Holographic Illumination for Computer Vision

Dorian Chan

CMU-CS-25-118
May 2025

Computer Science Department
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Matthew O'Toole, Chair
Ioannis Gkioulekas
Aswin Sankaranarayanan
Mohit Gupta (University of Wisconsin–Madison)

*Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy.*

This research was sponsored by: the National Science Foundation under award numbers IIS2107236 and IIS2238485; the Lockheed Martin Corporation under award number MRA19001RPS008; Phlux Technologies under award number 1990821; and Snap Inc.

Keywords: Holographic illumination, active sensing, computer vision

For my family

Abstract

Artificial illumination is ubiquitous in real vision systems. By coding extra information from a controlled light source into the images captured by a camera, so-called "active sensing" approaches robustly capture depth, reflectance and other visual cues crucial to tasks in robotics, manufacturing, consumer products and more. However, active sensors struggle with well-known challenges that lessen their practicality in modern systems. First, limited power in portable devices restricts range and outdoor performance of active sensors. Second, the slow speed of many active sensors precludes their usage for dynamic scenes. Finally, the lack of depth programmability in today's illumination sources reduces effective resolution in scenes with significant depth variation, and shrinks the space of potential future applications.

To tackle these challenges, this thesis explores using *holographic illumination*. Similar holographic systems have recently seen significant attention as displays in the augmented and virtual reality (AR/VR) literature. By combining a spatial-light modulator (SLM) with laser light, such devices can replicate natural 3D visual cues in a compact form factor, key aspects that are currently missing in modern AR/VR architectures.

In our work, we analyze how they can potentially be adapted as sources of active illumination. First, we show how holographic illumination can be used to build light-redistributive systems that allow for smarter energy usage in active sensing, enabling time-of-flight sensors with far-improved dynamic range. Next, we demonstrate how this light redistribution, when combined with the underlying fast speed of modern SLMs, allows for far faster projector systems, allowing for new types of triangulation light curtains. Finally, we test how the inherent coherent propagation of holographic illumination can be used to program meaningful, distinct content at multiple depths, enabling new user interfaces and depth-sensing methodologies.

Acknowledgments

First and foremost, I would like to thank my advisor Matt for all of his wonderful mentorship and feedback over the years. His research insights, his composure and more have meant so much to me throughout my PhD, and I will always appreciate that he took the time to listen to my stupid ideas. I'm also grateful for all of the advice, feedback and opportunities I received throughout this entire process from my committee members Yannis, Aswin and Mohit.

I would additionally like to thank all of my other collaborators during my PhD. Chapter 4 would not have been possible without Srinivasa Narasimhan's sage advice, and our work in optical vibration sensing over the years has been a real treat. I also spent an amazing summer with Jian Wang and Sizhuo Ma working on Chapter 5, who showed me what it means to be a mature researcher. Although our work did not quite make it into this thesis, Mark Sheinin has been an absolute role model for me — he taught me so much about the actual process of effective, impactful research. My work with Grace Kuo, Oliver Cossairt and Nathan Matsuda was also an incredibly rewarding experience — I look back fondly on the collaborative environment they fostered. And of course, thanks to Mariko Isogawa and Tianyuan Zhang for being awesome coauthors, whose dedication inspired me.

Furthermore, I want to acknowledge all of my "almost collaborators" — people who I never officially worked with, but shaped the way I think about computational photography today: Benjamin Attal, Yingsi Qin, Arjun Teh, Florian Schiffers, Bailey Miller, Sreekar Sai Ranganathan, Sally Chen, Arkadeep Narayan Chaudhury, Pradyumna Chari, Anthony Bisulco, Cindy Nguyen, Esther Lin, and Berthy Feng. I must also thank all of my other friends at CMU, New York, Seattle, and back home who supported and encouraged me throughout my PhD — there are too many to list here, but they know who they are.

Finally, I want to thank my family for their unwavering support. It's been quite a journey to get here, but they had my back the entire time.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A camera records light from a scene and converts it into a digital image, by using a lens with a sensor composed of millions of light-sensitive pixels. Each point in the output image measures the amount of incident light that arrives from a particular direction [127]. While cameras today are effective and ubiquitous, present in nearly every mobile device and computer, they do not record all the information that may be of interest for computer vision tasks. For example, many applications may require estimates of how far away every point in the output image is (also known as "depth"), rather than just the amount of light. Other applications may require accurate estimates of object texture independent of lights and shadows in the captured scene (also known as "reflectance"), rather than how an object currently looks from the point-of-view of the camera. A camera cannot directly capture any of this desired information.

To fill this gap, "active sensing" aims to encode more information into the images captured by a camera by introducing a controlled light source. For instance, perhaps the simplest instantiation of this idea is flash photography [42], where a bulb is flashed on during the camera exposure to better capture poorly-lit objects. Time-of-flight and laser ranging (LiDAR) systems measure how long emitted light takes to return to the camera, from which depth can be computed from the speed of light [72]. In photometric stereo, object normals are estimated by analyzing the change in brightness as a light source is moved to different angles [2]. In structured light, the distortion of a projected pattern is captured by a camera, from which object shape can be estimated [45]. These are just a couple of the ways in which artificial illumination from a controlled source is now leveraged in computer vision.

Outside the lab, such active sensing systems have found significant real-world use today. Nearly every mobile device provides flash photography functionality [42], and many now integrate structured light and time-of-flight sensors for biometric authentication and augmented reality [10, 39]. Autonomous vehicles and robots use LiDAR and structured light for precise depth measurements to help them navigate around their environment [137]. In manufacturing, photometric stereo and structured light are widely used on assembly lines for inspection purposes [3, 121].

Despite its ongoing success, active sensing is associated with well-known challenges that limit performance and practicality in the real world, as delineated below:

Brightness. Nearly every active sensor struggles with brightness. One reason is that all cameras will fundamentally record some amount of noise due to the photon nature of light as well as sensor non-idealities [56]. Thus, the controlled source must be sufficiently bright so that the sensor can resolve its contribution over this noise. This requirement can cause a number of problems in practice. For one, an object reflects different amounts of light depending on how far away it is from the light source — this variation can be calculated to be given by $1/R^2$, where R is the distance from the source [122]. In practice, this "falloff" imposes an effective maximum on the range of an active system, as an object that is too far away may reflect too little light. This problem is exacerbated in darker-colored objects, as they may reflect so little light that they can only be effectively measured when they are very close to the camera.

Secondly, outside the lab, the camera will also necessarily record some "ambient" light from other sources in the scene, which will introduce its own noise into the final measurement. Thus, the controlled source must also be bright enough to not become overwhelmed by this extra noise [55]. As a result, active sensors typically struggle outdoors, as bright sunlight can effectively wash out any artificial illumination in the captured images.

Unfortunately, simply using a brighter light source to mitigate these challenges is typically a non-trivial upgrade. Brighter sources use more power, which is often a luxury in modern standalone mobile systems with limited battery life. Simultaneously, thanks to the laws of thermodynamics, brighter sources produce more heat, which needs to either be dissipated via bulky cooling systems, or engineered for via heat resistant circuits [24, 100]. This results in added cost, complexity and form factor. Furthermore, on the theoretical end, increasing the power has diminishing returns, *i.e.*, if power is increased by N, range and SNR only increase by a factor $\sqrt{N}$. Finally, many camera sensors possess some capacity in terms of how much light can be recorded per pixel beyond which information is lost. Thus, a more powerful source can effectively saturate this capacity for any nearby, brighter objects that already reflect a significant amount of light, resulting in reduced overall performance. Thus, a brighter light source is typically not a desirable solution.

