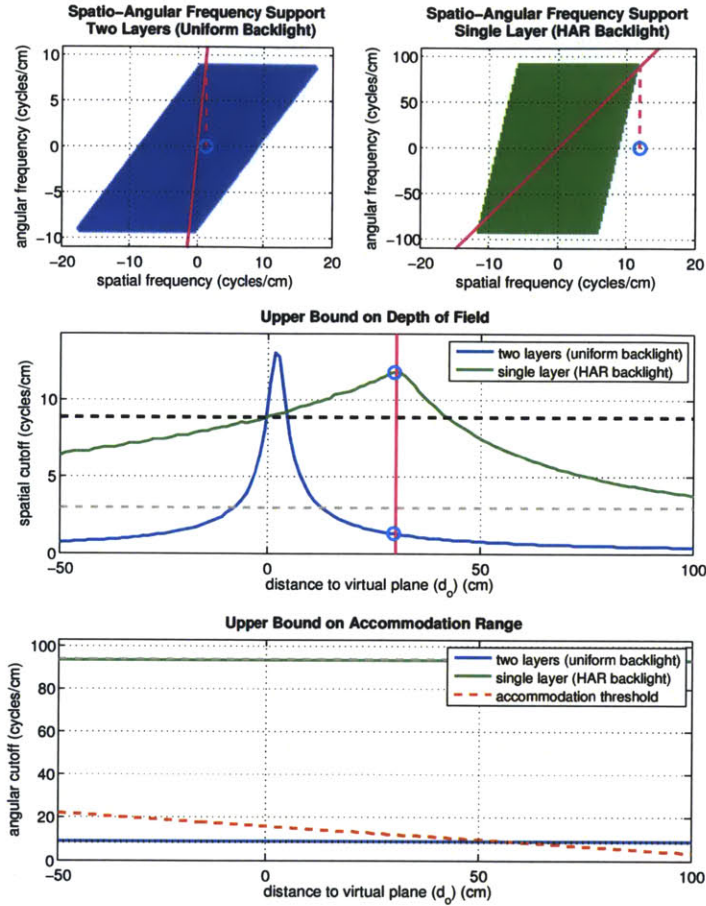


Figure 6-7: HAR backlighting is required to support accommodation within the depth of field of a multilayer display. We compare a two layer display with uniform backlighting and a single layer display with HAR backlighting. *Top:* The spatio-angular bandwidths, evaluated following Section 6.1.1. Note that HAR backlighting significantly increases the maximum angular frequency. *Middle:* Upper bounds of depth of field. The dashed black line denotes the maximum spatial frequency corresponding to the physical pixel pitch. The dashed gray line denotes the maximum spatial frequency supported by the virtual panel, given by the magnified image of the backlight layer. The magenta lines illustrate the relationship between the spatio-angular bandwidth and depth of field plots for a reference plane at $d_o = 25$ cm. *Bottom:* Accommodation is supported for virtual plane distances d_o where the display’s angular cutoff frequency (blue and green lines) is above the accommodation threshold (Equation 6.6, red dotted line). Note that without HAR backlighting a two-layer display only supports accommodation when the virtual layer is separated by $d_o \gtrsim 65$ cm from the display (well outside the depth of field). With HAR backlighting, accommodation is predicted throughout the depth of field, as reflected in experiments. The plots reflect our prototype testing configuration: a pupil diameter $a = 2.0$ cm, an eye to display distance $d_e = 127$ cm, and an $f = 31.8$ cm focal length lens. The two layer display used a layer separation of 4.0 cm.

A similar upper bound on the maximum angular frequency $\omega_{v\max}$ can be derived by analyzing the spatio-angular bandwidth of a given multilayer display. Depth-of-field analysis is facilitated by considering the frequency-domain properties of a Lambertian surface located a distance d_o in front of the display. For such a surface, the emitted light field, $\tilde{l}(x, v)$, equals $f(x + (d_o/d_r)v)$, corresponding to the line $\omega_v = (d_o/d_r)\omega_x$ in the frequency domain (see Figure 6-7, top and center row). Similarly, a uniform directional area source emits a light field $l(x, v)$ such that

$$\tilde{l}(x, v) = f(v). \quad (6.10)$$

Taking the two-dimensional Fourier transform of this expression yields an estimate for the corresponding light field spectrum:

$$\hat{l}(\omega_x, \omega_v) = \hat{f}(\omega_v) \delta(\omega_x), \quad (6.11)$$

where $\delta(\xi)$ is the Dirac delta function. Thus, the spectrum of a directional source located *any* distance d_o from a light field display is approximated by a vertical line in the emitted light field spectrum. As a result, the maximum angular frequency $\omega_{v\max}$ supported by any light field display is provided by the intersection of the spatio-angular bandwidth with a vertical line, evaluated along the ω_v -axis.

The above demonstrates a connection linking depth-of-field analysis to bounds on the accommodation range of a light field display. As shown in Figure 6-7, the accommodation range is found by intersecting the maximum angular frequency $\omega_{v\max}$ with the accommodation threshold given by Equation 6.6. In this example, we find that HAR backlighting is necessary to support accommodation within the depth of field centered near the display surface.

Diffraction

Light passing through an aperture spreads out angularly (diffracts) to a degree inversely related to the aperture size. For a multiview display, this relationship enforces a limit on the maximum angular resolution that can be achieved for a given spatial resolution; for a

given display pixel aperture size, views can be spaced no more closely than the corresponding angular spread of diffraction without overlapping. For multiview displays supporting correct accommodation, diffraction is an important consideration as ultra-high angular resolution is required.

A more thorough analysis of diffraction for the Focus 3D display is presented in Maimone et al. [134]. In summary Diffraction causes light to spread out to form an Airy disk. Adjacent views will not overlap due to diffraction at viewing distance d_e if the diameter of the central element of the Airy disk is less than or equal to the view spacing over the pupil, i.e.:

$$2d_e \tan \theta_d \leq \frac{a}{n-1}, \quad (6.12)$$

where a is the pupil diameter and n is the number of views spaced over the pupil. If the diameter of the central element of the Airy disk exceeds this value, adjacent views will begin to overlap and degrade. By the Rayleigh criterion, two point-light sources are considered “just resolved” when the central element of the Airy disk of one source coincides with the minimum of the other. By this definition, when the diameters of the Airy disk center elements exceed $4d_e \tan \theta_d$, the maximum of the disk corresponding to each view will extend beyond the first minimum of the neighboring views, and adjacent views are no longer resolvable.

Figure 6-8 shows the diffraction-limited spatial and angular resolution configuration space for multiview displays that support multiple focal depths. The analysis assumes a human-sized pupil diameter, $a = 5$ mm, and optimal viewing distance of our prototype display, $d_e = 127$ cm. The figure shows that reasonable configurations (spatial resolution of 20-30 cycles/degree, angular resolution of 2-3 views over pupil) are attainable, but lie close to the diffraction limits.

Section 6.1.1 provides simulations to show how diffraction affects the focus quality of a light field.

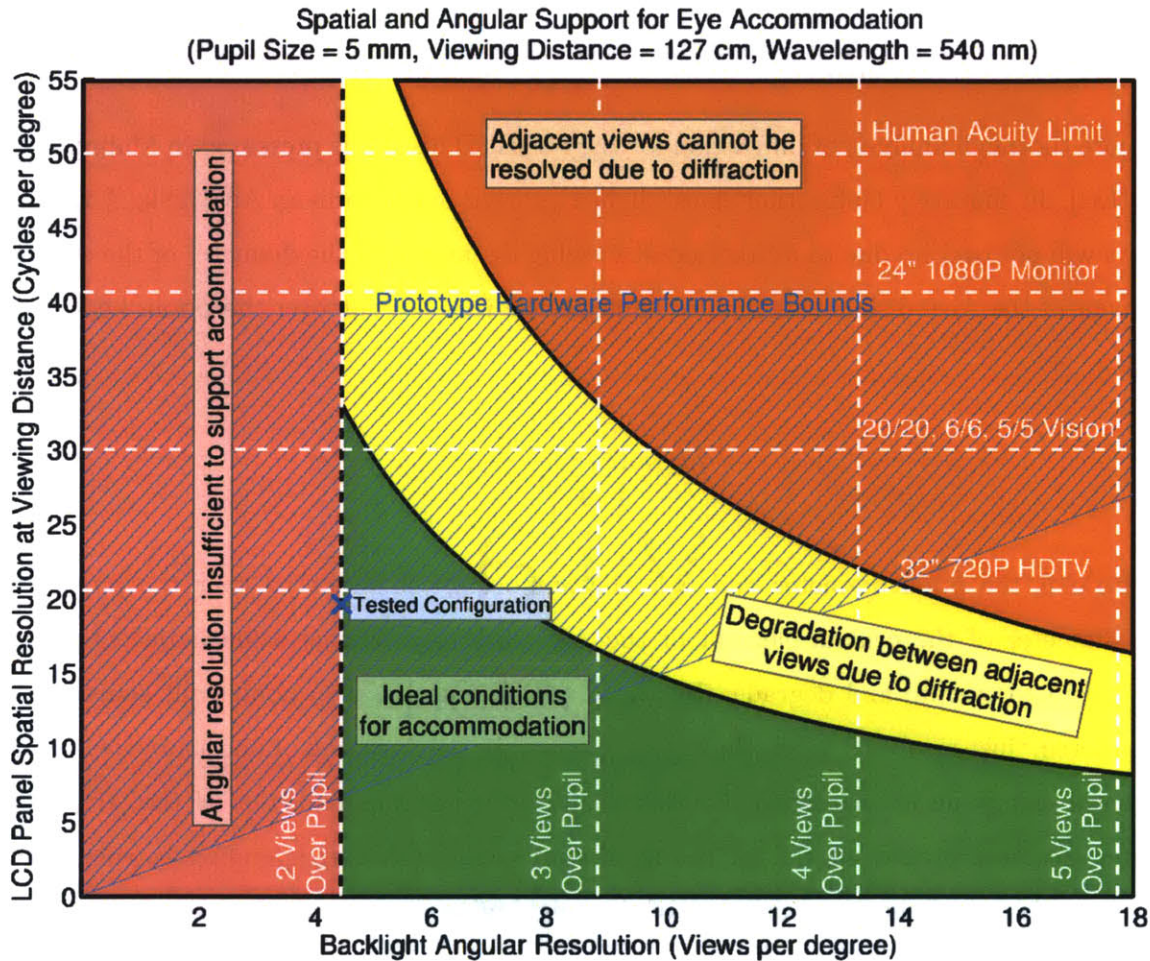


Figure 6-8: Diffraction limits on spatial and angular resolution. *Red area:* configurations with insufficient angular resolution to support correct eye accommodation (i.e. fewer than 2 views over the pupil). *Shaded blue area:* theoretical performance bounds of our prototype display (described in Section 6.1.2) with regards to the LCD panel resolution and backlight angular resolution. Plot assumes that the backlight LCD panel resolution matches the resolution of the front LCD layers; hence backlight angular resolution increases as LCD panel resolution increases. The tested configuration (see Section 6.1.3) is marked with a blue cross. *Green area:* configurations that support correct eye accommodation and have no overlap between views due to diffraction. *Yellow area:* configurations that support correct eye accommodation, but diffraction causes some crosstalk between adjacent views. *Orange area:* configurations that may support eye accommodation, but diffraction is so severe that adjacent views can no longer be resolved. Diffraction is approximated as in Section 6.1.1 for pupil size $a = 5$ mm, viewing distance $d_e = 127$ cm, and wavelength $\lambda = 540$ nm.

Retinal Blur

Along with the ability to focus at different depths about a display, it is also important that the blur of out-of-focus imagery is accurate; retinal blur has been shown to help the human visual system solve the binocular correspondence problem and interpret monocular occlusions [84]. Here again, we summarize a more thorough analysis presented in Maimone et al. [134].

From a theoretical standpoint, the quality of retinal blur in our proposed display design is influenced by two primary factors: the light field compression performance of the tensor factorization algorithm and diffraction. (In practice, the blur quality will also be affected by the performance of the optical components).

Figure 6-9 provides a comparison of retinal blur between a ground truth light field and light fields compressed through tensor factorization while simulating diffraction. Note that most of the test cases fall above the diffraction limits shown in Figure 6-8 in order to provide an estimate of maximum performance in a diffraction-limited system. We make the following observations from the results:

1. In the nominal compressed case, the average PSNR is 31 dB for the in-focus images and 37 dB for the out-of-focus images. It is clear that diffraction and compression limit the performance of our approach, but performance on the order of lossy video compression (≥ 30 dB) can still be achieved.
2. As expected, the in-focus performance decreases as the number of views and time-multiplexed frames are reduced. With too many constraints for the available degrees of freedom (e.g. 5×5 views, 2 frames), focus performance is poor.
3. High PSNR is not indicative of qualitative blur performance. The most numerically accurate out-of-focus blur occurred in the 2×2 view case, in which the radius appears most accurate. However, the blur accuracy is low as compared to the nominal 5×5 view case – two distinct out-of-focus images can be seen. This issue can be resolved in future work by employing error metrics inspired by the human visual system.



Figure 6-9: Simulated retinal blur and diffraction. Images show closeups of close and far matryoshka dolls from the light field shown in Figure 6-2, and inset images show further magnification. The larger doll is virtually positioned at 17 cm *in front of* the display and the rear doll at 18 cm *into* the display. The views of the light field are evenly spaced over a pupil of $a = 5$ mm for a single eye, with the outermost views centered at the pupil edges. Compressed images reflect the configuration of our prototype display: 1 LCD layer in front of a HARB, a $f = 31.8$ cm focal length lens, a viewing distance of $d_e = 127$ cm, and native panel resolution of 39.1 cycles/degree at this distance. Diffraction is approximated using the method described in Section 6.1.1 using the wavelengths $\lambda_{red} = 700$ nm, $\lambda_{green} = 546$ nm and $\lambda_{blue} = 435$ nm. *Rows:* Synthetically refocused images of front doll (*first rows*) and rear doll (*second rows*). *First column:* Source light field. *Following three columns:* Compressed version of the source light field using the decomposition algorithm described in Section 6.1.1 in the noted configurations. High quality retinal blur can be achieved in the presence of diffraction (*second column*), but quality suffers if compression is too high (*third column*) or angular resolution is too low (*fourth column*).

From these observations we conclude that the proposed design theoretically supports focus at multiple depths over a human sized pupil with high quality retinal blur. We also note that the most accurate blur required many views over the pupil, an approach that is only practical with a compressive framework. In Section 6.1.3, we describe the actual performance of a prototype display.

6.1.2 Implementation

Hardware

As shown in Figure 6-10, our Focus 3D prototype is constructed using off-the-shelf components – three spatial light modulating layers and a large Fresnel lens. The entire optical train is suspended from rails, enabling the placement of the lens and spatial light modulating layers at various distances from the viewer to support the experiments detailed in Section 6.1.3. The light modulating layers and backlight consist of modified Viewsonic VX2268wm

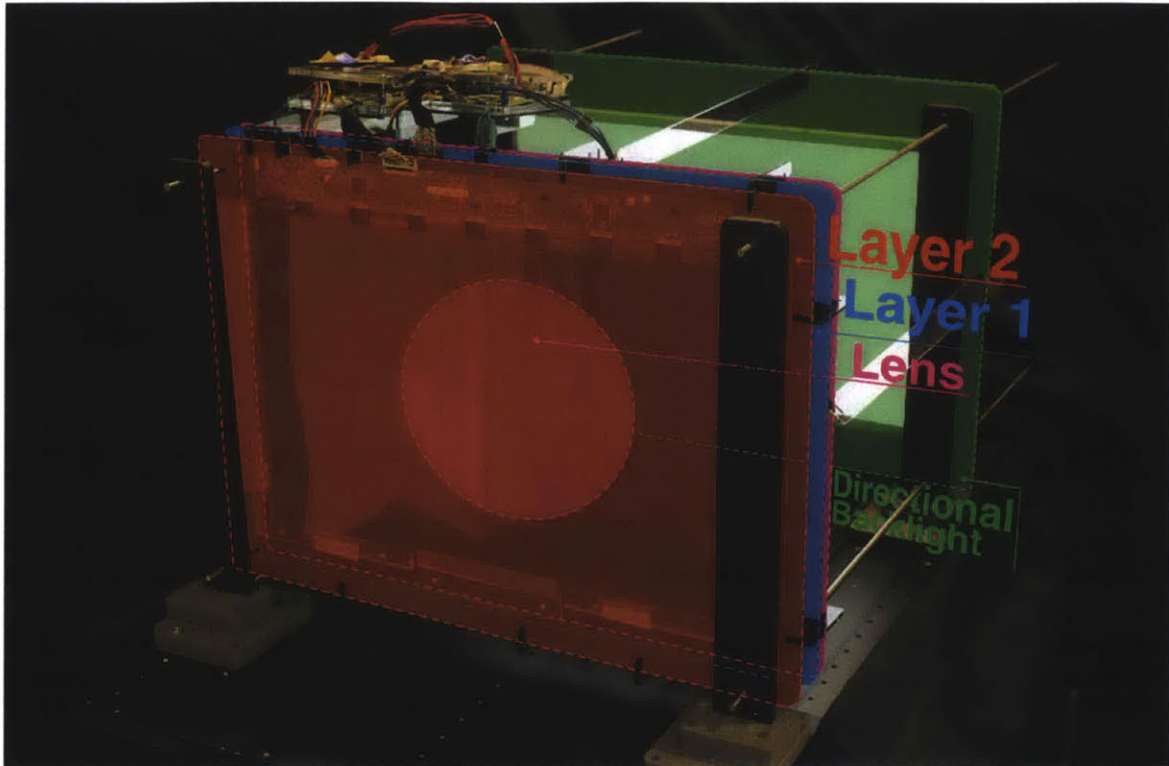


Figure 6-10: Focus 3D prototype. A stack of two transparent LCDs is mounted on rails in front of a Fresnel lens with an additional LCD monitor behind the lens. The rear monitor and the lens form a HAR backlight.

120 Hz LCD panels. The diffusing front polarizers were removed from the two front panels and replaced with clear polarizers, enabling image formation through the panels. The lens element is a Fresnel Technologies Inc. #32, 254 mm diameter Fresnel lens with $f = 318$ mm, optimized for conjugates at 424 mm and 1270 mm. We address the impact of the low optical quality of Fresnel lenses in Section 6.1.3.

Both simulation and driver software for our prototype run on an Intel Core i7 workstation with 6GB RAM and an external Nvidia QuadroPlex 7000 unit containing two Quadro 7000 GPUs and a G-Sync card. This configuration enables us to drive all three LCDs synchronously at 120 Hz over standard dual-link DVI connections.

Software

All light fields displayed on the prototype were generated by rendering multiple views of a 3D scene in OpenGL or POV-Ray. A total of 5×5 views were generated at a spatial resolution of 840×525 per eye. For stereoscopic image display, two sets of views were generated at an interocular distance of 64 mm. We note natural light fields can be captured efficiently using compressive techniques [135].

Following Section 3.4.3, we implement tensor factorization (NTF) on the GPU. This solver implements the multiplicative update rules outlined in Section 6.1.1 using OpenGL and Cg. These operations are computationally and memory efficient; the full light field matrix or tensor is never stored in memory – only the target views, 32-bit off-screen buffers for the decompositions, and intermediate buffers. The key insight allowing efficient computation is that the mathematically abstract matrix and tensor update rules applied to compressive light field synthesis directly map to hardware-accelerated operations such as perspective rendering and projective texture mapping. Solver runtimes for the above light field resolution are typically a few minutes for 100-200 iterative multiplicative updates. We note that light field rendering and factorization runtimes can be reduced by computing these stages jointly with adaptive sampling [76].

6.1.3 Assessment

We first compare Focus 3D to conventional, time-sequential displays – highlighting the increased display brightness and lower required display framerates. We then evaluate the display system with respect to the supported depth cues, optical design variations, and viewer position. All photographs of the prototype were taken as long exposures on a camera with a 2 cm lens aperture. The minimum aperture size is limited by the focal spot size of the low quality lens used in our prototype – Figure 6-9 and 6-16 provide simulations for human-sized (5 mm) pupils.

Focus 3D Architecture

Focus 3D fundamentally differs from conventional, time-sequential displays in its compressive approach to light field synthesis. As described in Section 6.1.1, the computational framework utilized in this paper allows a target light field with an arbitrary number of views to be compressed, in a numerically optimal manner, into the available display refresh rate. This approach enables both practical display architectures and brighter images. This is demonstrated in Figure 6-5 – the target light field, containing 25 views over the pupil size of a camera, is compressed into only six frames. An overall brightness gain factor of five was achieved by setting the brightness scaling factor to $\beta = 0.2$ during tensor factorization (see Equation 3.44).

Accommodation and Binocular Disparity

Near correct accommodation and binocular disparity are naturally supported by the proposed tensor framework. For this application, two light fields – each with a narrow angular baseline corresponding to one pupil – are rendered and decomposed with the mathematical framework introduced in Section 6.1.1. Figure 6-11 demonstrates the display prototype supporting both binocular disparity and multiple focal depths. The matryoshka doll images were photographed from two different positions spaced 64 mm apart and were optically focused on three different depths at each viewpoint. Note the focus/defocus effect in the closeups. This scene contains 5×5 viewpoints for each eye – 50 views total – and was successfully decomposed into 12 available frames displayed on a single LCD in front of a Fresnel lens and another LCD at the conjugate distance to the pupil plane behind the lens.

Figure 6-12 evaluates the image quality, using peak signal-to-noise ratio (PSNR) as a metric, for a varying number of time-multiplexed frames and light-attenuating layers placed within close proximity in front of the lens. Compared to the multilayer display presented in Section 3.4, the proposed work shows improved performance under the same conditions and a greater ability to scale with the number of time-multiplexed frames – even with a single layer. While the PSNR is theoretically improved for multiple stacked layers, design-

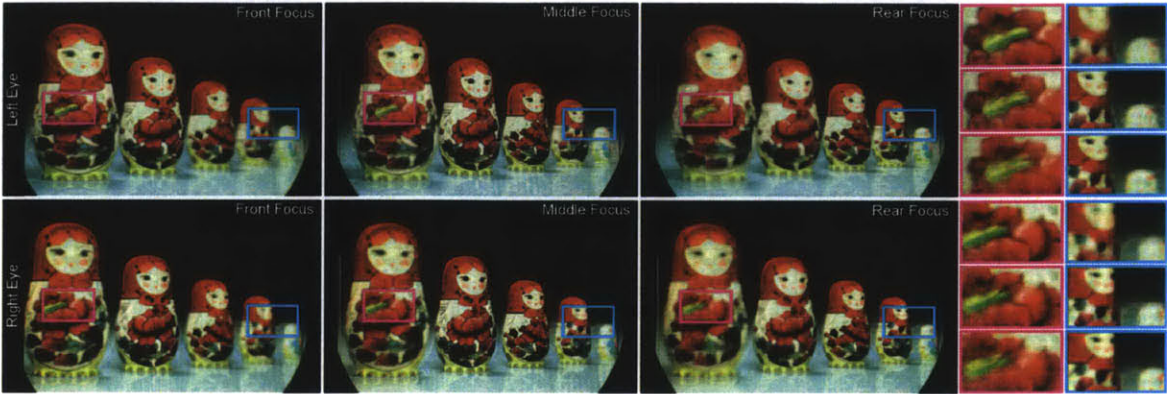


Figure 6-11: Photographs of prototype demonstrating binocular disparity (*rows*) and multiple depths of focus (*columns*). Rightmost columns show magnified inset regions. The prototype was configured with a single LCD layer placed directly in front of the lens and was photographed at a viewing distance of 127 cm.

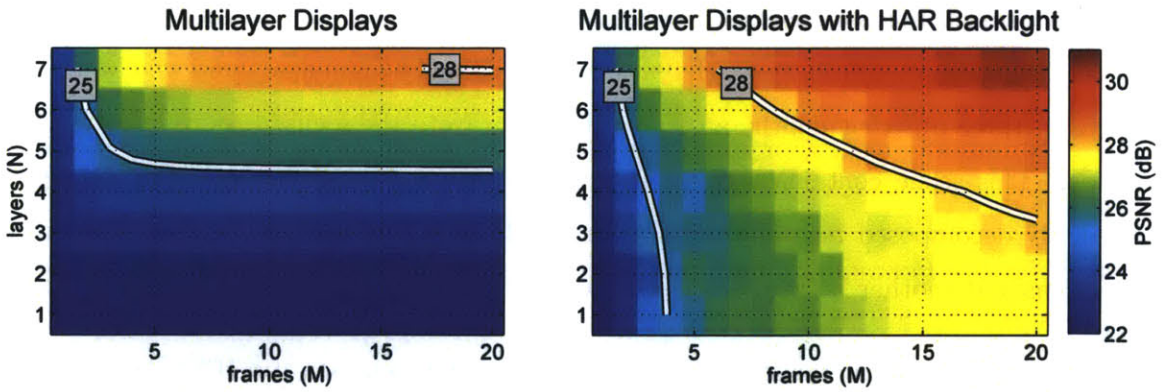


Figure 6-12: PSNR scaling with the number of attenuation layers and time-multiplexed frames for the dragon light field shown in Figure 6-4. *Left:* Multilayer displays. *Right:* Multilayer Display with HAR Backlight.



Figure 6-13: Failure case photographs for dual layer architectures. In practice imprecise alignment of prototype layers creates artifacts. Artifacts are also observed in simulation, as narrow view spacing poses a challenge for large LCD pixel sizes.

ing such systems in practice is challenging due to the necessity of precise layer alignment. Experiments with our prototype show the difficulty of achieving the necessary precision in practice (see Figure 6-13); hence, all photographs of the prototype display utilize only a single LCD and the directional backlight.

Motion Parallax and View Steering

Motion parallax and view steering are evaluated in Figure 6-14. We capture three different viewpoints, centered around the display normal, within a lateral range of 30 cm at a viewing distance of 127 cm. The display optically steers a small light cone into the direction of each view without consideration of any other view. Motion parallax is clearly visible in the three rows of Figure 6-14. Additionally, two different focal settings show, for each viewpoint, the front and rear of the shark in focus, respectively. The lateral range of supported viewpoints is practically limited by the quality of the refractive display element – the inexpensive Fresnel lens used in our prototype exhibits significant radial image distortion, coma, and dispersion for off-axis viewpoints at steeper angles. To show the theoretical performance of our system with higher quality optics, simulated results are shown in Figure 6-15.

Moving Away from the Conjugate Plane

Moving away from the conjugate plane results in an optical configuration in which the pupil plane does not correspond to the conjugate plane of the backlight. If the observer moves far

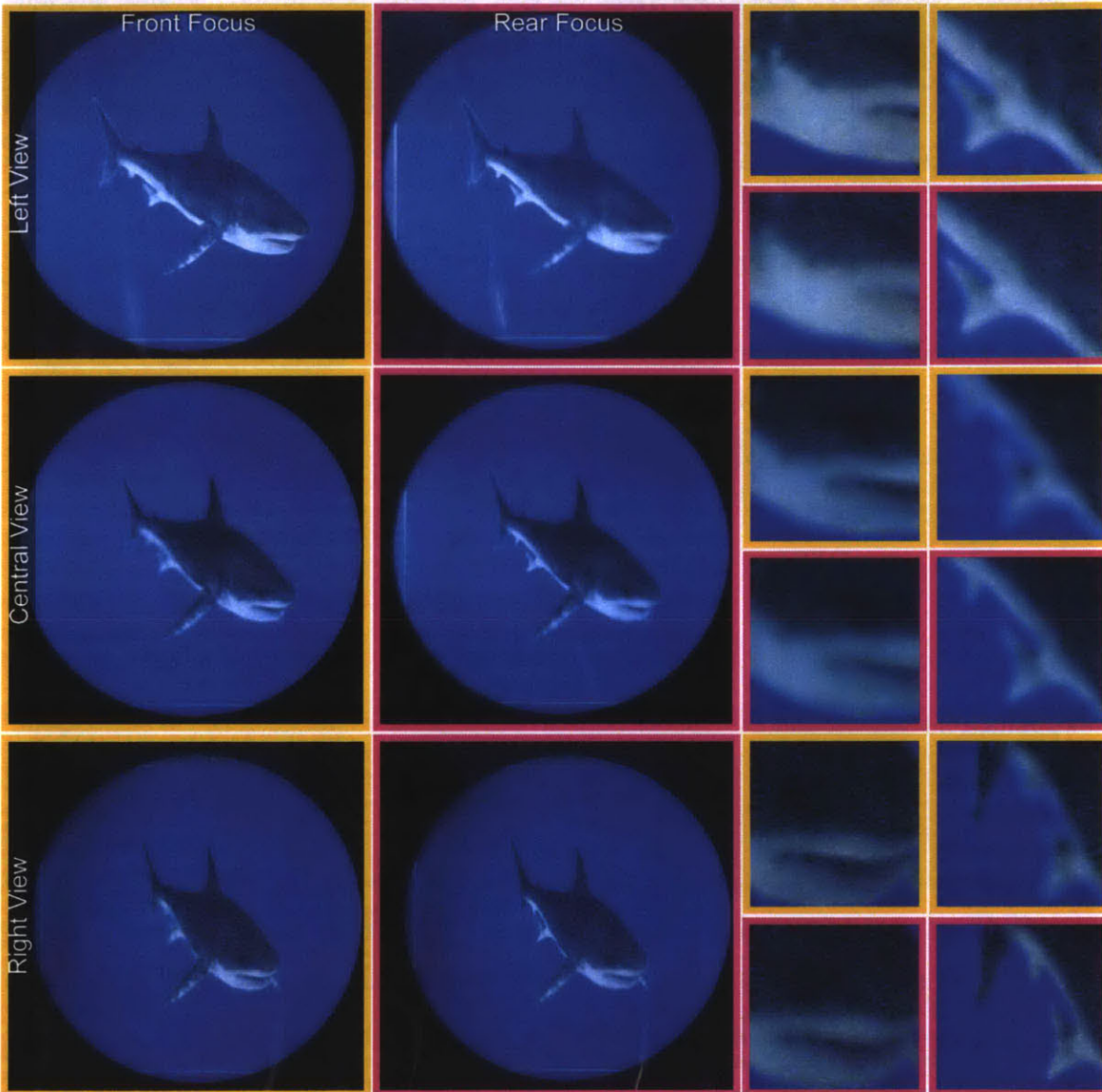


Figure 6-14: Photographs of prototype demonstrating motion parallax and multiple depths of focus. The prototype was configured with a single LCD layer placed directly in front of the lens and was photographed at a viewing distance of 127 cm. Three viewpoints, laterally shifted parallel to the display, are shown in the rows while the left and center columns show the front and rear of the shark in focus, respectively.

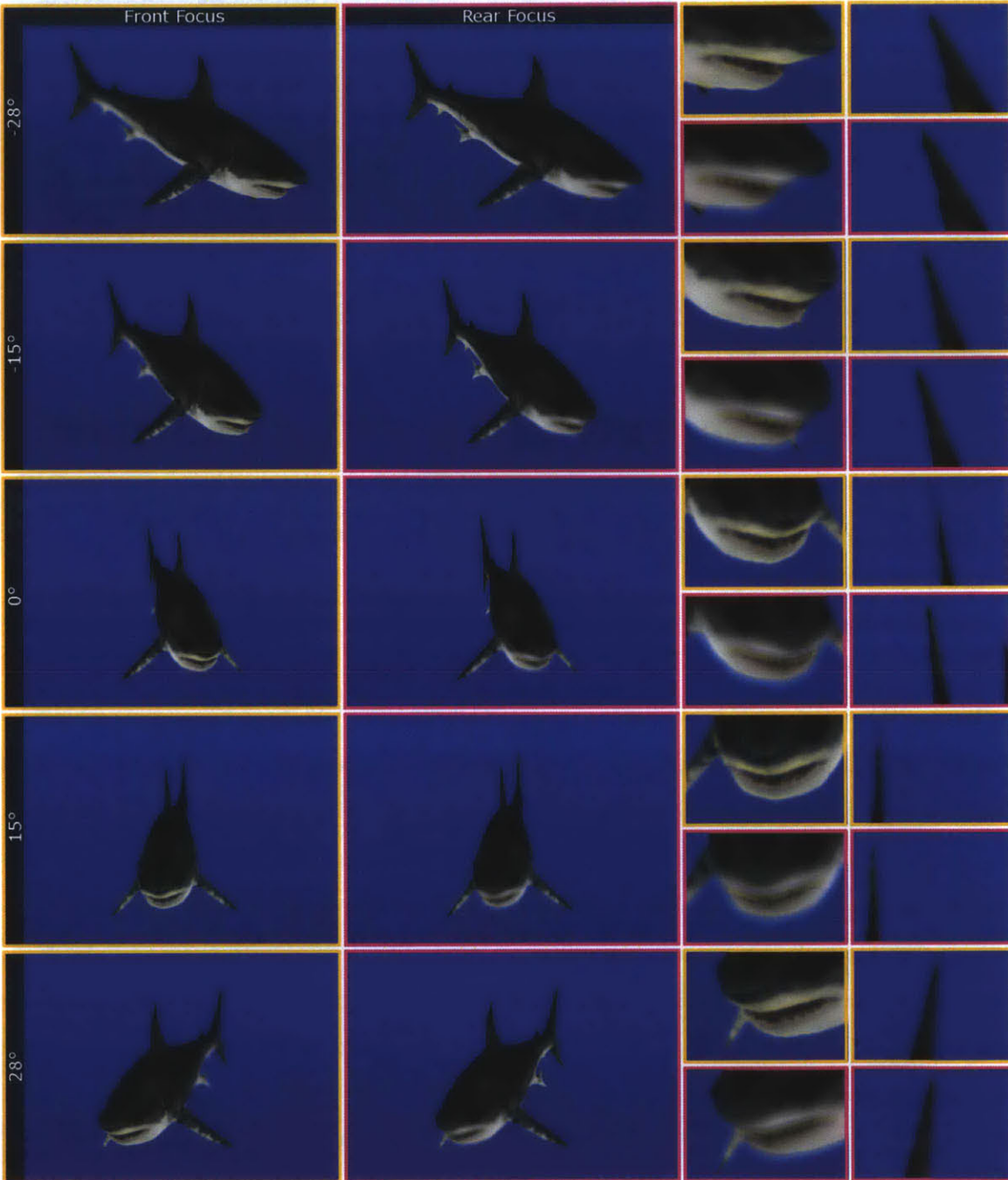


Figure 6-15: The field-of-view of the Focus 3D prototype was limited by the distortions of the inexpensive lens chosen. Simulation of a monoscopic wide field of view display with correct accommodation for a tracked user is shown here. Five viewpoints, laterally shifted parallel to the display, are shown in the rows while the left and center columns show the front and rear of the shark in focus, respectively. The simulations demonstrates focusability over a wide 56° field-of-view (136 cm laterally) for a user 127 cm from the display using six time-multiplexed frames.