Speed. A related challenge is the speed of active sensors, limiting practicality for moving objects and dynamic scenes. One reason why active sensors are slow is again thanks to challenges with brightness, where longer or multiple camera exposures are needed per frame to adequately resolve coded illumination [55]. This translates to slower overall framerates. Another reason is that many active sensing techniques rely on captures of the scene under different illumination conditions (*e.g.*, structured light or photometric stereo), and thus programmable illumination hardware is required. Unfortunately, such devices are often slow compared to modern sensors. For example, the liquid-crystal display projectors commonly used for structured light typically operate at $60\,\mathrm{Hz}$ [97], while modern quanta [126], coded-exposure [138] and event devices [90] provide kilohertz sensing rates. Brightness challenges aside, slow illumination hardware can therefore become the bottleneck in real active sensing.

Depth programmability. Many active sensing techniques rely on projector systems to illuminate scenes with structured patterns. Unfortunately, most modern projector architectures fundamentally produce a sharp image at a single plane and defocused versions elsewhere [144]. Thus, such systems lack depth programmability — the ability to simultaneously program meaningful content at multiple depths. A projector with this capability could unlock many new applications that are currently impractical for existing systems. For instance, such a system could be applied to scenes with significant depth variation, where traditional projector architectures produce blurry output [53]. More broadly, it could enable new modes of projection mapping or user interfaces where visible content changes as a function of depth, creating a new dimension of human-computer interaction. It could also be used as a depth cue, where the visible component of the depth-dependent pattern is used to estimate distance .

In this thesis, we explore *holographic illumination* as a potential panacea for the aforementioned problems. Such holographic systems have primarily been explored in the context of next-generation displays for AR/VR headsets or consumer projectors, but have seen little-to-no attention in the context of computer vision. Such devices are based on using coherent laser illumination with a spatial-light modulator (SLM) to fundamentally steer, rather than block, light to create desired patterns. This modality makes them an excellent tool for building brighter, faster illumination systems that are potentially depth programmable. To this end, in this thesis:

- In Chapter 2, we review the physics behind holographic illumination, how to apply them in practice, and the recent applications of similar systems outside computer vision.
- In Chapter 3, we analyze how the inherent light redistribution of these devices allows for far more light-efficient active sensing systems, which we demon-

strate in the context of time-of-flight cameras.

- In Chapter 4, we show that when combined with modern fast SLMs, holographic illumination can enable high-speed structured light, which we use to create new types of triangulation light curtains.
- In Chapter 5, we show that holographic illumination can produce depth-varying content, which we use to remedy defocus, create novel interfaces, and perform a new modality of depth sensing.
- Finally, in Chapter 6, we discuss various limitations with holographic illumination and potential future directions.

Chapter 2

A brief introduction to holographic illumination for vision/graphics researchers

In computer graphics and vision, we typically think of light in terms of photons that fly through space in straight lines [127]. We also assume these photons combine additively [127], *i.e.*, if a point is illuminated by two sources of light, it receives the sum of the energy from both sources. Such *geometric* models of light currently govern the design of modern systems, *e.g.*, the ray tracers we use to simulate virtual scenes, the reconstruction models we use to digitize the real world, or the light sources we use for active sensing. However, as Thomas Young and Francois Arago famously showed in the early 19th century, light can also behave like a wave. Thus, light bends around corners, and can interfere constructively or destructively, *i.e.*, two identical lights adding together can result in four times the brightness of a single source or zero brightness! Typically, such phenomena are visually insignificant in most scenarios, as the broadband light that dominates the physical world commonly washes out such effects. However, when using monochromatic laser light, such *diffraction* effects become prominent. For instance, in the famous double-slit experiment, illuminating two slits in a sheet with a laser results in a prominent "fringe" pattern on the other side, where light bends through the slits and then combines constructively or destructively as a function of position (Figure 2.1).

More concretely, when using a laser, light can be shown to satisfy the so-called Helmholtz equation [48] under certain assumptions on the medium of propagation:

$$(\nabla^2 + k^2)U = 0, \tag{2.1}$$

where $k = \frac{2\pi}{\lambda}$, with λ equal to the wavelength of light, and $U : (x, y, z) \rightarrow \mathbb{C}$ represents the so-called scalar field. A sensor measures the squared magnitude of

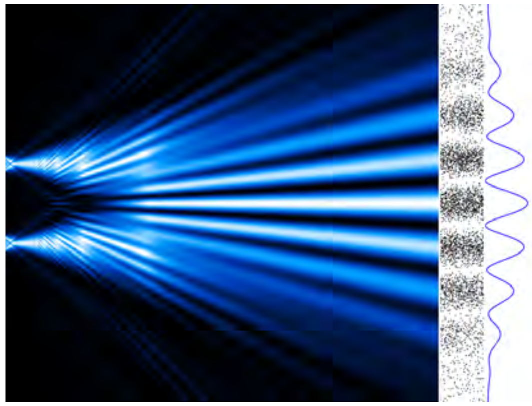


Figure 2.1: **Double-slit experiment.** Illuminating two slits with a laser results in a prominent fringe pattern on the other side, where light bends around the slits and adds together constructively or destructively. Image from Wikipedia under CC BY-SA 4.0.

this scalar field $|U|^2$.

Under certain simplifying assumptions, it can be shown that U at any point in space can be computed in closed form given U on a 2D plane via a "diffraction integral" [47]. This is the underlying principle of "holography", originally proposed by Dennis Gabor in 1948 [43]. In short, by properly encoding U over a 2D plane by using an optical element, 3D images of a desired scene can be produced when this so-called hologram is illuminated by laser light. Traditionally, such holograms were created by illuminating the target scene with a laser, and then interfering the scattered reflecting light with a reference beam onto a recording medium [43]. To lift the requirement of an initial recording step, Brown and Lohmann proposed the concept of computer-generated holography in the 1960s [17, 18], where a computer is used instead to compute holograms which are then manufactured by a plotter. Rather than manufacturing static holograms, the calculated holograms can also be displayed on programmable elements, which we will henceforth refer to as a holographic system for the purposes of this thesis.

In their simplest form, holographic systems consist of a laser source and a spatial-light modulator (SLM) that can programmably modulate either the phase or the amplitude of the scalar field in a spatially-varying fashion. Let $\mathcal{S}(\phi)$ represent the modulation imposed by the SLM given SLM pattern ϕ, *i.e.*, $\mathcal{S}_{\text{phase}}(\phi) = e^{j\phi}$ for a phase SLM, and $\mathcal{S}_{\text{amplitude}}(\phi) = \phi$ for an amplitude SLM. Then, given some diffraction integral/forward model $\mathcal{P}(\cdot)$ for wave propagation between the SLM plane and the target, displaying some target pattern T involves solving the following optimization problem:

$$\boldsymbol{\phi}^* = \underset{\boldsymbol{\phi}}{\operatorname{argmin}} \left\| T - |\mathcal{P}(\mathcal{S}(\phi))|^2 \right\|^2 . \tag{2.2}$$

Displaying the resulting $\boldsymbol{\phi}^*$ on the SLM then produces a close fascimile to T. The requisite SLM can be implemented with various technologies, including liquid-crystal displays (LCD), liquid crystal on silicon (LCoS), digital mirror devices (DMD), and programmable metasurfaces [79].