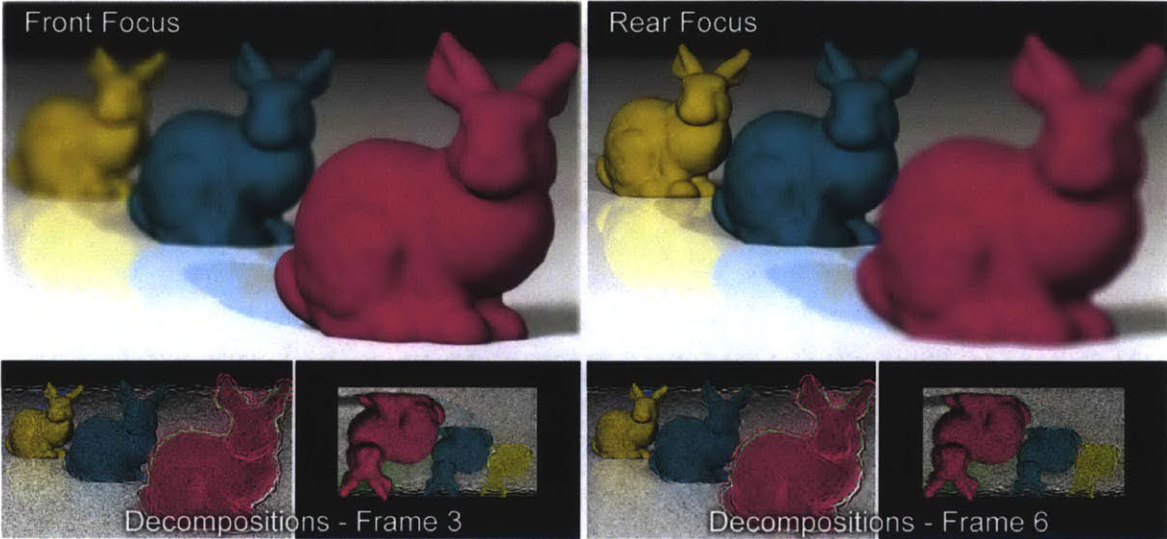


Figure 6-16: Multiple focal depths are also supported when the observer moves away from the conjugate plane of the backlight. *Top row:* Simulation shows two differently focused views. *Bottom row:* Two frames of the decomposed patterns for front layer and backlight.

enough from the display, this optical arrangement practically results in a multilayer display – the backlight is a virtual layer placed at the conjugate plane in front of the physical display enclosure. This approach is similar to that of Gotoda [65], who noted that placing a lens over an LCD in a multilayer display changes its apparent position. Figure 6-16 simulates this case for an observer at a distance of 127 cm, while the conjugate plane of the backlight is located 57 cm in front of the screen. The decompositions use six time-multiplexed frames and the target light field has 5×5 viewpoints over an eye aperture of 5 mm. As shown in the top row, multiple focal depths are still supported. The decompositions (see Figure 6-16, bottom row), however, differ from the case where the conjugate plane is in the pupil plane (see Figure 6-5) – they show a flipped version of the mask patterns that appear on the virtual layer floating in front of the other layers.

6.2 Compressive Light Field Projector

Within the last few years, 3D movie theaters have become so popular and wide-spread that most new movies are released in 3D; even classics are often re-rendered to fit the

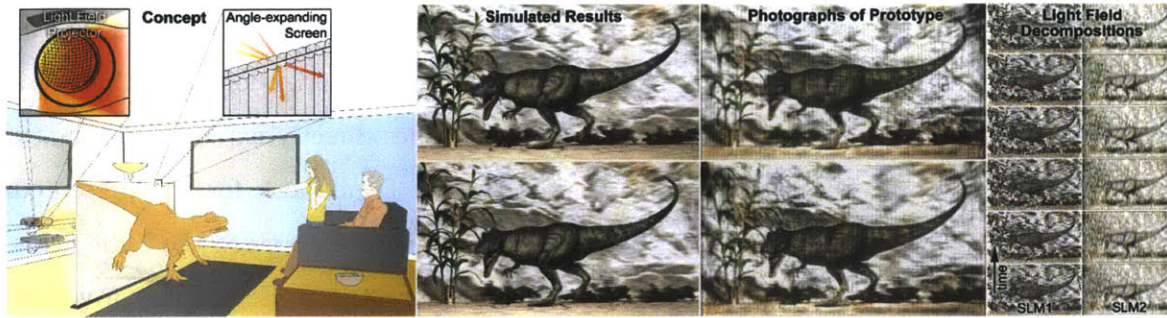


Figure 6-17: Compressive light field projection for glasses-free 3D display. The system comprises a single light field projector and a completely passive screen. The angular range of the light field emitted from the projector is limited to the size of the projection lens aperture, hence very small. Keplerian telescopes inspire our screen design—the angular range of incident light is expanded for an observer on the other side, creating a field of view that is suitable for glasses-free 3D display. A prototype projector was implemented from scratch using two high-speed spatial light modulators (SLMs); a prototype screen was fabricated from two lenticular sheets with different focal lengths, mounted back-to-back. With the implemented system, we achieve high-rank light field synthesis (center) for human observers with a critical flicker fusion threshold that is smaller than the product of the SLM refresh rates and the rank of the synthesized light field. Note that color results above are composited from multiple images captured from our grayscale prototype.

increasing demand for 3D content. For many people, the experience of watching a 3D movie on a large screen is significantly more immersive than conventional 2D screenings or watching smaller-scale 3D content on TV. Commercially available 3D projection technology is based on stereoscopic principles (review in Section 2.3.1), usually with special eye-wear. This approach can create viewer discomfort; furthermore, the correct perspective is only observed from a single sweet-spot in center of the theater.

As opposed to stereoscopic image generation, light field displays provide physically correct views for a wide range of perspectives and do not require an observer to wear special glasses (review in Section 2.3.2). Interestingly, inventors worldwide have investigated large-scale light field projection systems throughout the last century [59]. Several light field movie theaters were open to the public in Russia and France in the 1940s. Most of these and subsequent installations employ large parallax barrier-type screens, resulting in severe loss of image resolution and light throughput. Today, larger-scale light field projection systems are commercially available but require dozens of devices [12], making these systems expensive, power hungry, bulky, and difficult to calibrate.

In this section we present the first compressive light field projection system as an application

of the Tensor Display Framework. The proposed system combines a novel, passive screen, a single high-speed light field projector, and tensor light field factorization algorithms. As described in Section 3.4. the employed factorization routines directly exploit redundancy in the target content. Application of the Tensor Display Framework for the presented projection system not only reduces the memory footprint needed to store the light fields but also the number of projection devices required to display them. Hence, the proposed system is compressive in a computational and an optical sense. Through the co-design of display optics and careful application of the Tensor Display Framework, we devise a practical solution to large-scale light field display.

Superlenses One part of this system incorporates an angle expanding screen and one possible implementation of such a screen is a superlens composed of back-to-back lenticular sheets. Though not widely used today, Gabor superlenses, as such arrangements are described in optics literature [77], have been explored historically for their unique imaging properties. Dennis Gabor [60] demonstrated that varying the pitch and focal length of back-to-back lenticular sheets can create configurations that perform analogously to physically larger standard lens systems. In a closely related work Eichenlaub et al. [52] demonstrate that superlenses can be used to enlarge a volumetric display, though other means of generating large volumetric displays have been shown [112, 175].

6.2.1 Compressive Light Field Synthesis

In this section, we derive the optical image formation of the proposed system as well as related optimization techniques. The formulations are derived in 2D “flatland”, but extensions to the full 4D case are straightforward.

Optical Image Formation

Consider a conventional rear-projection system. The projection lens re-images and magnifies the pattern displayed on an internal spatial light modulator (SLM). A diffusing transmissive

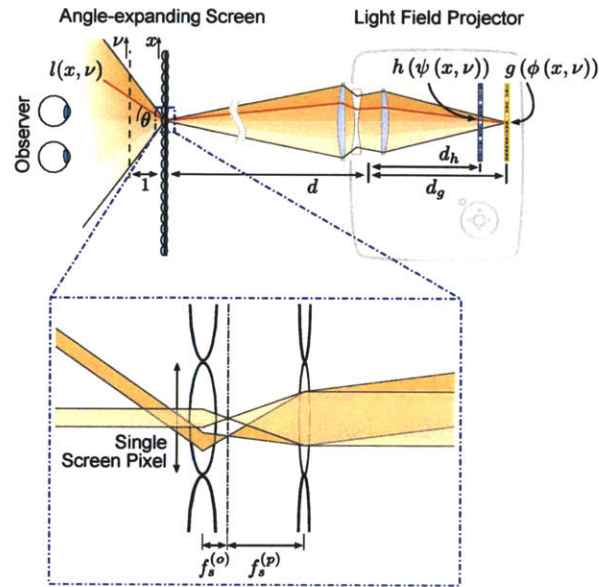


Figure 6-18: Overview of light field projection system. Two spatial light modulators, g and h , synthesize a light field inside a projector (top right). The projection screen is composed of an array of angle-expanding pixels (bottom). Inspired by Keplerian telescopes, these pixels expand the field of view of the emitted light field for an observer on the other side of the screen.

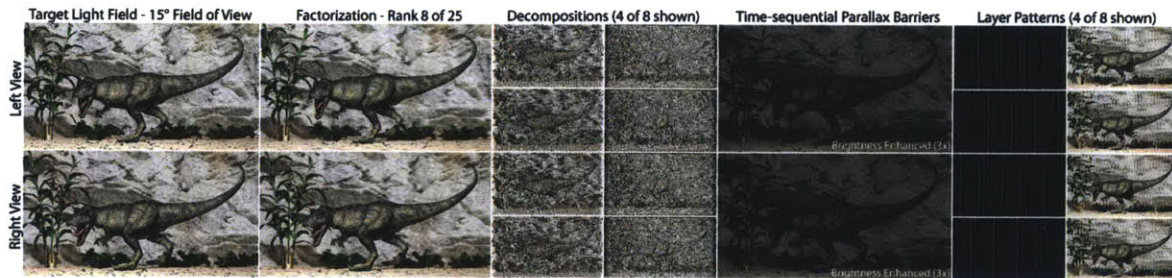


Figure 6-19: Light field factorization and comparison to time-sequential parallax barriers. Two views of a target light field with horizontal-only parallax and 25 views equally distributed over a field of view of 15° are shown on the left. Using the framework proposed in this paper, the light field is decomposed into a set of patterns for two spatial light modulators (SLMs) running at 480 Hz (center). When observed by a human, these decompositions create a rank-8 approximation of the light field (center left). The alternative to factorized image synthesis is display of time-sequential parallax barriers (right), which are $7.5\times$ darker than our method and require 1500 Hz SLMs to achieve the same resolution (center right).

screen is placed at the conjugate plane of the SLM, such that an image can be observed over a wide range of viewing angles from the other side of the screen. The light field on the viewer side is engineered to be as *view-independent* as possible and directly corresponds to the SLM image g :

$$\tilde{l}(x, \nu) = g(x). \quad (6.13)$$

While this approach is effective for presenting two-dimensional images, we are interested in emitting *view-dependent* 4D light fields. For this purpose, two modifications to conventional projection systems are necessary. First, the projector has to emit a light field and not just a 2D image. Second, the screen has to preserve the incident angular variation. Unfortunately, diffusing screens in most existing projection setups optically average an incident light field in the angular domain and eliminate high frequency directional variation. To overcome this limitation, we introduce the notion of an angle-preserving screen that changes the image formation to

$$\tilde{l}(x, \nu) = g(\phi(x, \nu)), \quad (6.14)$$

where each light ray (x, ν) on the observer side of the screen is mapped to the SLM inside the projector by the function $\phi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$. We adopt a two-plane parameterization of the light field, where x is the spatial coordinate on the screen and $\nu = \tan(\theta)$ the point of intersection with a relative plane at unit distance (see Figure 6-18).

In addition to the angle-preserving screen, the projector also needs to be modified so as to emit a light field. Such projectors have been proposed in the past; possible options include microlenses or a pinhole mask near the image SLM and coded projector apertures (e.g., [67]). We follow the design presented in Section 3.3 and use two programmable, light-attenuating SLMs inside the projector (see Figure 6-18). The image formation is now given by the multiplication the the patterns g and h shown on the two SLMs:

$$\tilde{l}(x, \nu) = g(\phi(x, \nu)) h(\psi(x, \nu)). \quad (6.15)$$

Similar to ϕ for g , $\psi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ maps each ray in the light field to a position on the second SLM h . Using ray transfer matrices [73], these mapping functions can easily be derived given the distance between screen and aperture d , the ray transfer matrix of the screen \mathbf{T}_s , the focal length of the projection lens f_p , and the distance d_g from the aperture to the SLM:

$$\begin{pmatrix} \phi(x, \nu) \\ \zeta \end{pmatrix} = \begin{pmatrix} 1 & d_g \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f_p & 1 \end{pmatrix} \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \mathbf{T}_s \begin{pmatrix} x \\ \nu \end{pmatrix} \quad (6.16)$$

The incident ray angle ζ is disregarded in the following; ψ is similar to ϕ but replaces d_g by d_h . All system parameters are illustrated in Figure 6-18. While Equation 6.16 models the ray transfer under the assumption of perfect optics, aberrations can be incorporated into ϕ as well.

Angle-expanding Screen Design

Independent of the specific method of light field synthesis within the projector, the resulting light field will have a narrow angular range that varies only over the aperture of the device. Unfortunately, this limited range is insufficient for an observer to freely move and enjoy glasses-free 3D display within a reasonable field of view . To address this problem, we propose a screen that not only preserves angular variation but *expands* it.

Angle expansion is a common technique in optics that is for instance used in Keplerian telescopes. These telescopes perform angle expansion with two lenses of different focal lengths mounted such that the distance between them is equal to the sum of their focal lengths. Inspired by this idea, we propose a screen that comprises an array of miniature angle-expanding telescopes—one for each screen pixel. This design is illustrated in Figure 6-18 (close-up). Whereas the spatial extent of a beam incident from the right is reduced, its

incident angle is amplified on the observer side of screen. The ray transfer matrix \mathbf{T} of such a Keplerian angle expander can be modeled as

$$\mathbf{T} = \begin{pmatrix} 1 & 0 \\ -1/f_s^{(p)} & 1 \end{pmatrix} \begin{pmatrix} 1 & f_s^{(p)} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & f_s^{(o)} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f_s^{(o)} & 1 \end{pmatrix} \quad (6.17)$$

where $f_s^{(o)}$ and $f_s^{(p)}$ are the focal lengths of screen lenslets facing the observer and the projector, respectively. As illustrated in Figure 6-18, a simple design of the proposed screen uses two lenslet or lenticular arrays with the same lens pitch but different focal lengths. Mounted back to back and with a lens size corresponding to the pixel size on the screen, \mathbf{T}_s becomes

$$\mathbf{T}_s = \begin{pmatrix} 1 & 0 \\ 0 & -f_s^{(o)}/f_s^{(p)} \end{pmatrix} \quad (6.18)$$

Note that the dependence on ray position in a single angle-expander vanishes for the entire screen in Equation 6.18, because each lenslet has the same size as a projected image pixel. The refractive effect of the proposed screen only depends on the incident ray angle (i.e. $\nu_p = -f_s^{(o)}/f_s^{(p)}\nu_o$), which is flipped and amplified by an angle-expansion factor of $M = f_s^{(p)}/f_s^{(o)}$. Although the screen is fundamentally limited by diffraction, this effect is negligible in the proposed system because pixels on projection screens are usually large (millimeters as opposed to microns).

Efficient Light Field Synthesis

The most intuitive way to generate a light field inside the duallayer projector is to display an array of pinholes on one screen and the interlaced views of the light field on the other [94]. Unfortunately, as discussed throughout this thesis, this approach generates low-resolution images and is also extremely light-inefficient. We follow the methods developed

in Section 3.4. Here we adapt the factorization algorithm to the proposed system by incorporating the effects of projection lens and angle-amplifying screen via mapping functions ϕ and ψ .

Specifically, following the formulation in Section 3.4.1, Equation 3.39, the image formation (Eq. 6.15) is discretized as

$$\tilde{\mathbf{l}} = (\mathbf{\Phi}\mathbf{g}) \circ (\mathbf{\Psi}\mathbf{h}), \quad (6.19)$$

where $\mathbf{\Phi} \in \mathbb{R}^{L \times N}$ and $\mathbf{\Psi} \in \mathbb{R}^{L \times M}$ are matrices that permute the rows of the discrete SLM patterns $\mathbf{g} \in \mathbb{R}^N$ and $\mathbf{h} \in \mathbb{R}^M$ according to the mapping in $\phi(x, \nu)$ and $\psi(x, \nu)$, respectively, and \circ is the Hadamard or element-wise product. In this notation, the emitted light field is represented as a discrete vector $\tilde{\mathbf{l}} \in \mathbb{R}^L$. The matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are sparse (usually one non-zero value per row) and constructed via raytracing for simulations (Eqs. 6.16, 6.17) or using calibration that accounts for optical aberrations in practice (Section 6.2.2).