Historically, such devices have been explored in the context of human-facing displays (colloquially, *holographic displays*) [32, 50, 69, 70, 84, 104] and less frequently, consumer projectors [21, 22, 30, 86, 87, 118]. In this thesis, we argue that such holographic systems may also be useful in the context of computer vision, as illumination sources for active sensing. This entails both their application as a point illumination source, like those used for time-of-flight sensors as we explore in Chapter 3, or as a projector like those used for structured light or user interfaces as we discuss in Chapter 4 and Chapter 5. We term holographic systems when used in this context as *holographic illumination*.

Before we dive into the applications of this holographic illumination in the rest of this thesis, in the rest of this chapter, we discuss technical preliminaries of holographic systems at large that may make them powerful tools in practice. We finish by discussing the calibration process for these types of systems, and computational approaches towards solving Equation (2.2).

2.1 Capabilities of holographic systems

With the above ideas in mind, why should one use a holographic approach over existing architectures in displays, projectors and illumination systems? In this section, we discuss two potential key reasons.

2.1.1 Holographic systems and lens-free 3D content

Theoretically, holographic systems can produce content at any distance from the SLM, all without a lens. In particular, the wavefront at a plane d away from the SLM can be computed by the angular-spectrum method [48]:

$$\mathcal{P}_{\text{ASM}}(U, d) = \iint_{-\infty}^{\infty} F(u, v) e^{j2\pi d \sqrt{\left(\frac{1}{\lambda}\right)^2 - u^2 - v^2}} \texttt{circ}(u, v) e^{j2\pi(ux+vy)} \mathrm{d}u \mathrm{d}v \tag{2.3}$$

where $F(u, v)$ is the Fourier transform of U, and $\texttt{circ}(u, v)$ returns 1 if $u^2 + v^2 < \frac{1}{\lambda}$, 0 otherwise. Thus, by setting T to an image and $\mathcal{P}(\cdot) = \mathcal{P}_{\text{ASM}}(\cdot, d)$,

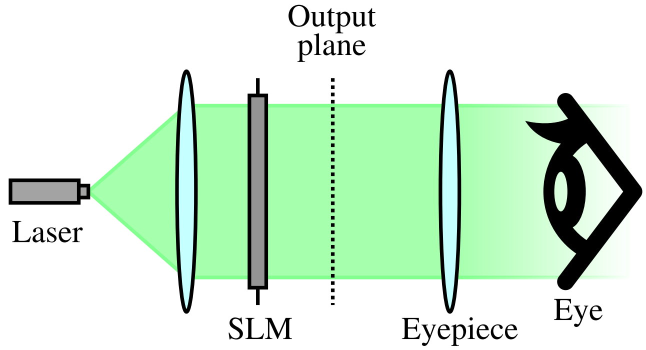


Figure 2.2: **A near-field holographic system.** Most holographic displays leverage a near-field configuration, where the SLM and output plane are separated by a small distance. When used in AR/VR configurations, an eyepiece is used to appropriately image the output for the human viewer.

Equation (2.2) can actually be used to display any desired 2D content at plane d by simply inverting this propagation, without the use of a lens — such a configuration is known as a "near-field" system when d is not particularly large, as visualized in Figure 2.2. By adjusting d in computation, the output content can be moved to a different distance.

Such an approach can also be generalized to display 3D content placed over a spectrum of depths. This is typically done by computing a loss between some target focal stack T_{stack} and the output of the holographic display at the corresponding depths:

$$\min_{\boldsymbol{\phi}} \sum_{d \in D} \left\| T_{\text{stack}}(d) - |\mathcal{P}_{\text{ASM}}(\mathcal{S}(\phi), d)|^2 \right\|^2, \tag{2.4}$$

where D is the set of depths of interest. If the behavior of out-of-focus points is inconsequential, 2.5D regularization can be specified via [32]:

$$\min_{\boldsymbol{\phi}} \sum_{d \in D} \left\| M(d) \cdot \left(T_{\text{stack}}(d) - |\mathcal{P}_{\text{ASM}}(\mathcal{S}(\phi), d)|^2 \right) \right\|^2, \tag{2.5}$$

where $M(d)$ masks out all regions of the output image except where the target stack is in-focus. As another option, holographic displays can also be directly used to reproduce a target light field T_{LF} via the Short Term Fourier Transform [102, 113, 145]:

$$\min_{\boldsymbol{\phi}} \sum_{d \in D} \left\| T_{\text{LF}} - |STFT(\mathcal{P}_{\text{ASM}}(\mathcal{S}(\phi), d))|^2 \right\|^2, \tag{2.6}$$

which can potentially improve visual realism [69].

Because of these capabilities, in recent years, such near-field systems have seen tremendous attention as holographic displays in the context of AR/VR [32, 50, 69, 70, 84, 104]. This is primarily due to three reasons. First, standard display architectures are limited to producing content at a single fixed depth. This limitation can result in nausea and visual discomfort when used to display 3D content in a binocular AR/VR headset. In contrast, holographic displays can in theory program 3D content at any distance, reproducing natural focal cues [84]. Second, these capabilities can be easily extended to provide vision correction by incorporating visual aberrations into $\mathcal{P}(\cdot)$ [84], allowing vision-impaired users to use a headset without needing to wear glasses. Finally, the form factor of AR/VR headsets could be potentially significantly reduced with such an architecture. More specifically, current near-eye display architectures typically require that the display element is placed about one focal length away from the eyepiece, imposing an effective minimum size on a headset. In contrast, in a holographic display, the SLM can be placed arbitrarily close to the eyepiece, as the desired content can be simply formed at the focal plane via an appropriate near-field propagation [70].

In the space of consumer projectors, past work has explored how such systems can be used to implement smaller form-factor projectors [30, 86, 87, 118] as well as correct for optical aberrations [21, 22]. However, no work has explored how the 3D capabilities of holographic systems can be used more broadly in a projector. In this thesis, we argue that this property actually enables more powerful types of illumination in the context of computer vision. We show that this allows for new types of user interfaces and depth sensing, but may require specially-modified holographic systems — we continue this discussion in Chapter 5.

2.1.2 Holographic systems and light redistribution

A second key characteristic is the *light redistributing properties* of a holographic system. In particular, consider the "far-field" configuration as illustrated in Figure 2.3, rather than the near-field setting typically used today in holographic displays. In this setting, collimated laser light illuminates an SLM that is placed one focal length in front of a lens. The output image one focal length behind the lens is then used as the output pattern, *e.g.*, to illuminate the scene via an objective lens under our active sensing setting. Under this configuration, it can be shown that the propagation from the SLM to the output plane is given by the Fourier transform [48]:

$$\mathcal{P}_{\text{lens}}(U) = \mathcal{F}(U). \tag{2.7}$$

The squared magnitude of this expression yields the illumination pattern. Such a lens system is called "far-field" because the resulting diffraction pattern is identical to the case where the SLM and output planes are simply separated by a large distance $d >> \frac{W^2}{\lambda}$, where W is the size of the SLM [48].