Equation 6.19 makes clear that the emitted light field is the product of two permuted vectors, hence rank-1. Following the approach developed in Section 3.4, we employ high-speed SLMs that operate at refresh rates beyond the critical flicker frequency of the human visual system. Images displayed at such refresh rates are perceptually averaged. In particular, we model high-speed SLMs in the proposed setup as

$$\tilde{\mathbf{l}} = \frac{1}{T} \sum_{t=1}^T (\mathbf{\Phi}\mathbf{g}_t) \circ (\mathbf{\Psi}\mathbf{h}_t) \quad (6.20)$$

Here, T pairs of displayed patterns are averaged by the visual system and create a perceived rank- T light field $\tilde{\mathbf{l}}$. The temporally-changing patterns on the SLMs at time t are \mathbf{g}_t and \mathbf{h}_t . Given a target light field $\mathbf{l} \in \mathbb{R}^L$, an optimization problem can be formulated to find the best set—in a least-squared error sense—of time-varying patterns as

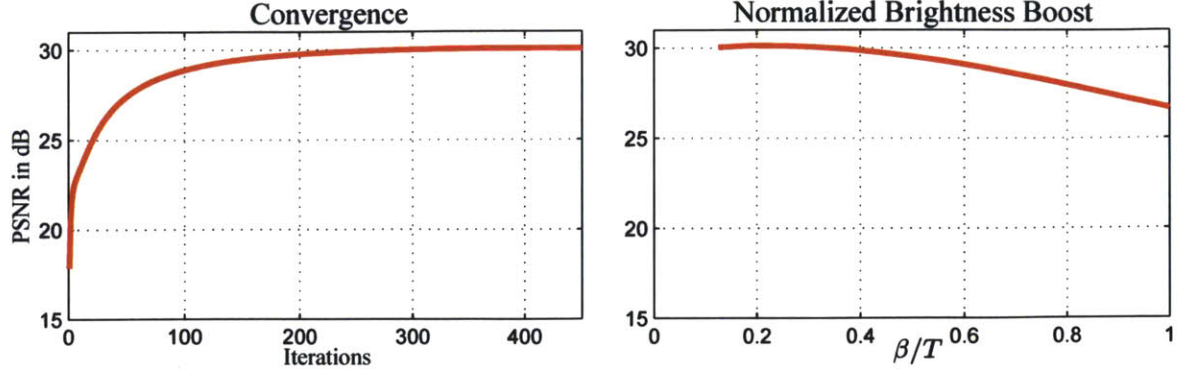


Figure 6-20: Quantitative evaluation of convergence and brightness boosting factor β in Equation 6.21 for the light field projector. In this example, the proposed update rules converge after about 200 iterations (left). The brightness of the target light field can be boosted as compared to conventional, time-sequential methods; a higher brightness, however, results in a slight decrease in reconstructed light field quality.

$$\begin{aligned}
 \arg \min_{\mathbf{g}, \mathbf{h}} \quad & \left\| \beta \mathbf{1} - \sum_{t=1}^T (\Phi \mathbf{g}_t) \circ (\Psi \mathbf{h}_t) \right\|_2^2 \\
 \text{subject to} \quad & 0 \leq g_{ik}, h_{jk} \leq 1, \forall i, j, k
 \end{aligned} \tag{6.21}$$

Note that β absorbs the factor $1/T$ as well as a user-defined brightness boost (see Figure 6-20). The nonnegativity constraints ensure that optimized patterns are physically feasible. Although this is a nonlinear and nonconvex problem, it is biconvex in \mathbf{g} and \mathbf{h} ; fixing one results in a convex problem for updating the other. Such updates are usually performed in an alternating and iterative manner. We derive multiplicative matrix update rules for our problem as:

$$\begin{aligned}
 \mathbf{g}_t & \leftarrow \mathbf{g}_t \circ \frac{\Phi^T (\beta \mathbf{1} \circ (\Psi \mathbf{h}_t))}{\Phi^T (\tilde{\mathbf{1}} \circ (\Psi \mathbf{h}_t)) + \epsilon} \\
 \mathbf{h}_t & \leftarrow \mathbf{h}_t \circ \frac{\Psi^T (\beta \mathbf{1} \circ (\Phi \mathbf{g}_t))}{\Psi^T (\tilde{\mathbf{1}} \circ (\Phi \mathbf{g}_t)) + \epsilon}
 \end{aligned} \tag{6.22}$$

where \circ and $-$ denote element-wise product and division, respectively, ϵ is a small value that prevents division by zero, and $\tilde{\mathbf{1}}$ is computed via Equation 6.20.

Multiplicative update rules for nonnegative matrix factorization problems have become increasingly popular in the scientific computing community (e.g., [122, 39]). We extend these methods by including the projection matrices Φ and Ψ into the solver. The update rules in Equation 6.22 are *mathematically distinct* but *numerically equivalent* to the conventional multiplicative update rules presented in Section 3.3.1. This extension has the advantage of not only allowing for an elegant mathematical formulation of arbitrary optical setups, but also for extremely *efficient implementations*. As discussed in more detail in Section 6.2.2, Φ and Ψ can be implemented as a multiview rendering step whereas Φ^T and Ψ^T correspond to projective texture mapping. These operations are hardware-accelerated on the GPU and can be implemented in real-time.

6.2.2 Implementation

Our prototype projection system comprises two optical hardware parts which can conceptually be implemented independently of one another: an angle-expanding screen and a light field projector. This section provides recipes for both parts.

Light Field Projector The projector places two spatial light modulators (SLMs) at different distances behind a projection lens to create angular variation across the lens aperture. As is apparent in Figure 6-18, the field of view of the system will be maximized by choosing a projection lens with a large aperture relative to its focal length or, similarly, a small f-number. For a fixed screen distance, the image size will be maximized with a shorter focal length lens. We choose a Nikon Nikkor 35mm f/1.4 AI-s lens for our prototype (Figure 6-21, b).

The SLMs are reflection mode Liquid Crystal on Silicon (LCoS) modulators (Silicon Micro Display ST1080, Figure 6-21 g). To achieve an optical path equivalent to that of Figure 6-18 with reflective modulators, we employ two polarizing beamsplitter cubes (Figure 6-21, c). The physical extent of the beamsplitter cubes requires an additional 1:1 relay lens to optically place both SLMs close to each other. The f-number of the relay lens should match

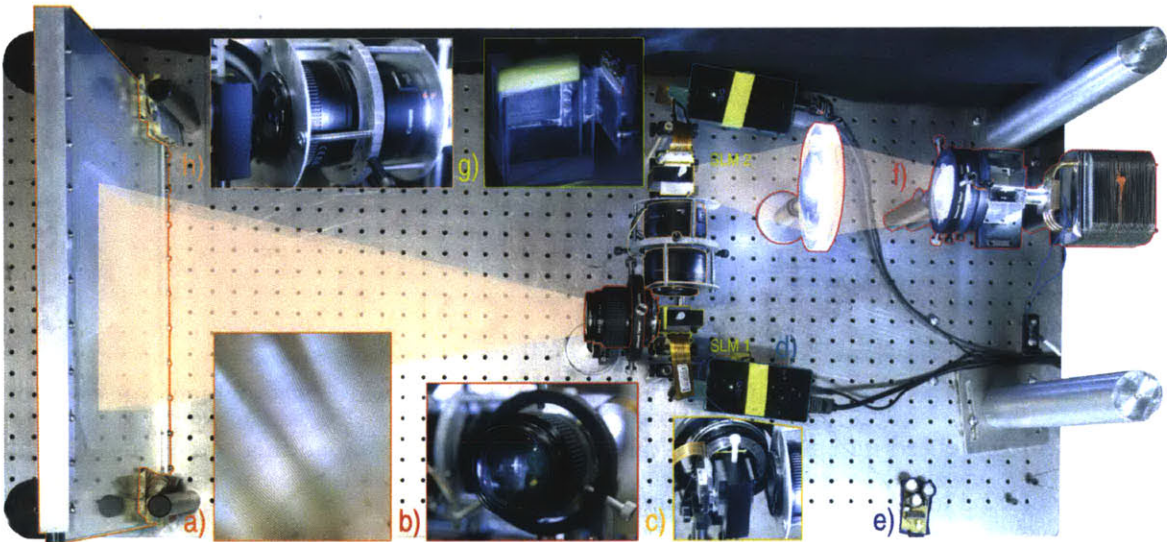


Figure 6-21: Overview of prototype light field projection system. The projector (right) emits a 4D light field with a narrow field of view that only varies over the projection lens (b, Nikkor 35mm f/1.4 AI-s). This angular range is expanded by the screen (left) for an observer on the other side. The screen (a) is composed of passive pixels that each expand the angles of all incident light, just like a Keplerian telescope. No special calibration w.r.t. the projector is necessary beyond focusing the latter on the screen. The projector emits a 4D light field, which is synthesized by two reflective spatial light modulators (SLMs, Silicon Micro Display ST1080). Their contribution is optically combined by a 1:1 relay lens (h, 2× Canon EF 50 mm f/1.8 mounted face-to-face). The light source (10W LED) is synchronized to the refresh rate (240 Hz) of the SLMs by a custom board (e). The SLMs use liquid crystal on silicon (LCoS) technology, which requires polarizing beam splitter cubes (c), and are connected to a standard graphics card via a driver board (d).

that of the projection lens. We use two Canon EF 50mm f/1.8 II lenses mounted face to face (Figure 6-21, h). Although this compound relay lens limits the f-number of the system, it provides high image quality and minimizes optical aberrations. The ST1080 modulator operates at 240Hz and is driven by a driver board (Figure 6-21, d) that is intended to run the LCoS for a head mounted display. Assuming a critical flicker fusion rate of about 40Hz for the human visual system, which is reasonable for low-light conditions, the available refresh rates allow a rank-6 monochrome light field decomposition.

The illumination unit in the projector has to match the f-number of the system. It should also be uniform over its spatio-angular extent and be synchronized with the frame updates of the SLMs, meaning the illumination source must be switchable at 240Hz. These constraints can be met with high-power LEDs; we place a 10W LED (similar models can be purchased from Cree, Inc.) mounted on a heat sink behind a mirrored light pipe (Figure 6-21, f). The light pipe is taken out of a conventional projector and acts as a “kaleidoscope”, virtually cloning the LED to a larger illumination range. Care is taken to place the LED image out-of-focus with any of the SLM planes, screen, or viewer location. Additional off-the-shelf lenses are used to form a converging beam on the rear SLM. A custom circuit board (Fig 6-21, e) employs a microcontroller and a power field-effect transistor to switch the LED in sync with the frame updates of the SLM.

Angle-expanding Screen In principle, a horizontal-only expander can be implemented by placing two lenticular sheets of different focal lengths back-to-back (Figure 6-21, a). However, the design tolerances of off-the-shelf lenticulars make it difficult to fabricate angle-expanding screens with suitable characteristics in practice. We were able to have a horizontal-only angle-expanding screen with an expansion power of $M = 3$ custom manufactured by Microsharp Innovation. To support a range of vertical viewpoints, the screen requires an additional vertical-only diffuser. We use a horizontally oriented 100 lpi 31° Lenstar Plus 3D lenticular from Pacur. Alternatively, holographic uni-directional diffusers, for instance from Fusion Optix or Luminit, can be used.

Unlike typical lenticular displays, the proposed screen does not decrease the resolution of

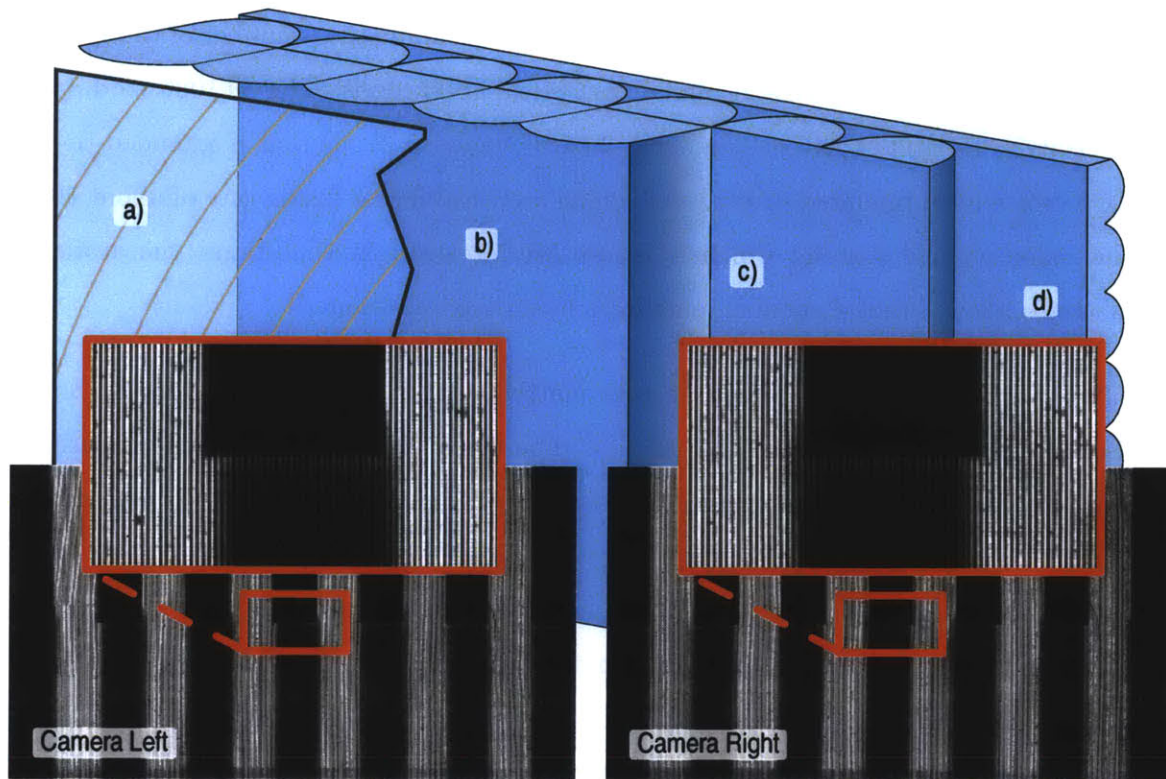


Figure 6-22: Top: parts of the prototype angle-expanding screen: a) Fresnel lens, back-to-back lenticular sheets on b) projector-side, and c) viewer-side, and d) overlaid vertical diffuser. Bottom: test images captured from the extreme viewing angles demonstrating parallax. The closeups show vertical stripes caused by the lenticular of the angle expanding screen. These are not apparent when observed by eye.

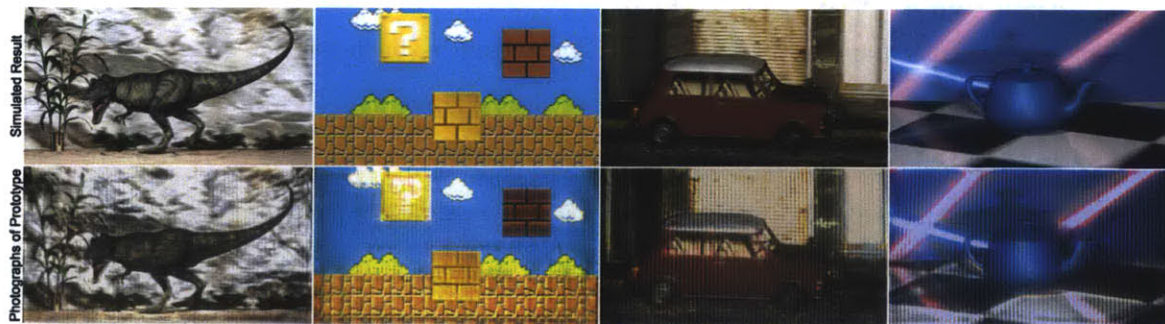


Figure 6-23: Overview of experimental results. Each column shows the central view of a light field that comprises eight views with horizontal-only parallax. Simulated results (top row) are compared with photographs of the prototype light field projector (bottom row). Color results are composited from three photos of our grayscale prototype.

the projected image. Ideally, each lenslet or lenticular has the size of a projected pixel on the screen. To maximize image resolution, these properties should be optically matched. In our current implementation, the projection lens is located 50cm away from the screen and produces an image with a size of 21.3×11.7 cm. The lenticular pitch of the screen is 0.5mm, which currently limits the achieved image resolution in the prototype to 426×720 pixels. A larger image size or smaller lenticulars could increase this resolution. Also note that, unlike a typical lenticular display, no precise horizontal alignment between the SLM image and the angle-expanding screen is required.

The screen lenticulars of our prototype have the same pitch. However, to achieve a viewing zone at a distance from the screen greater than that of the projector, the pitch should be adjusted such that the screen acts as an angle-expander and simultaneously as a lens focusing illumination into the viewing zone. For the prototype setup, we can achieve the same effect with an additional Fresnel lens mounted close to the screen on the projector side. The entire optical stack can be seen in Figure 6-22. With complete design freedom the entire screen optics could equivalently be fabricated as a single, large-scale sheet in a roll-to-roll process.

System Calibration Creating a projection system from scratch relies on careful calibration of each component. An optical rail system constraints many unnecessary degrees of freedom in the prototype projector. Approximate alignment of each component is achieved by probing with a laser. Once both SLM images can be observed through the projection lens, a checkerboard pattern is projected to focus both SLMs independently and overlay them precisely. As a final verification step, a bar target (Figure 6-22) is displayed on the prototype and photographed. The top bars, displayed on SLM 1 (Figure 6-21) are in sharp focus on the screen, while the bottom bars form a virtual image in front of the screen, and demonstrate motion parallax as the camera is moved.

Note that the front-focused image cannot be focused as sharply as the rear-focused image due to optical aberrations in the angle-expanding screen. We characterize the point spread function (PSF) of the angle-expanding screen by displaying a point on SLM 2 and taking

a RAW photograph with subtracted blacklevel. The recorded PSF is approximated as a 2D Gaussian and incorporated into the light field factorization. We also characterize the intensity transfer function of the SLMs, which are not well approximated by a standard Gamma curve. For this purpose, RAW photos of the screen are taken from the center of the viewing zone while the prototype displays a sequence of different intensities over the input range of the SLM driver. The inverses of the resulting curves are applied to the factorizations computed by the solver.

Software Implementation Target light fields are rendered using POV-Ray, but any graphics engine could be used alternatively. We implement the nonnegative light field factorization routines (Eq. 6.22) on the GPU using OpenGL and Cg. Decomposing a light field with eight horizontal views and an image resolution of 1280×720 pixels takes about one minute on an Intel Core i7-2600 PC with an Nvidia GeForce GTX 690 GPU. Including convolution operations with the point spread function modeling screen aberrations increases processing times by a factor of $10 - 20\times$, depending on the PSF size. The finite blacklevel of each SLM is taken into consideration by clamping the values of \mathbf{g} and \mathbf{h} to the feasible range after each iteration (see Eq. 6.22).

6.2.3 Assessment

We simulate factorized light field synthesis for the proposed projection system in Figure 6-19. For this experiment, we decompose a light field with 25 views (two of them shown) into eight pairs of time-multiplexed patterns. This choice simulates 480 Hz spatial light modulators that create a rank-8 light field approximation for an observer with a critical flicker frequency of 30 Hz. Device dimensions match those of the prototype (see Section 6.2.2). The light field can be reproduced with a high image quality (center left, PSNR 26.4 dB) using this configuration. In comparison, a time-sequential parallax barrier display would require 25 time-multiplexed images to achieve the same resolution and 1500 Hz SLMs. In addition, the normalized brightness boost β/T for the factorized result is chosen to be 0.3, which makes the observed light field $7.5\times$ brighter than the parallax-barrier display mode.

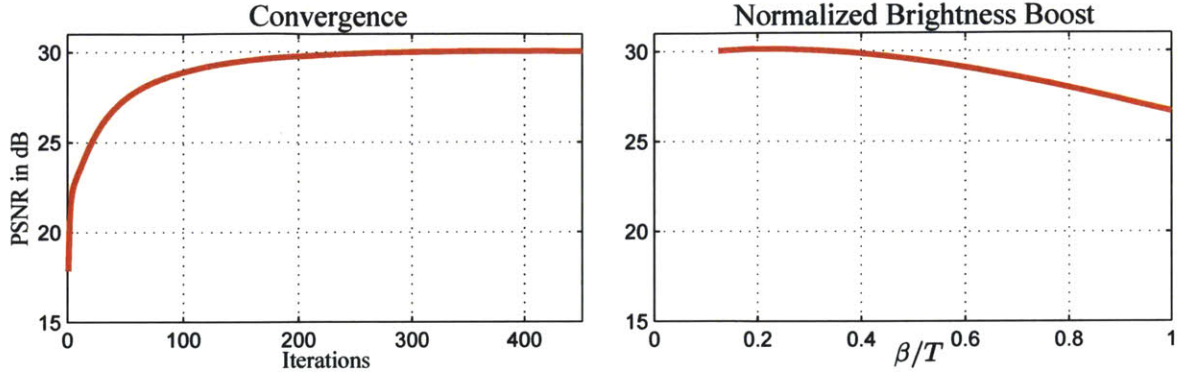


Figure 6-24: Quantitative evaluation of convergence and brightness boosting factor β in Equation 6.21. In this example, the proposed update rules converge after about 200 iterations (left). The brightness of the target light field can be boosted as compared to conventional, time-sequential methods; a higher brightness, however, results in a slight decrease in reconstructed light field quality.

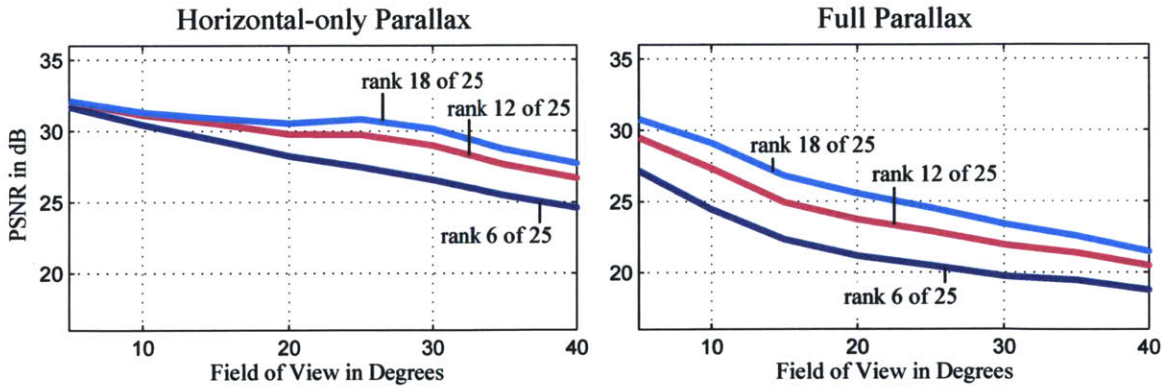


Figure 6-25: Light field compressibility. We simulate reconstructions of the “t-rex” scene for a varying field of view. The target light field has either 25 or 5×5 views equally distributed in a horizontal-only (left) or horizontal and vertical (right) viewing zone, respectively. Horizontal-only parallax (HOP) light fields are much more compressible; higher-rank decompositions achieve a better quality. We observe that rank-6 decompositions with HOP for fields of view up to 20° achieve high-quality reconstructions.

Looking at the layer decompositions (Figure 6-19) shows how light-inefficient parallax barriers are. While the patterns for one of the SLMs contain the interlaced views of the light field, the other comprises a set of vertical slits that block most of the light (images may appear black in printout). The factorized patterns are much more light efficient but less intuitive. As observed in previously in Section 3.3 and Section 3.4, we interpret the patterns as distributing low image frequencies in the 3D scene to the closest SLM while depth discontinuities in the light field create high-frequency, temporally-varying structures. These can be interpreted as content-adaptive parallax barriers that are automatically created where needed: around edges and scene features that extrude from the physical device.

We also plot the convergence of the proposed algorithm in Figure 6-20 (left). After about 200 iterations, no significant improvements in image quality, measured in peak signal-to-noise (PSNR) ratio, are observed. The brightness boosting factor β (see Eqs. 6.21, 6.22) can be freely chosen to trade 3D image quality for brightness. We analyze this tradeoff in Figure 6-20 (right). A normalized value of β/T of 0.2–0.3 results in high-quality reconstructions. For the example shown in Figure 6-19, we chose $\beta/T = 0.3$ which results in a direct brightness boost of factor $7.5\times$ over conventional time-sequential parallax barriers.

We also show a quantitative evaluation of light field compressibility in Figure 6-25. Both horizontal-only and full parallax light fields are considered for decompositions with rank 6, 12, and 18. In all cases, the target light field has 25 views equally spaced over the entire 2D field of view (FOV). Intuitively, light fields containing only horizontal parallax are much more compressible, which is confirmed by higher PSNR values. As the FOV increases, compressibility of the light field decreases due to larger parallax. The small “bumps” in the left plots are discretization artifacts.

Scaling the system

A system of the size depicted in the concept sketch to the left of Figure 6-17 (2m screen size) can reasonably be obtained by scaling the optical properties of the current projector components. The field of view (FOV) of a projection system following the schematic shown

in Figure 6-18 is given by

$$\text{FOV} = 2M \arctan \left(\frac{h_i}{2N(h_s - h_i)} \right), \quad (6.23)$$

where h_i and h_s are the SLM and screen heights, respectively, N is the effective f-number of the projection optics, and M is the power of the angle expanding screen. Note that the dependence of FOV on f-number alone, rather than the focal length or exit pupil size of the projection lens, remains so long as the projector-to-screen distance is not fixed.

Plausible real-world values for the variables in Equation 6.23 can be obtained from commercial catalogs and published academic work. The catalogs of Pacur and Micro Lens Technology Inc. contain commodity microlenses that range in focal length from 6.35mm to 0.26mm (although they differ in pitch), suggesting a plausible value for M approaching 25. A conservative estimate derived from analysis of similar screen optics in Eichenlaub et al. [52] is $M = 10$. Commodity 35mm camera lenses with f-number as small as f/1.1 exist, such as the Voigtlander Nokton 50mm f/1.1. Therefore, from Equation 6.23 it follows that with an $N = 1.1$ lens, $M = 10$ power screen, $h_s = 2000\text{mm}$ and $h_i = 36\text{mm}$, equivalent to a typical working area for a 35mm lens, a projector placed 2.7m from the screen would produce a 2m wide light field image with a 10° FOV.

As previously discussed, wider fields of view can be obtained by employing multiple devices, such that a 20° FOV can be obtained with two devices, a 30° with three, and so on. Real-time color can be achieved using three devices of the type described, or one device with an SLM capable of switching at $3 \times 240\text{Hz} = 720\text{Hz}$. Achieving wider field of view with a single device will require smaller f-number projection lenses, or alternative screen optics, which we leave to future work.

Limitations

The major limitation of the proposed system is the image quality achieved with the prototype setup. A maximum refresh rate of 240 Hz limits us to show rank-4–6 grayscale light

fields for a human observing the prototype. Higher-speed SLMs with field sequential color or multi-device setups could address this limitation. The image quality of the prototype projector is limited by vignetting, optical field curvature by the beam splitter cubes, scattering in the screen, as well as color aberrations from the Fresnel lens. Further, the f-number of the projection system is currently limited to $f/1.8$ by the relay lens. The contrast of the SLMs is reduced by low f-number illumination, but this is inherently addressed and partially corrected for by the solver. The prototype screen provides an angular amplification factor of approx. $3\times$, resulting in a total field of view of approx. 5° achieved with the prototype. Future screen implementations should significantly increase this factor. Finally, the factorization adds additional computational cost to the system, but we are confident that real-time implementations are possible with optimized software on modern GPUs.

Summary

We have introduced a compressive light field projection system. Through the careful application of the Tensor Display Framework, and a novel passive screen design, we present the first single device approach to glasses-free 3D projection that does not require mechanical movement of screen elements. We believe that the proposed system has promise to scale to large sizes, such as movie theaters, although additional engineering efforts are necessary to achieve the required image quality and dimensions.

6.3 Soundaround: An 8D Display for Audio

This section presents a conceptual extension of the concepts of 8D display presented in Chapter 5 to audio.

Multi-view display hardware has made compelling progress recently in graphics. While multi-view stereoscopic displays have a direct functional audio analogy in multi-channel surround sound systems, to date, robust tools do not exist to design multi-listener audio systems to accompany multi-viewer display systems. In the multi-viewer use-case, a display

produces a light field that typically projects a distinct 2D plane into the view zone of each viewer, such that multiple observers perceive different images on the display. Similarly separating audio channels for each viewer currently requires headphones, which run contrary to the spirit of unencumbered (glasses free) multi-view displays. Other solutions such as parabolic reflectors or directional ultrasonic transducers [215] are expensive and can only be directed to single, fixed spatial locations.

In Section 6.3.2 we present a method for obtaining complex-valued weights describing the phase and amplitude of coherent audio frequency emitters arranged in an array to direct separate audio channels to each viewer in a multi-viewer system. We formulate the problem using a straightforward mathematical framework based on linear algebra. We use a straightforward path length metric, consistent with ray tracing and other common formulations in the graphics and optics communities, to populate a transfer function between emitters and observation points in the environment. Our formulation enables us to map the optimal solution to this over-constrained problem to an efficient quadratic program or pseudo inverse that can be solved in seconds in a Matlab script. The result of the optimization is a set of filters that form the basis for the associated real-time signal processing system.

We further show that our framework for addressing multi-view audio is general, and can handle exotic arrangements of emitters and observation planes.

We have implemented a prototype (Figure 6-26) to validate the framework presented in this paper. The prototype construction, operation, and validation are described in Section 6.3.3. This work will enable graphics and display researchers to use familiar tools from optics to design low-cost multi-viewer audio systems to accompany their multi-view displays.

Our work builds on existing techniques for creating directional audio fields. Techniques such as wavefront synthesis work well for simple situations, but do not provide control over nulls (i.e. minima in the energy response), or the ability to steer energy towards arbitrary configurations of users.

We do not explicitly account for reflections in the environment of the array in our wave propagation model. This is not a limitation of our system framework, as a more accurate

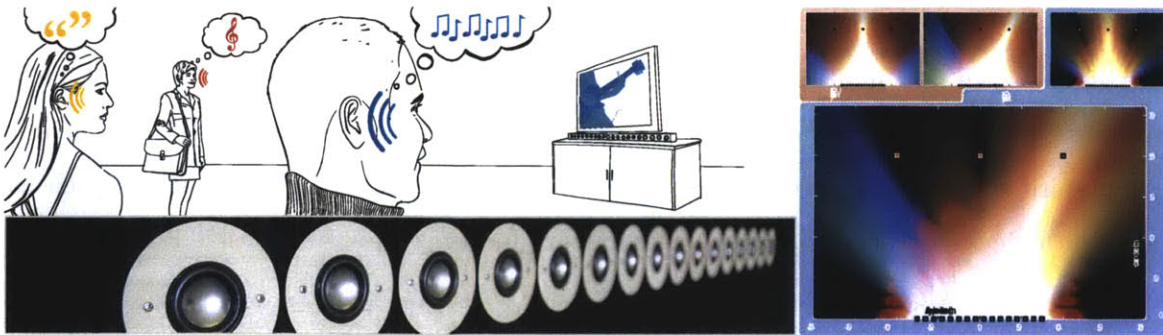


Figure 6-26: (*Top, Left*) Concept Sketch. A multi-view television is integrated with a multi-view audio system as described in this paper. Viewers of the system each receive a different audio and video stream, depending on their location. (*Bottom, Left*) Our linear 16-element emitter/receiver array prototype. Similar arrays are standard practice in applications from consumer audio to RADAR. We present a general optimization framework for directional audio, and apply it to a multi-user interactive audio system. We show that our technique is general, and can be applied to multiple observation planes and simulated or measured room responses. (*Right*) False color renderings comparing wavefront synthesis (*Red*) and the output of one possible optimization handled by our framework (*Blue*). Intensity at each spatial location corresponds to audio intensity. The color at each location denotes frequency (red to blue denoting 200 Hz to 3.4 kHz). Note that the optimization can be posed to introduce deeper nulls in the energy response than with the heuristic wavefront synthesis method.

acoustic model could be used if additional information about the viewing area is known, but this does limit the results achievable in our examples. The assumption of absorptive surfaces in our model will reduce the signal to noise ratio of the result in practice. We have shown that our prototype matches theory in open spaces, even in the presence of highly reflective surfaces. The contribution of reflection terms is further reduced by constraining the side lobes in our quadratic program formulation, as weak side lobes will be less likely to create strong diffuse reflections.

Our framework also accommodates placing microphone elements at observer locations to eliminate the need for modeling altogether. We describe this area of future work in more detail in Section 6.3.4.

6.3.1 Related Work

Phase-based steering using an array of transducers has long been common practice in many fields [193]. Terrestrial radars were among the earliest applications for delay-and-sum beam

formers, which sense directional plane waves to within the limit of the array's aperture size. This simple technique is effective in the far-field, when the observer or target is far from the array.

In the near-field, wavefront synthesis has been used by commercial audio equipment manufacturers to create focused point sources for large scale performance. We also see many consumer products, such as speaker phones, the Microsoft Kinect, and Sony PS3Eye, which use microphone arrays to locate users and isolate voice signals. Because these are commercial products we do not know what techniques are employed. Wavefront synthesis is a likely choice, however, as it is especially effective for the receiver problem where the roles of reflection and other environmental interference are reduced. Wavefront synthesis cannot achieve the goal of a true multi-user audio system, as it is not possible to directly control nulls. A wavefront synthesis system can focus different signals to multiple viewer locations, but cannot isolate the audio from one signal to another viewing location.

Recent work in the graphics community has considered novel tools such as the Wigner distribution [217] and the augmented light field [150] for modeling the propagation of coherent or partially coherent light. Cuypers et.al [46] describe the use of a ray-tracing engine, augmented with the Wigner distribution, for modeling sound wave propagation. While these tools are ideal for applying a graphics intuition to modeling the propagation of sound waves, we have found that mathematically simpler, phase-preserving tools, presented in Section 6.3.2, are better suited for posing sound field emission as an optimization problem.

6.3.2 Multi-View Audio

System framework

The overall system framework is illustrated in Figure 6-27. It consists of three stages, designed to process and direct independent audio sources toward viewers located at various positions in the viewing area using a fixed array of transducers. The first stage consists of a multiple-input, multiple-output (MIMO) signal processing system that processes the audio

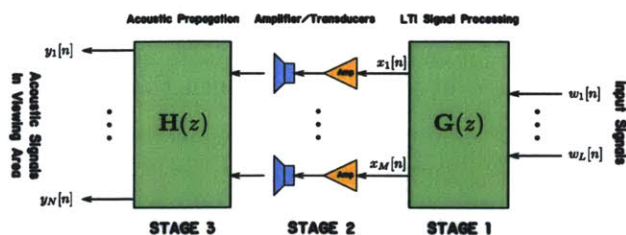


Figure 6-27: Multi-view audio system framework.

sources, the second stage consists of an array of transducers that convert the processed signals into acoustic signals, and the third stage represents acoustic propagation from the array to various points in the viewing area relevant to the formulation of the problem. As we are concerned with processing bandlimited signals all three stages are represented as discrete-time systems without loss of generality.