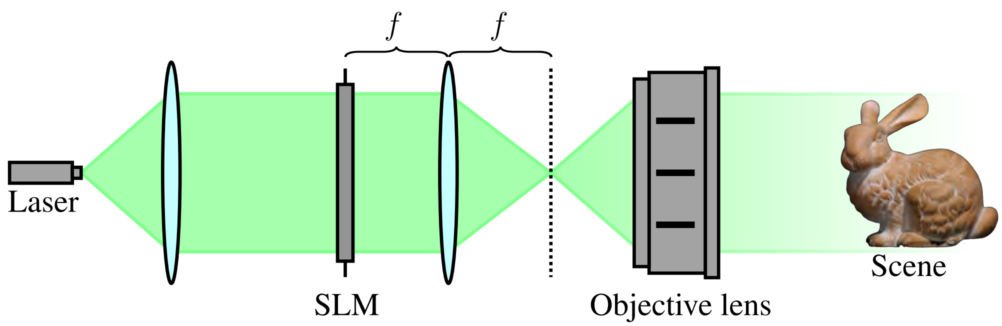


Figure 2.3: **A far-field holographic system.** For the holographic illumination systems we use for vision applications, we primarily leverage a far-field configuration. A collimated laser illuminates an SLM one focal length away from a lens. An objective lens images the output pattern one focal length behind the lens into the scene.

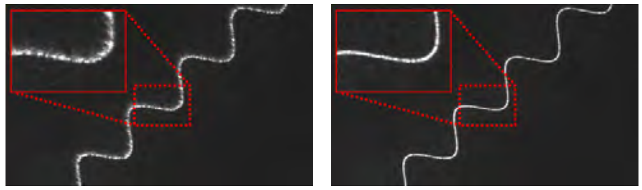
This Fourier relationship yields important conclusions on the behavior of far-field holographic systems. For one, thanks to Parseval's theorem, such displays can be shown to "steer" or "redistribute" light to form patterns:

$$\iint |U_{\text{SLM}}(x,y)|^2 \mathrm{d}x\mathrm{d}y = \iint |\mathcal{P}_{\text{lens}}(U_{\text{SLM}})(x,y)|^2 \mathrm{d}x\mathrm{d}y \tag{2.8}$$

Intuitively, this expression shows that all input energy is used to form the desired output pattern. For instance, forming a pattern consisting of a single point results in all the input energy being concentrated into that point — in effect, the SLM has steered all of its incident light to create this point, without wasting any energy. This inherent "light efficiency" makes these systems a powerful tool for vision applications, which we revisit in Chapter 3 and Chapter 4.

In our work, we primarily leverage this far-field configuration to design our vision-specific holographic illumination systems. Why is this functionally different than the near-field configuration typically used for holographic displays? In the far-field setting, every point on the SLM affects every output point, thanks to the Fourier transform relationship. In contrast, this is not necessarily a given for near-field configurations, where each SLM pixel can only steer light to a limited "diffraction cone" thanks to finite SLM resolution [31, 63]. In practice, this results in reduced contrast and light redistribution for near-field setups [31, 63], making them less desirable for the vision tasks we tackle.

We note that we are not the first to recognize the light redistribution of such systems. Past research has acknowledged the resultant increase in contrast of consumer projectors implemented via holography [21, 22]. Recent work has explored



(a) Without phase correction (b) With phase correction

Figure 2.4: **Effect of calibrating for the phase distortion of a SLM. (a)** A sinusoid pattern generated without taking into account the phase. **(b)** A sinusoid that corrects for the phase when solving Equation (2.2). Calibrating for the distortion caused by phase aberrations at the SLM plane creates a sharper curve.

the use of incoherent versions of these systems to build high-dynamic range consumer projectors [36, 59]. In our work, we leverage this light efficiency to tackle key challenges in computer vision, which we discuss in further detail in Chapter 3 and Chapter 4.

2.2 Calibrating holographic systems

While the aforementioned physical models hold in theory, in practice real systems are far more complicated, where various non-idealities add extra complexity. Accurately modeling these aspects can be crucial to good image/pattern quality in holographic systems. Perhaps the most significant of these effects is any type of distortion resulting from imperfectly-collimated incident light or non-flat SLM surface, as visualized in Figure 2.4. For instance, in the far field setting, these effects can be modeled as:

$$\mathcal{P}_{\text{distortion}}(U) = \mathcal{F}(\mathbf{A} \cdot U) \tag{2.9}$$

where $\mathbf{A}$ is an extra complex modulation capturing the aforementioned distortions. If $\mathbf{A}$ is ignored or incorrectly estimated when solving Equation (2.2), it would result in an extra blur in the output pattern thanks to the Fourier convolution theorem, dramatically reducing resolution. Other effects include undiffracted light, lens non-idealities, and misalignment that can be incorporated into some propagation model $\mathcal{P}_{\text{nonideal}}$ via additional terms or even a neural network [27, 31, 32, 33, 104].

To calibrate these extra factors, two classes of approaches exist. First, one could simply capture a large dataset of input patterns ϕ and output real captured patterns $I(\phi)$. Then, $\mathbf{A}$ and other such factors can be simply optimized through first order optimization methods, *e.g.*, gradient descent [27, 31, 32, 33, 104]. For instance, for the distortion-only model given by Equation (2.9), $\mathbf{A}$ could be optimized for via:

$$\mathbf{A}^* = \underset{\mathbf{A}}{\operatorname{argmin}} \sum_{\boldsymbol{\phi}} \left\| I(\boldsymbol{\phi}) - |\mathcal{P}_{\text{distortion}}(\mathcal{S}(\boldsymbol{\phi}))|^2 \right\|^2 . \tag{2.10}$$

This is the style of approach that we take in Chapter 5. While a simple approach that can capture nearly any effect of interest, it can be a data hungry one, and often results in poor convergence thanks to many local minima.

A second class of approaches relies on more principled methodologies. For instance, to calibrate for $\mathbf{A}$, one can measure changes in the diffraction pattern for different regions on the SLM. More formally, let m, n be the dimensions of the SLM, and $q \ll m$ and $r \ll n$. Let $P(x, y)$ be some function with non-zero values only in range $(-\frac{q}{2}, \frac{q}{2}) \times (-\frac{r}{2}, \frac{r}{2})$. Then, we can write a shifted version of $P(x, y)$ as follows:

$$P_{(a,b)}(x, y) = P(x - a, y - b) \tag{2.11}$$

where a and b represent a shift of the pattern such that the $q \times r$ patch is centered over point (a, b). The 2D Fourier transform of $P_{(a,b)}$ will be given by:

$$\mathcal{F}\{P_{(a,b)}\}(s, t) = \mathcal{F}\{P\}(s, t) e^{-i2\pi(as+bt)} \tag{2.12}$$

Consider the case where there are no distortions at the SLM. Then, if we display $P_{(a,b)}$ at the SLM, by Equation (2.9), the pattern formed at the image plane will simply be a 2D Fourier Transform of P:

$$I_{\text{ideal}}(s, t) \propto \left| \mathcal{F}\{P\}\,(s, t)\, e^{i2\pi(as+bt)} \right|^2 \tag{2.13}$$

$$\propto |\mathcal{F}\{P\}\,(s, t)|^2 \tag{2.14}$$

Note that the created pattern is invariant to the exact values of a and b; shifts in the Fourier plane translate to phase modulation in the image plane, which is dropped in the intensity calculation.