The overall strategy behind system design within this framework is to choose the array geometry and signal processing to best isolate the audio sources from the perspective of the viewers. While techniques for designing generally non-uniform array geometries can be used in achieving this goal, e.g. [13], this paper focuses on designing appropriate signal processing for use with linear arrays in particular.

Referring again to Figure 6-27, we specifically allow the processing performed in Stage 1 of our system framework to be linear, time-invariant (LTI) and discrete-time. We are concerned with designing a MIMO system that processes each of the input audio sources $w_k[n]$, $k = 1, \dots, L$. The output of the processing consists of M signals $x_k[n]$, $k = 1, \dots, M$, each of which is independently amplified and then converted into an acoustic signal using each of the N transducers. The overall MIMO signal processing system is represented by an M -by- L matrix \mathbf{G} of LTI system functions, i.e.

$$\underbrace{\begin{bmatrix} X_1(z) \\ \vdots \\ X_M(z) \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} G_{1,1}(z) & \dots & G_{1,L}(z) \\ \vdots & \ddots & \vdots \\ G_{M,1}(z) & \dots & G_{M,L}(z) \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} W_1(z) \\ \vdots \\ W_L(z) \end{bmatrix}}_{\mathbf{w}}, \quad (6.24)$$

where $X_k(z)$ and $W_k(z)$ respectively denote the z -transforms of $x_k[n]$ and $w_k[n]$, and where $H_{k,\ell}(z)$ denotes the z -transform of the impulse response $h_{k,\ell}[n]$ from the ℓ th input to the k th output, i.e. from $w_\ell[n]$ to $x_k[n]$. The column vectors \mathbf{x} and \mathbf{w} contain the z -transforms of the M signals $x_k[n]$ and L signals $w_k[n]$, respectively.

The amplifier-transducer systems in Stage 2 are designed to be individually identical single-input, single-output LTI systems. They therefore have a common system function, which can be shown to commute backwards through the \mathbf{H} matrix and may thus be compensated for by appropriately equalizing the input signals $w_k[n]$. Techniques for performing this compensation are addressed by the problem of acoustic equalization, the details of which constitute a broad area of research in the acoustics community.

We are interested in the response of the overall system at various viewing locations, where N denotes the number of relevant sample locations in the viewing area. Note that N will generally be different from the number of input signals L , also corresponding to the number of viewers. (This will particularly be the case in the later examples that are pertinent to multi-view audio systems designed to be robust to small changes in the specified viewing positions.) The acoustic propagation from each of the M transducers to the signals $y_k[n]$, $n = 1, \dots, N$ at the sample locations is modeled as a MIMO, LTI system with M inputs and N outputs. The system is notated as a matrix \mathbf{H} of system functions that map from the $X_k(z)$ to the $Y_k(z)$, the z -transforms of the signals $y_k[n]$. Specifically,

$$\underbrace{\begin{bmatrix} Y_1(z) \\ \vdots \\ Y_N(z) \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} H_{1,1}(z) & \dots & H_{1,M}(z) \\ \vdots & \ddots & \vdots \\ H_{N,1}(z) & \dots & H_{N,M}(z) \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} X_1(z) \\ \vdots \\ X_M(z) \end{bmatrix}}_{\mathbf{x}}. \quad (6.25)$$

System design

Combining Eqns. 6.24 and 6.25 results in the following equation relating the input signals $w_k[n]$ to the relevant acoustic signals $y_k[n]$ in the viewing area:

$$\mathbf{y} = \mathbf{H}\mathbf{G}\mathbf{w}. \quad (6.26)$$

The vector \mathbf{w} is the set of input sources to the system, and the matrix \mathbf{H} is determined by the physics underlying wave propagation in the viewing area. The strategy in designing the multi-view audio system is therefore to design \mathbf{G} so that the cascaded system $\mathbf{H}\mathbf{G}$ exhibits the desired response from the input signals \mathbf{w} to the relevant acoustic signals \mathbf{y} .

The overall system from \mathbf{w} to \mathbf{y} is LTI, and consequently the effect of the system on \mathbf{w} is fully-characterized by its discrete-time Fourier transform (DTFT), which is related to the matrices of z -transforms \mathbf{H} and \mathbf{G} by substituting $z = e^{i\omega}$. For a fixed value of $\omega = \omega_0$, Eq. 6.26 reduces to a matrix of complex scalars, and the system is fully characterized by evaluating a parameterization of the matrices over the range $-\pi < \omega \leq \pi$.

As we are concerned with designing \mathbf{G} to result in a desired response from \mathbf{w} to \mathbf{y} , an important consideration is how \mathbf{H} , the matrix mapping from the transducer signals to the acoustic signals in the viewing area, is obtained. There are many potential techniques for modeling \mathbf{H} or obtaining it via acoustic measurements. While the examples in this paper focus on using the wave equation and propagation without reflections in determining \mathbf{H} , the presented framework allows for the use of other techniques as well.

With a model for \mathbf{H} in place, the matrix \mathbf{G} representing the MIMO signal processing may be designed to obtain a desired response from \mathbf{w} to \mathbf{y} . As the number of signals N in \mathbf{y} generally be much larger than the number of input signals L in \mathbf{w} , the problem will generally be overcomplete, necessitating some type of trade-off between constraints. There are many potential techniques for addressing this, including convex optimization, as well as heuristic methods such as wavefront synthesis. In the later examples we focus on the use of wavefront synthesis and weighted least-squares optimization to obtain \mathbf{G} by substituting $z = e^{i\omega}$ and designing an ensemble of matrices \mathbf{G} parameterized by ω , sampled densely over the interval $-\pi < \omega \leq \pi$.

6.3.3 Implementation

Prototype

We have implemented a 16 element receiver and emitter array (Figure 6-26) to verify the theoretical results presented in Section 6.3.2. In this section, we describe the hardware and software necessary to construct the array presented in this work.

Hardware

Our speaker array is comprised of 16 Aurasound 2" NSW2-326 drivers, driven by 8 2-channel T-Amp amplifiers. The T-Amps are driven in their low THD region to prevent distortion. The receiver array uses 16 custom microphone boards, designed around the low-noise Panasonic WM61 electret microphone. The emitter and receiver elements are separated by 10 cm, making the total extent of the array 150 cm. Two 8 channel M-Audio Delta 1010lt sound cards are used to drive the array. Each card has 8 input, and 8 output channels. The cards sample clocks are synchronized using a S/PDIF cable to ensure that the phase of the audio sent to and received from the array is consistent between the two cards. The processing requirements for driving the array are minimal. The array is driven by a dual-core Pentium 4 3.4Ghz desktop computer with 1GB of RAM.

Software

The computer driving the array is running Fedora 14 Linux, using the CCRMA realtime kernel. The JACK Audio Connection Kit (JACK) is a low-latency audio routing API, which is used to route audio signals to and from the array. Software to drive the array is written in Matlab and Pure Data (PD). We have additionally implemented real-time filters, using jconvolver, to apply optimization-derived complex weights to each array element, per frequency band.

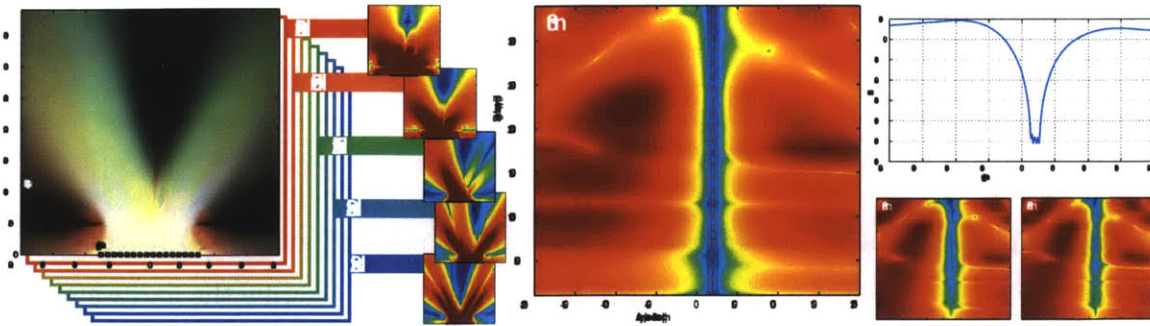


Figure 6-28: (*Left*) A null is aimed 300 cm from the array, and 20 cm to the right. This false-color rendering indicates the energy content at a range of frequencies from 200 Hz to 3.4 kHz. Lower frequency regions are more red, while higher frequency areas are more blue. Five contributing frequencies, along with their intensity distributions, are shown to the right. Colored tabs indicate the color used to label the frequency band in the false-color rendering. (*Middle, Bottom Right*) The intensity at each frequency band is plotted as a function of distance along the array at the target location of 300 cm. The same plot is also generated for a distance of 280 cm and 320 cm from the array. (*Top, Right*) The theoretical response of the array is plotted, in dB at 300 cm.

6.3.4 Assessment

The framework presented herein will enable graphics researchers to develop and integrate true walk-up multi-user audio and video systems. Complex room geometries or multi-level spaces will benefit from extending the techniques presented here to two-dimensional arrays.

While our framework is independent of the technique used to generate the transfer matrix between emitters and listeners, using more sophisticated models or real-time measurements will improve the agreement between theory and practice.

Results

As with any optimization, the validity of the result will depend on the suitability of the chosen objective function. In this section we apply our framework to the problem of creating a wide null. We limit ourselves to equality constraints, resulting in a weighted least-squares problem to be solved over a dense sampling of discrete-time frequencies ω (corresponding to the range of pertinent wavelengths). This problem is both relevant for multi-viewer audio systems, and difficult to achieve using heuristic methods.

In Figure 6-28, we demonstrate the result of running our optimization framework to target the creation of a null at 300 cm from the array, and 20 cm to the right. The transfer matrix, \mathbf{H} , between the output of the speakers and the viewer locations is populated by spherical wave propagation. The distance $d = \sqrt{(x - x_s)^2 + d_o^2}$ is evaluated for each pair of speaker locations and samples on the observer plane, where x_s is the position of the speaker in the plane of the array, and d_o is the orthogonal distance between the plane of the array and the observer plane. Elements of $\mathbf{H}|_{z=e^{i\omega}}$ are then of the form $\frac{1}{d^2}e^{-i\omega}$, where $\omega = \frac{2\pi d}{\lambda f_s}$, and where f_s is the sampling rate of the system.

In the interest of computational complexity, we elect to fill the \mathbf{G} matrix using results from weighted least squares optimizations. In this formulation we create a vector of equality constraints which represent the desired values in the observation plane and a vector of weights, indicating the relative importance of the target plane constraints. The optimal least squares solution can be obtained using standard approaches, such as the Moore-Penrose pseudo inverse. We have implemented the pseudo inverse-based optimization in Matlab, which produced the results in Figures 6-26 and 6-28 using a 306-band decomposition in approximately 10 seconds on an 8-core 3.2-GHz Intel-based machine, optimizing $M = 16$ complex-valued variables over $N = 802$ complex-valued weighted constraints in each band.

When working in environments that are highly acoustically reflective, it is helpful to limit the endfire energy propagation from the array, as this energy may be manifest at the locations of intended nulls as diffuse reflections. To obtain the results shown in Figures 6-26 and 6-28, we create a guard band to constrain the energy in the region just above and outside the support of the array. The guard band is implemented as an additional constraint plane represented in our \mathbf{H} matrix. Weights and target values are also chosen for the guard band. The \mathbf{H} matrix is populated using spherical wave propagation, as described above for the propagation plane, where d_o is replaced by d_g , the distance from the array to the guard plane. The black lines at 50cm, seen in the left of Figure 6-28 is a result of the guard band.

Chapter 7

Conclusion

In this thesis we have presented frameworks for compressive light field capture and compressive light field display. We have demonstrated prototypes of advanced displays, capable of creating glasses-free 3D effects, such as parallax and accommodation across a range of size scales (4D displays), and other prototypes capable of simultaneously measuring and emitting light field data within a desired view-cone (8D displays). These techniques can be adapted to other domains, such as audio. We have also demonstrated that recently developed dictionary based sparse light field reconstruction is applicable to new diffractive camera pixel structures such as angle sensitive pixels. These developments suggest the near-term viability of thin form-factor compressive 8D displays with the potential to revolutionize the way we interact with computation, the world, and one another.

Combining these frameworks with future electro-optical devices will enable a new class of advanced display, capable of addressing the human visual system in ways that equal or surpass the stimuli generated from physical light transport under some circumstances. The frameworks we have developed are broadly applicable, driving many form-factors and enabling diverse applications in many domains. The techniques presented will be the underpinnings of new forms of entertainment enabled by blurring the lines between rendered objects and the real world. As a fundamental interface technology the frameworks presented in this thesis have the potential to deeply impact fields that rely on data visualization—from

education, to medicine, to scientific research, to financial services, government and policy makers.

New optical systems will inspire new applications beyond human interaction. As optical systems are fundamental to a wide variety of scientific and industrial equipment, the frameworks presented herein will have broad impact going forward. Areas such as rapid fabrication, microscopy, medicine, communications, and transportation all rely deeply on the performance of optical systems. The core concept of this thesis, in abstract, is the application of formal optimization to the problem of light transport. While the methods developed around this idea may not directly apply to the above fields, certainly as the abundance of computation increases, optimization methods will become ever more integral to the optical and computational aspects of our technical endeavors.

Large challenges lie ahead. In order to develop advanced display systems further work will be required in electro-optics and algorithms. Creating computational hardware to accelerate the expanded computational requirements of such display systems will enable the optimization framework to tackle ever more challenging problems. Similarly, algorithmic improvements resulting in faster convergence to more optimal solutions will be required to achieve widespread commercial adoption of the type of advanced display systems presented in this thesis. It may be possible to overcome fundamental limitations of ray-based systems such as diffraction and narrow depth-of-field by extending the methods developed in this thesis to diffractive systems. Finally, the advanced displays presented herein, and their conceptual cousins, will necessitate the rethinking of core concepts in the field human-computer interaction as general purpose computation moves to these more information-rich platforms.

By considering computation as a fundamental aspect of optical information display and capture, we will ensure a bright future in display research and industry.

Bibliography

- [1] E. Adelson. On seeing stuff: The perception of materials by humans and machines. In *Proc. of the SPIE*, volume 4299, pages 1–12. Citeseer, 2001.
- [2] E. Adelson and J. Wang. Single lens stereo with a plenoptic camera. *IEEE Trans. PAMI*, 14(2):99–106, 1992.
- [3] T. Agocs, T. Balogh, T. Forgacs, F. Bettio, E. Gobbetti, G. Zanetti, and E. Bouvier. A large scale interactive holographic display. In *IEEE Virtual Reality*, pages 311–312, 2006.
- [4] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: Design of dictionaries for sparse representation. *Proceedings of SPARS*, 5:9–12, 2005.
- [5] K. Akeley, S. J. Watt, A. Girshick, and M. Banks. A stereo display prototype with multiple focal distances. *ACM Trans. Graph. (SIGGRAPH)*, 23:804–813, 2004.
- [6] Toshiba America. Toshiba america electronic components to showcase 3D and other innovative lcd technologies at SID 2011, May 16 2011.
- [7] A.H. Andersen and A.C. Kak. Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm. *Ultrasonic Imaging*, 6(1):81–94, 1984.
- [8] R. Anderson, W. Knowles, and J. Culley. Methods and apparatus involving light pen interaction with a real time display, April 12 1977. US Patent 4,017,680.
- [9] A. Ashok and M. Neifeld. Compressive light field imaging. In *SPIE Defense, Security, and Sensing*, pages 76900Q–76900Q. International Society for Optics and Photonics, 2010.
- [10] D. Babacan, R. Ansorge, M. Luessi, P. Ruiz, R. Molina, and A. Katsaggelos. Compressive light field sensing. 2012.
- [11] G. Bader, P. Ott, E. Lueder, and V. Schmid. Hybrid shape recognition system with microlens array processor and direct optical input. volume 3073, pages 277–287, 1997.
- [12] T. Balogh. The HoloVizio system. In *Proc. SPIE 6055*, volume 60550U, 2006.