Now, consider the case where there is distortion $D_{(a,b)}$ that can be approximated in a $q \times r$ region around point (a, b) as a linear phase ramp:

$$D_{(a,b)}(x, y) \approx e^{i2\pi\left(u_{(a,b)}(x-a)+v_{(a,b)}(y-b)\right)} \tag{2.15}$$

where $u_{(a,b)}$ and $v_{(a,b)}$ are the slopes of the phase ramp. Then, if pattern $P_{(a,b)}$ is displayed, the corresponding wavefront at the DMD will be given by $P_{(a,b)}D_{(a,b)}$. By the Fourier shift theorem, the resultant intensity captured by a camera at the image plane will be shifted by the slope of the phase ramp:

$$I_{(a,b)}(s, t) \approx I_{\text{ideal}}\left(s - u_{(a,b)}, t - v_{(a,b)}\right) \tag{2.16}$$

This equation suggests a simple method to calibrate for $\mathbf{A}$. If I_{ideal} can be found, then to estimate the phase distortion around point (a, b), the resulting projected image $I_{a,b}$ can be simply cross-correlated with I_{ideal} to determine the shift. The peak of this cross correlation will give an estimate of the gradient of the phase distortion $\left(u_{(a,b)}, v_{(a,b)}\right)$. These gradients can then be recombined using a Poisson solve to recover a map of the phase distortion [120]. We leverage this approach to calibrate our SLM in Chapter 4. Another set of approaches divides the SLM into blocks of pixels and interferes pairs of blocks together [123, 131]. By measuring these interference patterns, one can compute the relative phase between the two blocks of pixels, which can again be integrated to recover a phase distortion map.

In general, this class of principled phenomenon-specific calibration is often effective for simpler systems, but does not trivially generalize to more complicated optical configurations, where multiple non-idealities occur in tandem. In such cases, the aforementioned gradient descent-based approach may be more suitable.

2.3 Solving for the right SLM pattern

Most modern systems minimize Equation (2.2) via gradient-descent type algorithms [26, 104], *e.g.*, Adam [71], implemented via automatic differentiation in modern machine learning frameworks. Such formulations allow for easy integration with more complex forward models as discussed in Section 2.2 — we use this approach in Chapter 5. While such approaches typically produce extremely high-quality results, they require a significant number of iterations to converge. Alternatively, other methods rely on ping-pong algorithms that can be seen to be equivalent to projected gradient descent [41, 46]. For example, for a phase SLM, a Gerchberg-Saxton approach [46] would iterate the following steps:

$$T_{\text{curr}} = \mathcal{P}(e^{j\phi_{\text{curr}}}), \tag{2.17}$$

$$T_{\text{curr}} = T \cdot e^{j\angle T_{\text{curr}}}, \tag{2.18}$$

$$U_{\text{curr}} = \mathcal{P}^H(T_{\text{curr}}), \tag{2.19}$$

$$\phi_{\text{curr}} = \angle U_{\text{curr}}, \tag{2.20}$$

where $\mathcal{P}^H$ indicates the Hermitian transpose operator of $\mathcal{P}$. Intuitively, such an approach repeatedly propagates between the SLM and output planes, and applies the desired "constraints" at each of these planes, *e.g.*, the wavefront is phase-only at the SLM, and the wavefront is constrained to the desired target at the output plane. These approaches typically require a fewer number of iterations to produce a reasonable SLM pattern, but overall output quality is decreased compared to a gradient-descent approach. We use this class of techniques in Chapter 3 and

Chapter 4. In Chapter 6, we discuss future directions in which computation can be potentially reduced for holographic illumination while preserving output quality.

Chapter 3

Light-efficient holographic illumination

As described in Chapter 1, the active sensors used in computer vision struggle with brightness. In general, active sensors require objects to reflect sufficient light back to the sensor, in order to resolve the desired signal over any background noise or ambient light. This requirement can severely limit practical performance of active sensors. For one, objects that are too far away can reflect too little light to the camera, thanks to distance-squared falloff. Simultaneously, objects that have lower albedo can underexpose the camera, reducing performance. Furthermore, sunlight in outdoor scenarios can effectively wash out any coded signal [55].

One naive solution to this problem is to simply use a brighter active source, that emits more light into the scene. With more emitted light, more light reflects back to the camera, mitigating the above issues. Although theoretically effective, such an approach raises a number of practical challenges. For one, brighter sources draw more power, which can be a luxury in modern applications. For instance, subsystems in mobile devices and AR/VR headsets are currently designed under extreme power constraints, thanks to the limited battery life of these devices. Allocating more power to sensing is not an easy choice for the AR/VR architect, given the power requirements of other headset components. Secondly, thanks to thermodynamics, drawing more power inherently results in more heat, which needs to be appropriately dissipated. The requisite thermal engineering can dramatically increase form factor and add complexity [24, 100]. Thirdly, using brighter sources has diminishing returns. For one, thanks to distance-squared falloff, an increase in power by f results in an increase in range by $\sqrt{f}$. Simultaneously, thanks to the statistics of Poisson noise, a power increase by f also only results in an increase in signal-to-noise ratio by $\sqrt{f}$ for an object at a fixed distance. Finally, active sensors can also be sensitive to saturation, where objects that reflect too much light

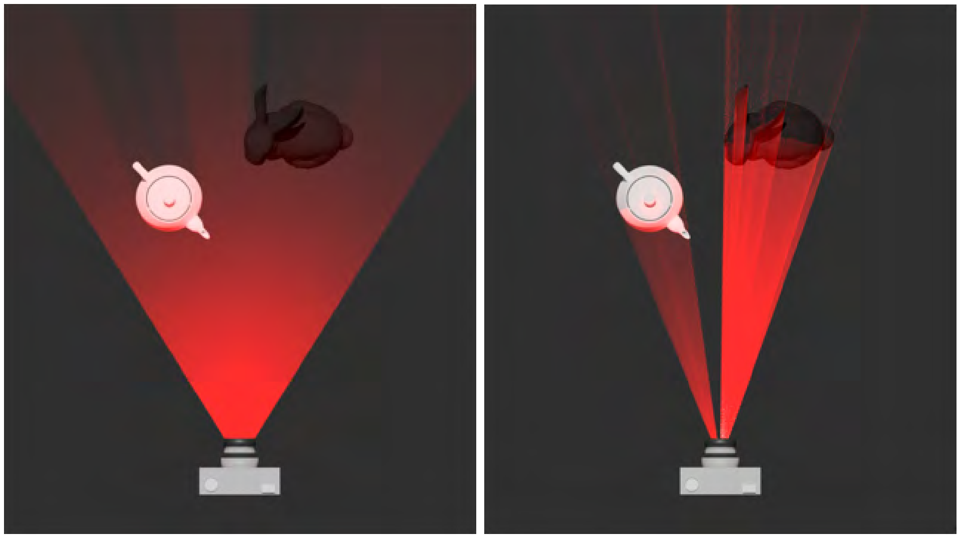


Figure 3.1: **Concept of light redistribution.** Typical active sensors uniformly illuminate the world. In contrast, we propose distributing this energy as necessary, *e.g.*, redirecting energy from bright objects to dark objects.

can overwhelm the camera pixels. Thus, increasing the power naively can saturate parts of the scene that were previously well-exposed, resulting in reduced overall performance.

How then, can we mitigate these brightness challenges without using a brighter source that draws more power? In our work, we rely on the key insight of *light redistribution*, as visualized in Figure 3.1. In particular, current systems typically evenly distribute their energy over the scene. However, if we intelligently concentrate our energy into particular regions of the scene, we can potentially improve overall performance. For example, given some target signal-to-noise ratio (SNR), we can potentially redistribute energy from bright, nearby objects that reflect far sufficient light, to darker, farther objects that reflect too little. In outdoor settings, such a system could move energy from shaded to sunlit regions dominated by ambient light. Alternatively, we can concentrate light according to some region-of-interest defined by another algorithm or system — for example, if an active sensor was used in conjunction with a stereo camera, the active sensor could concentrate energy in textureless regions [128]. Such scene-adaptive light concentration clearly could yield significant benefits.

We are not the first to note the benefits of light concentration for computer vision. Gupta et al. demonstrated that concentrated structured light patterns allow for much better 3D reconstructions in outdoor scenarios [55]. Sun et al. showed

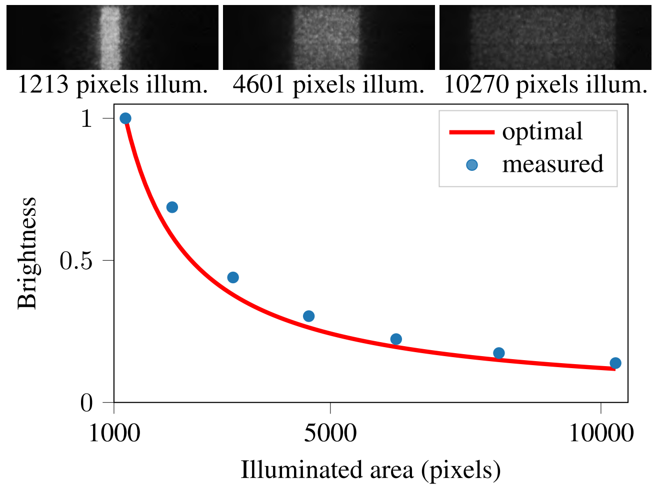


Figure 3.2: **Light redistribution efficiency.** We plot the average brightness of our holographic illumination for different sizes of illuminated area, normalized by the largest brightness. Our system follows an inverse relationship between output brightness and illuminated area, as predicted theoretically.

that using a patterned flash system with light concentrated into a sparse dot pattern increased effective range [124]. O'Toole et al. temporally concentrated light in conjunction with fast sensor modulation in the context of light transport probing, for applications in 3D reconstruction under ambient light, direct/indirect separation and imaging through scattering media [99]. Achar et al. later applied these probing ideas to time-of-flight cameras [1]. However, all of this work has focused primarily on generic concentration patterns that are globally applied, rather than the spatially-adaptive per-scene concentration we propose.

On the side of per-scene adaptation, one related line of work adaptively controls the sensitivity of individual camera pixels according to scene content [44, 88, 94, 96]. Another line of work adjusts projector illumination such that a camera or user sees the desired image [6, 7, 15, 16, 52, 95], which can be applied to high dynamic range capture [139]. While these approaches can successfully ensure that no pixels are saturated in a single measurement, such methodologies based on reducing the signal at individual pixels cannot fundamentally mitigate brightness challenges as we have suggested.

With these ideas in mind, how do we build an illumination system that can implement our scene-adaptive light concentration? As mentioned before, most of the past research in light concentration has relied on static concentration patterns [1, 99, 124], precluding the systems they built from any adaptive concentration. Gupta et al. relied on a scanning mirror to move a laser point across the scene, where light was concentrated by adjusting scanning speed [55]. We could poten-

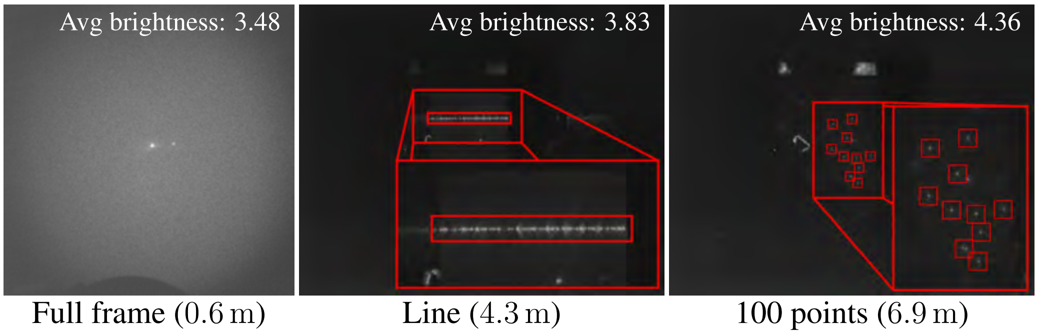


Figure 3.3: **Implementing illumination schemes on a holographic illumination prototype.** Our holographic approach can potentially be used to implement any illumination scheme. We use schemes on a white poster board placed at different distances. The sparser the pattern, the farther we can project it while receiving similar amounts of light.

tially use a similar idea for our scene-adaptive light concentration, by scanning a laser more slowly in areas where we need more concentration. Unfortunately, most scanning mirror systems are designed for certain spatiotemporal scanning patterns, and operating them for other patterns can dramatically reduce speed [34]. Thus, such a methodology can only concentrate light in very simple patterns when operated in real time, *e.g.*, 1D patterns in the case of the laser line scanner used by Gupta et al. [54].

Most related to our to-be-proposed methodology, outside of computer vision, Hoskinson et al. [59] proposed the concept of a coarse light-reallocating analog mirror array combined with a fine-resolution DMD for the purposes of a high-dynamic range consumer projector. Damberg et al. [36] replaced the analog mirror array with a phase SLM for increased resolution. While effective for producing high quality output, these configurations still fundamentally mask light to produce desired patterns, and the use of multiple programmable modulators can result in prohibitive cost and form factor.

To fill this gap, in our work, we propose the use of holographic illumination, where a far-field holographic system with a phase SLM is used as the illumination source. As discussed in Chapter 2, such a system inherently redistributes light to create a desired pattern, and by using a phase SLM no light is lost in the image formation process. Thus, all we need to build a light-concentrating illumination source is plug in our desired concentration pattern into Equation (2.2) and display the resulting SLM pattern. It can be shown that the output of this light-efficient holographic illumination will be governed by the following constraint:

$$\sum_x p_x = A, \tag{3.1}$$

where A is the total input energy and p_x is the output brightness for pixel x. The above equation demonstrates the potential of how a holographic approach can dramatically increase the apparent brightness of an illumination system. If all other pixels are turned off, we can theoretically increase the brightness of a single point by a factor of N over flat illumination that evenly distributes energy, assuming that the output concentration pattern contains N total pixels. To achieve the same effect, a traditional system would need to increase light power N-fold.

We empirically study the light redistribution of our system in Figure 3.2. More broadly speaking, thanks to its redistribution capabilities, this holographic system can be effectively viewed as a generic light-concentrating system that can be used to implement any desired illumination scheme, as illustrated in Figure 3.3. This makes holographic illumination a powerful, reconfigurable tool for active sensing.

3.1 HoloTOF: light-efficient holographic time-of-flight imaging

To demonstrate this holographic approach, we applied it to the context of time-of-flight (TOF) sensors. The emergence of these cameras as the leading imaging technology for 3D sensing has had a significant impact on various scientific and consumer sectors. TOF cameras excel in capturing 3D shapes in diverse situations, making them extremely valuable for a wide range of commercial uses. Notable instances of these commercial applications include the use of Microsoft Azure Kinect in manufacturing, retail, and healthcare industries; the integration of 3D sensors in iPhones and iPads for user authentication and augmented reality experiences; and the implementation of LiDAR systems in robotics and autonomous driving.

Despite their success, like other active sensors, TOF cameras suffer from the aforementioned brightness challenges. They often struggle to provide reliable depth information for objects that are too close or too brightly-colored, as they saturate the sensor. Simultaneously, darker-colored objects at a distance may not reflect enough light, producing noisy range measurements. Another drawback is their sensitivity to ambient lighting; for example, the illumination emitted by these systems can be completely overpowered by direct sunlight, leading to inaccurate depth measurements. Moreover, the need to emit light results in more power consumption compared to passive systems, making them less suitable for applications where energy efficiency is a crucial factor.