- [13] T. Baran, D. Wei, and A.V. Oppenheim. Linear programming algorithms for sparse filter design. *IEEE Trans. Signal Processing*, 58(3):1605–1617, 2010.
- [14] R. Baraniuk. Compressive sensing [lecture notes]. *Signal Processing Magazine, IEEE*, 24(4):118–121, 2007.
- [15] L. Beiser. Anaglyph stereoscopy, September 22 1981. US Patent 4,290,675.
- [16] G. P. Bell, R. Craig, R. Paxton, G. Wong, and D. Galbraith. Beyond flat panels: multi-layered displays with real depth. *SID Digest*, 39(1), 2008.
- [17] S. Benton and V. M. Bove Jr. *Holographic imaging*. Wiley-Interscience, 2008.
- [18] T.E Bishop, S. Zanetti, and P. Favaro. Light field superresolution. In *Proc. ICCP*, pages 1–9. IEEE, 2009.
- [19] F. Blais. Review of 20 years of range sensor development. *Journal of Electronic Imaging*, 13(1), 2004.
- [20] V. Blondel, N. Ho, and P. van Dooren. Weighted nonnegative matrix factorization and face feature extraction. *Image and Vision Computing*, 2008.
- [21] V. Blondel, N. Ho, and P. van Dooren. Weighted nonnegative matrix factorization and face feature extraction. In *Image and Vision Computing*, pages 1–17, 2008.
- [22] L. Bogaert, Y. Meuret, S. Roelandt, A. Avci, H. De Smet, and H. Thienpont. Single projector multiview displays: Directional illumination compared to beam steering. In *Proc. SPIE 7524*, volume 75241R, 2010.
- [23] W. Bosking, Y. Zhang, B. Schofield, and D. Fitzpatrick. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *The Journal of Neuroscience*, 17(6):2112–2127, 1997.
- [24] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Matlab scripts for alternating direction method of multipliers. Technical report, Technical report, <http://www.stanford.edu/boyd/papers/admm>, 2012.
- [25] D. Brady, N. Pitsianis, and X. Sun. Reference structure tomography. *J. Opt. Soc. Am. A*, 21(7):1140–1147, 2004.
- [26] R. Brott and J. Schultz. Directional backlight lightguide considerations for full resolution autostereoscopic 3D displays. *SID Digest*, pages 218–221, 2010.
- [27] C. Brown, H. Kato, K. Maeda, and B. Hadwen. A continuous-grain silicon-system lcd with optical input function. *Solid-State Circuits, IEEE Journal of*, 42(12):2904–2912, dec. 2007.
- [28] M. Broxton, L. Grosenick, S. Yang, N. Cohen, A. Andalman, K. Deisseroth, and M. Levoy. Wave optics theory and 3-D deconvolution for the light field microscope. *Optics express*, 21(21):25418–25439, 2013.

- [29] G. Burdea and P. Coiffet. Virtual reality technology. *Presence: Teleoperators & Virtual Environments*, 12(6):663–664, 2003.
- [30] B. Cabral and L. C. Leedom. Imaging vector fields using line integral convolution. In *Proc. SIGGRAPH*, pages 263–270, 1993.
- [31] E. Candès, Y. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1):59–73, 2011.
- [32] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59:1207–1223, 2006.
- [33] J. Chai, X. Tong, S. Chan, and H. Shum. Plenoptic sampling. In *Proc. SIGGRAPH*, pages 307–318, 2000.
- [34] C. Chen, F. Lin, Y. Hsu, Y. Huang, and H. D. Shieh. A field sequential color LCD based on color fields arrangement for color breakup and flicker reduction. *Display Technology*, 5(1):34–39, 2009.
- [35] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1998.
- [36] K. Chien and H. Shieh. Time-multiplexed three-dimensional displays based on directional backlights with fast-switching liquid-crystal displays. *Applied Optics*, 45(13):3106–3110, 2006.
- [37] M. Chu, F. Diele, R. Plemmons, and S. Ragni. Optimality, computation, and interpretation of nonnegative matrix factorizations. *SIAM Journal on Matrix Analysis*, 2004.
- [38] Y. M. Chu, K. W. Chien, H. P. D. Shieh, J. M. Chang, Y. C. Shiu A. Hu, and V. Yang. 3D mobile display based on dual-directional light guides with a fast-switching liquid-crystal panel.
- [39] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Wiley, 2009.
- [40] T. F. Coleman and Y. Li. A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables. *SIAM Journal on Optimization*, (4):1040–1058, 1996.
- [41] O. Coles. NVIDIA GeForce 3D vision, January 9 2009.
- [42] O. Cossairt and G. Favalora. Minimized-thickness angular scanner of electromagnetic radiation, April 26 2006. US Patent App. 11/380,296.
- [43] O. Cossairt, J. Napoli, S. Hill, R. Dorval, and G. Favalora. Occlusion-capable multi-view volumetric three-dimensional display. *Applied Optics*, 46(8):1244–1250, 2007.

- [44] O. Cossairt, S. Nayar, and R. Ramamoorthi. Light field transfer: global illumination between real and synthetic objects. *ACM Trans. Graph. (TOG)*, 27(3):57, 2008.
- [45] H. Crane and T. Piantanida. On seeing reddish green and yellowish blue. *Science*, 221(4615):1078–1080, 1983.
- [46] T. Cuypers, S. B. Oh, T. Haber, P. Bekaert, and R. Raskar. WBSDF for simulating wave effects of light and audio. In *ACM SIGGRAPH 2010 Posters*, page 1. ACM, 2010.
- [47] M. Date, T. Hisaki, H. Takada, S. Suyama, and K. Nakazawa. Luminance addition of a stack of multidomain liquid-crystal displays and capability for depth-fused three-dimensional display application. *Applied Optics*, 44(6):898–905, 2005.
- [48] J. Davis, D. McNamara, D. Cottrell, and T. Sonehara. Two-dimensional polarization encoding with a phase-only liquid-crystal spatial light modulator. *Applied Optics*, 39(10):1549–1554, 2000.
- [49] N. A. Dodgson, J. R. Moore, S. R. Lang, G. Martin, and P. Canepa. A time-sequential multi-projector autostereoscopic display. *Journal of the SID*, 8(2):169–176, 2000.
- [50] D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
- [51] F. Durand, N. Holzschuch, C. Soler, E. Chan, and F. Sillion. A frequency analysis of light transport. *ACM Trans. Graph. (SIGGRAPH)*, 24(3):1115–1126, 2005.
- [52] J. Eichenlaub. Optical system which projects small volumetric images to very large size. In *Electronic Imaging 2005*, pages 313–322. International Society for Optics and Photonics, 2005.
- [53] H. Farid and E. Simoncelli. Range estimation by optical differentiation. *JOSA A*, 15(7):1777–1786, 1998.
- [54] G. Favalora. Volumetric 3D displays and application infrastructure. *IEEE Computer*, 38:37–44, 2005.
- [55] M. Feigin, D. Feldman, and N. Sochen. From high definition image to low space optimization. In *Scale Space and Variational Methods in Computer Vision*, pages 459–470. Springer, 2012.
- [56] R. W. Fleming, R. O. Dror, and E. Adelson. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, 3(5), 2003.
- [57] S. Follmer, M. Johnson, E. Adelson, and H. Ishii. deform: An interactive malleable surface for capturing 2.5d arbitrary objects, tools and touch. In *Proc. 24th ACM symp. on UIST*, pages 527–536. ACM, 2011.
- [58] M. Fuchs, R. Raskar, H. Seidel, and H. Lensch. Towards passive 6d reflectance field displays. *ACM Trans. Graph.*, 27(3):58:1–58:8, August 2008.

- [59] W. Funk. History of autostereoscopic cinema. In *Proc. SPIE 8288*, volume 82880R, 2012.
- [60] D. Gabor. Optical system composed of lenticules, June 13 1944. US Patent 2,351,034.
- [61] P. Gill, C. Lee, D. Lee, A. Wang, and A. Molnar. A microscale camera using direct fourier-domain scene capture. *Optics letters*, 36(15):2949–2951, 2011.
- [62] M. Goodale and A. D. Milner. Separate visual pathways for perception and action. *Trends in neurosciences*, 15(1):20–25, 1992.
- [63] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen. The lumigraph. In *Proc. SIGGRAPH*, pages 43–54, 1996.
- [64] H. Gotoda. A multilayer liquid crystal display for autostereoscopic 3D viewing. In *SPIE Stereoscopic Displays and Applications XXI*, volume 7524, pages 1–8, 2010.
- [65] H. Gotoda. Reduction of image blurring in an autostereoscopic multilayer liquid crystal display. In *SPIE Stereoscopic Displays and Applications XXII*, volume 7863, pages 1–7, 2011.
- [66] A. Greengard, Y. Schechner, and R. Piestun. Depth from diffracted rotation. *Optics letters*, 31(2):181–183, 2006.
- [67] M. Grosse, G. Wetzstein, A. Grundhöfer, and O. Bimber. Coded aperture projection. *ACM Trans. Graph.*, 29:22:1–22:12, 2010.
- [68] W. Gruber. Stereoscopic viewing device, January 20 1939. US Patent 2,189,285.
- [69] N. Hagood, R. Barton, T. Brosnihan, J. Fijol, J. Gandhi, M. Halfman, R. Payne, and J. L. Steyn. A direct-view MEMS display for mobile applications. *SID Symposium Digest of Technical Papers*, 38(1):1278–1281, 2007.
- [70] M. Halle. Holographic stereograms as discrete imaging systems. In *IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology*, pages 73–84. International Society for Optics and Photonics, 1994.
- [71] W. M. Hart. The temporal responsiveness of vision. In R. A. Moses and W. M. Hart, editors, *Adler’s Physiology of the Eye*. C.V. Moseby Company, 1987.
- [72] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [73] Eugene Hecht. *Optics (4th Edition)*. Addison Wesley, 2001.
- [74] S. Hecht and S. Schlaer. Intermittent stimulation by light v. the relation between intensity and critical frequency for different parts of the spectrum. *The Journal of general physiology*, 19(6):965–977, 1936.
- [75] D. Heeger. Perception lecture notes: Attention and awareness. <http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/attention/attention.html>. Accessed 2014-05-20.

- [76] F. Heide, G. Wetzstein, R. Raskar, and W. Heidrich. Adaptive image synthesis for compressive displays. *ACM Trans. Graph. (SIGGRAPH)*, 2013.
- [77] C. Hembd-Sölner, R. Stevens, and M. Hutley. Imaging Properties of the Gabor Superlens. *Journal of Optics A: Pure and Applied Optics*, 1(1):94, 1999.
- [78] G. Herman. Image reconstruction from projections. *Real-Time Imaging*, 1(1):3–18, 1995.
- [79] R. D. Hersch and S. Chosson. Band moiré images. *ACM Trans. Graph.*, 23(3):239–247, 2004.
- [80] M. Hirsch and D. Lanman. In *ACM SIGGRAPH ASIA 2010 Courses*, page 16. ACM, 2010.
- [81] M. Hirsch, D. Lanman, H. Holtzman, and R. Raskar. BiDi screen: A thin, depth-sensing lcd for 3D interaction using light fields. In *ACM Trans. Graph. (TOG)*, volume 28, page 159. ACM, 2009.
- [82] M. Hirsch, G. Wetzstein, and R. Raskar. A compressive light field projection system. *ACM Trans. Graph.*, 33(4):58:1–58:12, July 2014.
- [83] D. Hoffman and M. Banks. Stereo display with time-multiplexed focal adjustment. In *SPIE Stereoscopic Displays and Applications XX*, volume 7237, pages 1–8, 2009.
- [84] D. Hoffman and M. Banks. Focus information is used to interpret binocular images. *Journal of Vision*, 10(5):13, 2010.
- [85] D. Hoffman, A. Girshick, K. Akeley, and M. Banks. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision*, 8(3), 2008.
- [86] D. Hoffman, A. Girshick, K. Akeley, and M. Banks. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3):33, March 2008.
- [87] M. Holroyd, I. Baran, J. Lawrence, and W. Matusik. Computing and fabricating multilayer models. *ACM Trans. Graph. (SIGGRAPH Asia)*, 30:187:1–187:8, 2011.
- [88] J. Hong, Y. Kim, S. Park, J. Hong, S. Min, S. Lee, and B. Lee. 3D/2D convertible projection-type integral imaging using concave half mirror array. *Optics Express*, 18, 2010.
- [89] H. Hoshino, F. Okano, H. Isono, and I. Yuyama. Analysis of resolution limitation of integral photography. *J. Opt. Soc. Am. A*, 15(8):2059–2065, 1998.
- [90] F. Hsu. Three-dimensional (3D) image projection. US patent 7425070 B2, 2008.
- [91] M. Hullin, H. Lensch, R. Raskar, H. Seidel, and I. Ihrke. Dynamic display of brdfs. *Computer Graphics Forum*, 30(2):475–483, 2011.

- [92] M. C. Hutley, R. Hunt, R. F. Stevens, and P. Savander. The moiré magnifier. *Pure and Applied Optics: Journal of the European Optical Society Part A*, 3(2):133–142, 1994.
- [93] F. Ives. Parallax stereogram and process of making same. U.S. Patent 725,567, 1903.
- [94] H. Ives. Camera for making parallax panoramagrams. *J. Opt. Soc. Amer.*, 17:435–439, 1928.
- [95] S. Izadi, S. Hodges, S. Taylor, D. Rosenfeld, N. Villar, A. Butler, and J. Westhues. Going beyond the display: a surface technology with an electronically switchable diffuser. In *Proc. 21st ACM symp. on UIST*, pages 269–278. ACM, 2008.
- [96] A. Jacobs, J. Mather, R. Winlow, D. Montgomery, G. Jones, M. Willis, M. Tillin, L. Hill, M. Khazova, H. Stevenson, and G. Bourhill. 2D/3D switchable displays. *Sharp Technical Journal*, (4):1–5, 2003.
- [97] H. W. Jensen, S. Marschner, M. Levoy, and P. Hanrahan. A practical model for subsurface light transport. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '01*, pages 511–518, New York, NY, USA, 2001. ACM.
- [98] A. Jones, J. Liu, J. Busch, P. Debevec, M. Bolas, and X. Yu. An autostereoscopic projector array optimized for 3D facial display. SIGGRAPH Emerging Technologies, 2013.
- [99] A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec. Rendering for an interactive 360° light field display. *ACM Trans. Graph. (SIGGRAPH)*, 26:40:1–40:10, 2007.
- [100] R. C. Jones. A new calculus for the treatment of optical systems. *J. Opt. Soc. Am.*, 31(7):488–493, 1941.
- [101] H. Jorke and M. Fritz. Infitec-a new stereoscopic visualisation tool by wavelength multiplex imaging. *Proceedings of Electronic Displays*, 2003, 2003.
- [102] B. Julesz, T. Pappathomas, and F. Phillips. *Foundations of cyclopean perception*, volume 4. University of Chicago Press Chicago, 1971.
- [103] J. Jurik, A. Jones, M. Bolas, and P. Debevec. Prototyping a light field display involving direct observation of a video projector array. In *Proc. ProCams. IEEE*, 2011.
- [104] S. Kaczmarz. Angenäherte auflösung von systemen linearer gleichungen. *Bull. Acad. Pol. Sci. Lett. A*, 35:335–357, 1937.
- [105] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar. Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM Trans. Graph. (TOG)*, 32(6):167, 2013.

- [106] J. Kajiya. The rendering equation. In *ACM Siggraph Computer Graphics*, volume 20, pages 143–150. ACM, 1986.
- [107] A. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. Society for Industrial Mathematics, 2001.
- [108] M. H. Kamal, M. Golbabaee, and P. Vandergheynst. Light field compressive sensing in camera arrays. In *Proc. ICASSP*, pages 5413–5416, 2012.
- [109] B. Keck, H. Hofmann, H. Scherl, M. Kowarschik, and J. Hornegger. GPU-accelerated SART reconstruction using the CUDA programming environment. In *SPIE*, volume 7258, 2009.
- [110] Y. Kim, K. Hong, J. Yeom, J. Hong, J. Jung, Y. W. Lee, J. Park, and B. Lee. A frontal projection-type three-dimensional display. *Optics Express*, 20, 2012.
- [111] Y. Kim, J. Kim, J. Kang, J. Jung, H. Choi, and B. Lee. Point light source integral imaging with improved resolution and viewing angle by the use of electrically movable pinhole array. *Optics Express*, 15(26):18253–18267, 2007.
- [112] H. Kimura, T. Uchiyama, and H. Yoshikawa. Laser produced 3D display in the air. In *SIGGRAPH Emerging Technologies*, page 20. ACM, 2006.
- [113] M. Klug, M. Holzbach, and A. Ferdman. Method and apparatus for recording one-step, full-color, full-parallax, holographic stereograms, December 11 2001. US Patent 6,330,088.
- [114] T. Kolda and B. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.
- [115] H. Kwon and H. Choi. A time-sequential multiview autostereoscopic display without resolution loss using a multi-directional backlight unit and an LCD panel. In *SPIE Stereoscopic Displays and Applications XXIII*, volume 8288, pages 1–6, 2012.
- [116] Stanford Computer Graphics Laboratory. The Stanford light field archive. <http://lightfield.stanford.edu>, 2008.
- [117] M. Land and D. Nilsson. *Animal eyes*. Oxford University Press, 2012.
- [118] D. Lanman and M. Hirsch. Build your own glasses-free 3D display. In *ACM SIGGRAPH 2011 Courses*, page 4. ACM, 2011.
- [119] D. Lanman, M. Hirsch, Y. Kim, and R. Raskar. Content-adaptive parallax barriers: Optimizing dual-layer 3D displays using low-rank light field factorization. *ACM Trans. Graph. (SIGGRAPH Asia)*, 29:163:1–163:10, 2010.
- [120] D. Lanman, R. Raskar, A. Agrawal, and G. Taubin. Shield fields: modeling and capturing 3D occluders. In *ACM Trans. Graph. (SIGGRAPH)*, volume 27, page 131, 2008.