In our work, we propose an approach that adaptively redistributes light according to sensing needs as previously discussed, in the context of continuous-wave time-of-flight (CWTOF). Via holographic illumination, this "HoloTOF" system has

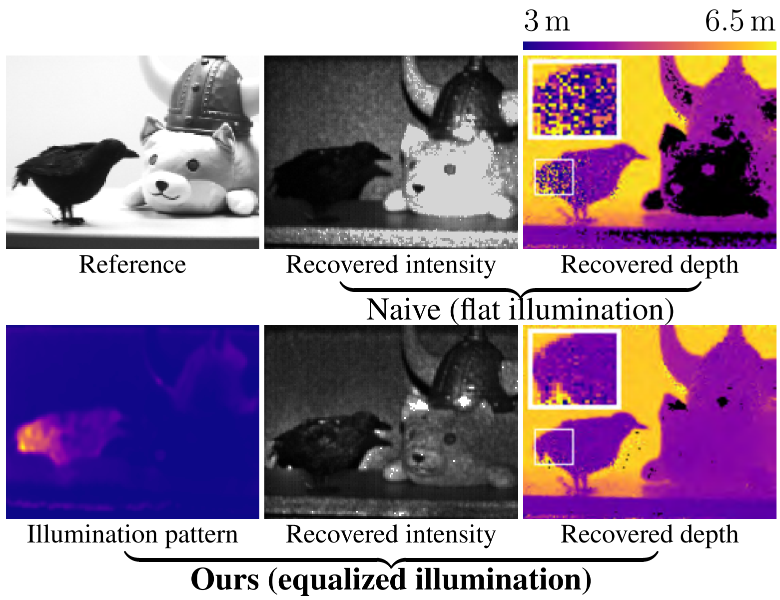


Figure 3.4: **Light-efficient CWTOF via holographic illumination.** Traditional continuous-wave time-of-flight (CWTOF) devices struggle with brightness in real scenes. Illuminating the scene uniformly (flat illumination) results in nearby, brighter objects saturating the pixels on the sensor (shown as black pixels in the recovered depth image), while underexposing farther, darker objects. Our approach redirects light from brighter to darker regions of a target scene using holographic illumination. In short, we can use prior information, like a previous capture, to compute a new relighting pattern that balances light across the scene. This methodology improves the depth measurements of the previously underexposed crow and saturated plush dog. For this result, we used the equalized illumination scheme discussed in Section 3.1.2.

the advantage of reducing the amount of light in areas that are overexposed (*e.g.*, due to retroreflective objects), and increasing the amount of light in underexposed regions (*e.g.*, due to their distance or low reflectivity). Here, we start by providing background on these CWTOF devices, and how past research has tried to address these brightness challenges. We then discuss our adaptive approach, and show results demonstrating its benefits.

3.1.1 CWTOF image formation model

CWTOF imaging devices capture depth information by temporally modulating both the outgoing light from a light source and the incident light on the sensor. If the

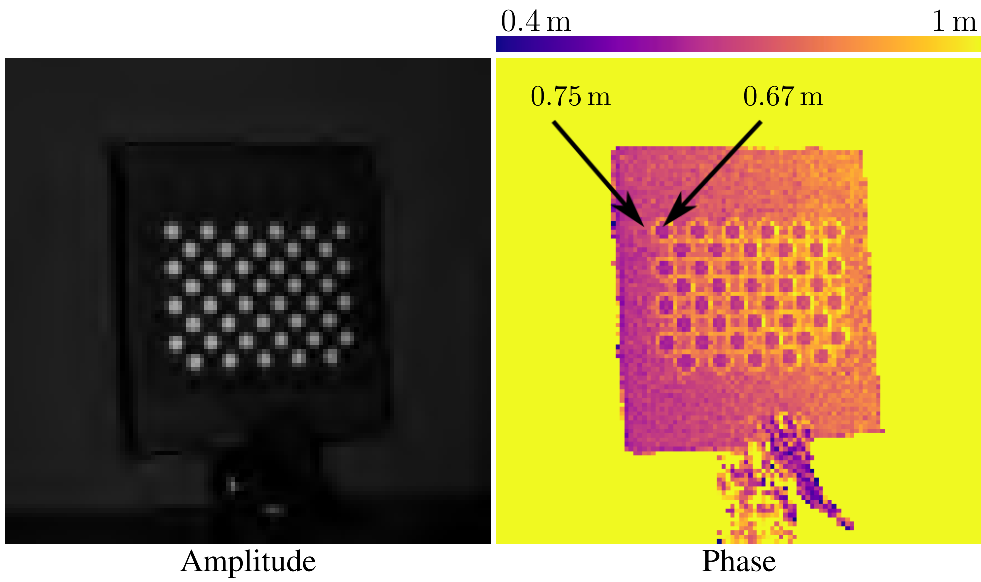


Figure 3.5: **Low-intensity artifacts on off-the-shelf CWTOF measurements.** We used an off-the-shelf Texas Instruments OPT8241 CWTOF sensor to capture an inverted circleboard pattern mounted on a dark stand. The reported distances are lower within the bright circles than the darker regions on the order of many centimeters, demonstrating how intensity information can bleed into CWTOF depth measurements. Simultaneously, the dark stand produces extremely noisy depth as it reflects very little light.

outgoing light is modulated periodically every T by $g(t)$ and the incident light by $f(t)$, the signal seen by a camera pixel q imaging a point with time-of-travel τ is given by [98]:

$$\begin{aligned} q &= \int_0^{MT} \theta p f(t) g(t-\tau)\, dt = M\theta p \int_0^T f(t) g(t-\tau)\, dt, \\ &= M\theta p h(\tau), \end{aligned} \tag{3.2}$$

where $h(\cdot)$ is the correlation of f and g, θ is the light throughput between the source and camera, p is the light projected towards the scene point, and M is the number of periods during a single exposure. Equation (3.2) ignores indirect light transport for simplicity. By appropriately designing f and g, a desired h can be created. Many available devices rely on sinusoidal functions, such as $g(t) = \frac{1}{2}\sin(2\pi\omega t) + \frac{1}{2}$ and $f(t) = \sin(2\pi\omega t + \phi)$ with ϕ a programmable phase offset and $\omega = \frac{1}{T}$, resulting in $h_\phi(\tau) = \frac{T}{4}\cos(2\pi\omega\tau + \phi)$.

By capturing multiple measurements q_ϕ with different h_ϕ, the time-of-travel τ can be found. For example, if q_0, $q_{\frac{\pi}{2}}$, q_π, and $q_{\frac{3\pi}{2}}$ are imaged (colloquially known as *quads*), τ as well as the intensity $a = M\theta p$ can be recovered as follows:

$$\tau = \frac{T}{2\pi} \arctan\left(\frac{q_{\frac{3\pi}{2}} - q_{\frac{\pi}{2}}}{q_0 - q_\pi}\right), \tag{3.3}$$

$$a = \sqrt{\left(q_{\frac{3\pi}{2}} - q_{\frac{\pi}{2}}\right)^2 + (q_0 - q_\pi)^2}. \tag{3.4}$$

Note that the differences between quads removes the contribution of ambient light. We denote this set of four measurements as $i = \left[q_0, q_{\frac{\pi}{2}}, q_\pi, q_{\frac{3\pi}{2}}\right]$. In the rest of this work, we let the boldface versions of the aforementioned variables represent entire images of the same values, *e.g.*, $\mathbf{q}, \boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{a} \in \mathbb{R}^N$ and $\mathbf{i} \in \mathbb{R}^{4\times N}$, where N is the number of pixels in the camera, and $\mathbf{p} \in \mathbb{R}^N$ denotes the projected pattern, assuming the projector has the same number of pixels as the camera.