- [121] D. Lanman, G. Wetzstein, M. Hirsch, W. Heidrich, and R. Raskar. Polarization fields: Dynamic light field display using multi-layer LCDs. *ACM Trans. Graph. (SIGGRAPH Asia)*, 3:1–9, 2011.
- [122] D. Lee and S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.
- [123] J. Leitner, J. Powell, P. Brandl, T. Seifried, M. Haller, B. Dorray, and P. To. Flux: A tilting multi-touch and pen based surface. In *Proc. 27th intl. conf. extended abstracts on Human factors in computing systems*, pages 3211–3216. ACM, 2009.
- [124] A. Levin, R. Fergus, F. Durand, and W. Freeman. Image and depth from a conventional camera with a coded aperture. In *ACM Trans. Graph. (TOG)*, volume 26, page 70. ACM, 2007.
- [125] A. Levin, W. Freeman, and F. Durand. Understanding camera trade-offs through a bayesian analysis of light field projections. In *Computer Vision–ECCV 2008*, pages 88–101. Springer, 2008.
- [126] M. Levoy and P. Hanrahan. Light field rendering. In *ACM SIGGRAPH*, pages 31–42, 1996.
- [127] C. Liang, T. Lin, B. Wong, C. Liu, and H. Chen. Programmable aperture photography: multiplexed light field acquisition. In *ACM Trans. Graph. (SIGGRAPH)*, volume 27, page 55, 2008.
- [128] G. Lippmann. Épreuves réversibles donnant la sensation du relief. *Journal of Physics*, 7(4):821–825, 1908.
- [129] L. Lipton. Stereoscopic motion picture projection system, January 2 1996. US Patent 5,481,321.
- [130] C. Lu, S. Muenzel, and J. Fleischer. High-resolution light-field microscopy. In *Proc. OSA COSI*, 2013.
- [131] M. Lucente. Electronic holographic displays–20 years of interactive spatial imaging. *Handbook of Visual Display Technology*, pages 1963–1978, 2012.
- [132] A. Lumsdaine and T. Georgiev. The focused plenoptic camera. In *Proc. ICCP*, pages 1–8, 2009.
- [133] B. Ma, B. Yao, T. Ye, and M. Lei. Prediction of optical modulation properties of twisted-nematic liquid-crystal display by improved measurement of Jones matrix. *Appl. Physics*, 107, 2010.
- [134] A. Maimone, G. Wetzstein, M. Hirsch, D. Lanman, R. Raskar, and H. Fuchs. Focus 3D: compressive accommodation display. *ACM Trans. Graph. (TOG)*, 32(5):153, 2013.
- [135] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar. Compressive light field photography using overcomplete dictionaries and optimized projections. *ACM Trans. Graph.*, 32(4), 2013.

- [136] J. Mather, N. Barratt, D. Kean, E. Walton, and G. Bourhill. Directional backlight, a multiple view display and a multi-direction display. U.S. Patent Application 11/814,383, 2009.
- [137] W. Matusik and H. Pfister. 3D TV: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. *ACM Trans. Graph. (SIGGRAPH)*, 23:814–824, 2004.
- [138] R. McIntosh and T. Schenk. Two visual streams for perception and action: Current trends. *Neuropsychologia*, 47(6):1391–1396, 2009.
- [139] S. McQuaide, E. Seibel, R. Burstein, and T. Furness. Three-dimensional virtual retinal display system using a deformable membrane mirror. In *SID Symposium Digest of Technical Papers*, volume 33, pages 1324–1327. Wiley Online Library, 2002.
- [140] J. Messnerl, S. Yerrapathrunil, A. Baratta, and D. Riley. Cost and schedule reduction of nuclear power plant construction using 4d cad and immersive display technologies. volume 2002, pages 136–144, 2002.
- [141] Y. Meuret, L. Bogaert, S. Roelandt, J. Vanderheijden, A. Avci, H. De Smet, and H. Thienpont. LED projection architectures for stereoscopic and multiview 3D displays. In *Proc. SPIE 7690*, volume 769007, 2010.
- [142] I. Moreno, J. L. Martínez, and J. A. Davis. Two-dimensional polarization rotator using a twisted-nematic liquid-crystal display. *Applied Optics*, 46(6):881–887, 2007.
- [143] I. Moreno, P. Velásquez, C. R. Fernández-Pousa, M. M. Sánchez-López, and F. Mateos. Jones matrix method for predicting and optimizing the optical modulation properties of a liquid-crystal display. *Appl. Physics*, 94, 2003.
- [144] Y. Munz, K. Moorthy, A. Dosis, J.D. Hernandez, S. Bann, F. Bello, S. Martin, A. Darzi, and T. Rockall. The benefits of stereoscopic vision in robotic-assisted performance on bench models. *Surgical Endoscopy And Other Interventional Techniques*, 18(4):611–616, 2004.
- [145] S. Nayar, P. Belhumeur, and T. Boult. Lighting sensitive display. *ACM Trans. Grap.*, 23(4):963–979, 2004.
- [146] S. Nayar and Y. Nakagawa. Shape from focus. *IEEE Trans. Pattern analysis and machine intelligence*, 16(8):824–831, 1994.
- [147] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2:11, 2005.
- [148] J. Nims and A. Lo. 3-D screen and system. US patent 3,814,513, 1972.
- [149] K. Oh, S. Yea, and Y. Ho. Hole filling method using depth based in-painting for view synthesis in free viewpoint television and 3-d video. In *Picture Coding Symposium, 2009. PCS 2009*, pages 1–4. IEEE, 2009.

- [150] S. B. Oh, S. Kashyap, R. Garg, S. Chandran, and R. Raskar. Rendering wave effects with augmented light field. In *Computer Graphics Forum*, volume 29, pages 507–516. Wiley Online Library, 2010.
- [151] G. Oster, M. Wasserman, and C. Zwierling. Theoretical interpretation of moiré patterns. *JOSA*, 54(2):169–175, 1964.
- [152] M. O’Toole and K. Kutulakos. Optical computing for fast light transport analysis. *ACM Trans. Graph.*, 29(6):164, 2010.
- [153] M. O’Toole, R. Raskar, and K. Kutulakos. Primal-dual coding to probe light transport. *ACM Trans. Graph.*, 31(4):39, 2012.
- [154] V. Pamplona, M. Oliveira, D. Aliaga, and R. Raskar. Tailored displays to compensate for visual aberrations. *ACM Trans. Graph.*, 31(4):81, 2012.
- [155] Y. Pan, X. Xu, and X. Liang. Fast distributed large-pixel-count hologram computation using a gpu cluster. *Applied optics*, 52(26):6562–6571, 2013.
- [156] K. Perlin, S. Paxia, and J. Kollin. An autostereoscopic display. In *Proc. SIGGRAPH*, pages 319–326, 2000.
- [157] Persistence of Vision Pty. Ltd. Persistence of vision raytracer (version 3.6). <http://www.povray.org>, 2004.
- [158] C. Perwass and L. Wietzke. Single lens 3D-camera with extended depth-of-field. In *Proc. SPIE 8291*, pages 29–36, 2012.
- [159] T. Peterka, R. Kooima, D. Sandin, A. Johnson, J. Leigh, and T. DeFanti. Advances in the Dynallax solid-state dynamic parallax barrier autostereoscopic visualization display system. *IEEE TVCG*, 14(3):487–499, 2008.
- [160] N. Peyghambarian, S. Tay, P. Blanche, R. Norwood, and M. Yamamoto. Rewritable holographic 3D displays. *Optics and Photonics News*, 19(7):22–27, 2008.
- [161] R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried. Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045):1102–1107, 2005.
- [162] A. Roorda, A. Metha, P. Lennie, and D. Williams. Packing arrangement of the three cone classes in primate retina. *Vision research*, 41(10):1291–1306, 2001.
- [163] A. Roorda and D. Williams. The arrangement of the three cone classes in the living human eye. *Nature*, 397(6719):520–522, 1999.
- [164] E. Rossi, P. Weiser, J. Tarrant, and A. Roorda. Visual performance in emmetropia and low myopia after correction of high-order aberrations. *Journal of Vision*, 7(8):14, 2007.
- [165] T. Sato, H. Mamiya, H. Koike, and K. Fukuchi. Photoelasticitytouch: Transparent rubbery tangible interface using an lcd and photoelasticity. In *Proc. 22nd ACM symp. on UIST*, pages 43–50. ACM, 2009.

- [166] A. Schwerdtner. Video hologram and device for reconstructing video holograms for large objects, January 1 2008. US Patent 7,315,408.
- [167] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Computer vision and pattern recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 519–528. IEEE, 2006.
- [168] P. Sen, B. Chen, G. Garg, S. Marschner, M. Horowitz, M. Levoy, and H. Lensch. Dual photography. *ACM Trans. Graph. (TOG)*, 24(3):745–755, 2005.
- [169] P. M. Shankar, W. C. Hasenplaugh, R. L. Morrison, R. A. Stack, and M. A. Neifeld. Multiaperture imaging. *Appl. Opt.*, 45(13):2871–2883, 2006.
- [170] T. Shibata, T. Kawai, K. Ohta, M. Otsuki, N. Miyake, Y. Yoshihara, and T. Iwasaki. Stereoscopic 3-D display with optical correction for the reduction of the discrepancy between accommodation and convergence. *SID*, 13(8):665–671, 2005.
- [171] T. Sielhorst, M. Feuerstein, and N. Navab. Advanced medical displays: A literature review of augmented reality. *Display Technology, Journal of*, 4(4):451–467, 2008.
- [172] S. Sivaramakrishnan, A. Wang, P. Gill, and A. Molnar. Enhanced angle sensitive pixels for light field imaging. In *Proc. IEEE IEDM*, pages 8–6, 2011.
- [173] C. Slinger, C. Cameron, and M. Stanley. Computer-generated holography as a generic display technology. *IEEE Computer*, 38(8):46–53, 2005.
- [174] D. Smalley, Q. Smithwick, V. Bove, J. Barabas, and S. Jolly. Anisotropic leaky-mode modulator for holographic video displays. *Nature*, 498(7454):313–317, 2013.
- [175] L. Smoot, Q. Smithwick, and D. Reetz. A volumetric display based on a rim-driven varifocal beamsplitter and LED backlit LCD. In *SIGGRAPH Emerging Technologies*, page 22. ACM, 2011.
- [176] N. Srebro and T. Jaakkola. Weighted low-rank approximations. In *ICML*, pages 720–727, 2003.
- [177] R. Stewart and W. Roach. Field-sequential display system utilizing a backlit LCD pixel array and method for forming an image. U.S. Patent 5,337,068, 1994.
- [178] H. Stolle, J. Olaya, S. Buschbeck, H. Sahm, and A. Schwerdtner. Technical solutions for a full-resolution autostereoscopic 2D/3D display technology. In *Proc. SPIE*, pages 1–12, 2008.
- [179] A. Sullivan. A solid-state multi-planar volumetric display. In *SID Digest*, volume 32, pages 207–211, 2003.
- [180] Y. Takaki. High-density directional display for generating natural three-dimensional images. *Proc. IEEE*, 94(3), 2006.

- [181] Y. Takaki, K. Tanaka, and J. Nakamura. Super multi-view display with a lower resolution flat-panel display. *Opt. Express*, 19(5):4129–4139, 2011.
- [182] K. Tanaka, J. Hayashi, M. Inami, and S. Tachi. Twister: An immersive autostereoscopic display. In *Virtual Reality, 2004. Proceedings. IEEE*, pages 59–278. IEEE, 2004.
- [183] J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka. Thin observation module by bound optics (TOMBO): Concept and experimental verification. *Appl. Opt.*, 40(11):1806–1813, 2001.
- [184] J. Tompkin, S. Muff, S. Jakushevskij, J. McCann, J. Kautz, M. Alexa, and W. Matusik. Interactive light field painting. In *ACM SIGGRAPH 2012 Emerging Technologies*, page 12. ACM, 2012.
- [185] R. Tootell, E. Switkes, M. S. Silverman, and S. L. Hamilton. Functional anatomy of the macaque striate cortex. ii. retinotopic organization. *Journal of Neuroscience*, 8(5):1531–1568, 1988.
- [186] K. Toyooka, T. Miyashita, and T. Uchida. The 3D display using field-sequential LCD with light direction controlling backlight. *SID Digest*, pages 177–180, 2001.
- [187] A. Travis. Autostereoscopic 3-D display. *Applied Optics*, 29(29):4341–4342, 1990.
- [188] A. Travis, T. Large, N. Emerton, and S. Bathiche. Collimated light from a waveguide for a display backlight. *Optics Express*, 17(22):19714–19719, 2009.
- [189] A. Travis, T. Large, N. Emerton, and S. Bathiche. Wedge optics in flat panel displays. *Proceedings of the IEEE*, 101(1):45–60, 2013.
- [190] A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, 1980.
- [191] T. Turner and R. Hellbaum. Lc shutter glasses provide 3-D display for simulated flight. *Inf. Disp.*, 2(9):22–24, September 1986.
- [192] R. Vaillant and O. Faugeras. Using extremal boundaries for 3-D object modeling. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(2):157–173, 1992.
- [193] H. L. Van Trees. *Optimum array processing*. Wiley Online Library, 2002.
- [194] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph.*, 26(3):69, 2007.
- [195] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. Bawendi, and R. Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, 3:745, 2012.

- [196] K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar. Picam: an ultra-thin high performance monolithic camera array. *ACM Trans. Graph. (SIGGRAPH Asia)*, 32(6):166, 2013.
- [197] A. Wang, P. Gill, and A. Molnar. Light field image sensors based on the talbot effect. *Applied optics*, 48(31):5897–5905, 2009.
- [198] A. Wang, P. Gill, and A. Molnar. An angle-sensitive CMOS imager for single-sensor 3D photography. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, pages 412–414. IEEE, 2011.
- [199] A. Wang, S. Sivaramakrishnan, and A. Molnar. A 180nm CMOS image sensor with on-chip optoelectronic image compression. In *Proc. IEEE Custom Integrated Circuits Conference (CICC)*, pages 1–4, 2012.
- [200] S. Wanner and B. Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE Trans. PAMI*, 2013.
- [201] G. Ward. Measuring and modeling anisotropic reflection. *ACM SIGGRAPH Computer Graphics*, 26(2):265–272, 1992.
- [202] C. Ware. *Information visualization: perception for design*. Elsevier, 2013.
- [203] G. Wendt, F. Faul, and R. Mausfeld. Highlight disparity contributes to the authenticity and strength of perceived glossiness. *Journal of Vision*, 8(1):14, 2008.
- [204] G. Wetzstein, I. Ihrke, and W. Heidrich. On plenoptic multiplexing and reconstruction. *IJCV*, 101:384–400, 2013.
- [205] G. Wetzstein, D. Lanman, W. Heidrich, and R. Raskar. Layered 3D: Tomographic image synthesis for attenuation-based light field and high dynamic range displays. *ACM Trans. Graph. (SIGGRAPH)*, 30:1–11, 2011.
- [206] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar. Tensor Displays: Compressive light field synthesis using multilayer displays with directional backlighting. *ACM Trans. Graph. (SIGGRAPH)*, 31(4):80, 2012.
- [207] C. Wheatstone. Contributions to the physiology of vision. part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Trans. of the Royal Soc. of London*, 128:371–394, January 1838.
- [208] A. Wichansky. User benefits of visualization with 3-D stereoscopic displays. volume 1457, pages 267–271, 1991.
- [209] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. High performance imaging using large camera arrays. *ACM Trans. Graph. (SIGGRAPH)*, 24(3):765–776, 2005.
- [210] A. J. Woods and A. Sehic. The compatibility of LCD TVs with time-sequential stereoscopic 3D visualization. In *SPIE Stereoscopic Displays and Applications XX*, 2009.

- [211] X. Xu, S. Solanki, X. Liang, Y. Pan, and T. Chong. Full high-definition digital 3D holographic display and its enabling technologies. In *Optical Data Storage 2010*, pages 77301C–77301C. International Society for Optics and Photonics, 2010.
- [212] Z. Xu and E. Lam. A high-resolution lightfield camera with dual-mask design. In *SPIE Optical Engineering+Applications*, pages 85000U–85000U, 2012.
- [213] R. Yang, X. Huang, S. Li, and C. Jaynes. Toward the light field display: Autostereoscopic rendering via a cluster of projectors. *IEEE TVCG*, 14(1):84–96, 2008.
- [214] P. Yeh and C. Gu. *Optics of Liquid Crystal Displays*. John Wiley and Sons, 2009.
- [215] M. Yoneyama, J. Fujimoto, Y. Kawamo, and S. Sasabe. The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design. *J. Acoust. Soc. Am*, 73(5):1532–1536, 1983.
- [216] T. Zhang, B. Fang, W. Liu, Y. Y. Tang, G. He, and J. Wen. Total variation norm-based nonnegative matrix factorization for identifying discriminant representation of image patterns. *Neurocomputing*, 71(10–12):1824–1831, 2008.
- [217] Z. Zhang and M. Levoy. Wigner distributions and how they relate to the light field. In *Computational Photography (ICCP), 2009 IEEE International Conference on*, pages 1–10. IEEE, 2010.
- [218] R. Zone. *Stereoscopic Cinema & the Origins of 3-D Film, 1838-1952*. University Press of Kentucky, 2007.
- [219] M. Zwicker, W. Matusik, F. Durand, and H. Pfister. Antialiasing for automultiscopic 3D displays. In *EGSR*, 2006.
- [220] M. Zwicker, A. Vetro, Sehoon Yea, W. Matusik, H. Pfister, and F. Durand. Resampling, antialiasing, and compression in multiview 3-D displays. *IEEE Signal Processing Magazine*, 24(6):88–96, 2007.