In practice, a given $f(t)$ and $g(t)$ cannot be perfectly replicated in hardware due to physical limitations, and $h_\phi(\tau)$ often needs to be exhaustively calibrated for each individual device in order to translate a measurement (Equation (3.3)) to accurate depth [103]. Furthermore, the recovered depth can depend heavily on intensity, as shown in Figure 3.5. For instance, like other camera measurements, $\mathbf{i}$ is affected by shot and dark current noise, resulting in noisy depth estimates for parts of the scene that reflect insufficient light. Furthermore, reconstructed depth is often biased under low light as a function of distance [38]. This can cause intensity information to bleed into the recovered depths, causing systematic errors which cannot be temporally averaged over multiple measurements. To calibrate for this phenomenon, a number of approaches propose building models that map measured intensity and range to ground-truth depth [40, 78, 110, 111]. But, these techniques require capturing large datasets for each individual device, limiting their practicality.

Simultaneously, like other cameras, such CWTOF devices are sensitive to saturation. As a result, regions that reflect a significant amount of light back towards the camera will not be accurately ranged. As a result, balancing scenes that contain a mix of objects that reflect a large amount of light back to the camera and objects that reflect very little light are fundamentally challenging for CWTOF devices, as can be seen in Figure 3.4. To address such issues, past work have translated high-dynamic range imaging approaches to the CWTOF context [35, 119]. However, such approaches either require multiple exposures of different durations, limiting the overall capture rate, or theoretical computational sensors, limiting practicality.

In this work, we propose a parallel approach for optically side-stepping these challenges by changing the illumination—an approach that, to our knowledge, has not yet been attempted in the context of CWTOF. In short, if we can redirect light from overexposed regions to underexposed regions, we can ensure that the entire camera is well-exposed, without needing to capture extra images or use experimental not-yet-realized sensors.

3.1.2 Relighting the scene

In our work, we leverage a light-redistributive system, via holographic illumination (Equation (3.1)), to more intelligently relight a scene given prior information. For instance, if a particular region of the scene is known to reflect little light back to the sensor, we can use our HoloTOF illumination system to *concentrate* energy in that region to ensure that a good measurement is captured. With this idea in mind, we propose a simple three-step methodology for CWTOF relighting:

1. **Estimate scene throughput**. Get an estimate of the scene throughput $\boldsymbol{\theta}$. This can be from a prior capture or from another device like an RGB camera.
2. **Compute new illumination pattern**. Using $\boldsymbol{\theta}$, compute a new illumination pattern $\mathbf{p}_{\text{relight}}$ that redistributes light as needed. Compute the equivalent SLM pattern for holographic illumination, and then capture an image $\mathbf{i}_{\text{relight}}$ with this pattern.
3. **(Optional) Fuse measurements.** Combine $\mathbf{i}_{\text{relight}}$ with past captures to create an improved range measurement.

We describe each step in more detail in the rest of this section.

Estimating the scene throughput

Consider a pixel x, that images a point $\mathbf{l}$ in the world. Ignoring indirect transport, the light throughput between the CWTOF light source and x can be written as $\theta_x = \frac{1}{d_i^2} f_r(\mathbf{l}, \omega_i, \omega_c, \lambda)(\omega_i \cdot \mathbf{n_l})$, where d_i is the distance between the source and $\mathbf{l}$, ω_i is the direction of the source from $\mathbf{l}$, ω_c is the direction of the camera from $\mathbf{l}$, λ is the wavelength, and $\mathbf{n_l}$ is the geometric normal at $\mathbf{l}$. $f_r(\cdot)$ denotes the bidirectional reflectance distribution function (BRDF). For simplicity, we treat all parameters in this section as dimensionless.

These parameters could potentially be estimated in a number of ways. For instance, a previous CWTOF capture $\mathbf{i}_c$ could be utilized. After computing the recovered intensity image $\mathbf{a}_c$, the throughput can be approximated by the element-wise division of this intensity image with the previous illumination $\mathbf{p}_c$:

$$\boldsymbol{\theta} \propto \boldsymbol{\theta}_c = \frac{\mathbf{a}_c}{\mathbf{p}_c} \,. \tag{3.5}$$

Computing the new illumination pattern

With an estimate of the scene throughput $\boldsymbol{\theta}$, we can now determine how to relight the scene. Ignoring global illumination, projecting a pattern proportional to $\mathbf{p}_{\text{equalized}} \propto \frac{1}{\boldsymbol{\theta}}$ will equalize the brightness of the entire scene, such that the quality of the depth measurement is uniform over the entire sensor. Under the constraint given in Equation (3.1), regions with low throughput will now receive more light, while regions with higher throughput will receive less. We term this our *equalized* illumination scheme.

Alternatively, if another CWTOF measurement was previously captured to estimate $\boldsymbol{\theta}$, good estimates of the depths may already be known at most pixels, and thus do not need to be remeasured. The light from these pixels can then be used for darker regions of the scene. We propose an alternate *clipped* illumination scheme:

$$\mathbf{p}_{\text{clipped}} = \begin{cases} \frac{\kappa}{\boldsymbol{\theta}_c} & \mathbf{a}_c < \kappa \\ f & \text{otherwise} \end{cases} \tag{3.6}$$

where κ is a predetermined intensity level for which good depth measurements are captured, and f is an appropriately computed fill value in order to disperse the extra light from Equation (3.1):

$$\begin{aligned} E &= \int_0^{\theta_{\max}} \frac{\kappa}{\theta} N_\theta \, d\theta, \\ f &= \frac{(N - E)}{N_{\mathbf{a} > \kappa}}, \end{aligned} \tag{3.7}$$

where θ_{max} is the maximum throughput, N_θ denotes the number of pixels with throughput θ and previously underexposed $a_c < \kappa$, and $N_{\mathbf{a}>\kappa}$ denotes the number of pixels that were previously well exposed. In short, E represents the amount of energy required to relight the underexposed pixels, while the expression f spreads the remaining energy over the already well-captured pixels.

Once we have computed our desired illumination pattern, we apply Equation (2.2) to identify an SLM pattern that produces the desired output. We then capture a measurement of the scene with this SLM pattern in our holographic illumination.

Fusing measurements

If the scene is static, after relighting, we now have two corresponding measurements, $\mathbf{i}_c$ and $\mathbf{i}_{\text{relight}}$. Inspired by the simple observation in Section 3.1.1 that brighter pixels, as long as they are unsaturated, have better SNR than dimmer pixels, we propose a simple fusion:

$$\boldsymbol{\tau}_{\text{fused}} = \begin{cases} \boldsymbol{\tau}_{\text{relight}} & ((\mathbf{a}_c < \mathbf{a}_{\text{relight}}) \vee (\mathbf{i}_c \in \Omega_{\text{sat.}})) \wedge (\mathbf{i}_{\text{relight}} \notin \Omega_{\text{sat.}}) \\ \boldsymbol{\tau}_c & \text{otherwise} \end{cases} \tag{3.8}$$

where $\Omega_{\text{sat.}}$ denotes saturated pixels. This rule simply selects the depth from the captured measurement with larger intensity, ignoring saturated values, and can be applied recursively.

This simple approach does not require careful calibration of the noise characteristics of the time-of-flight camera, making it easily applicable to any off-the-shelf device. If a more accurate camera model is known, such as the probabilities of a particular measurement given a ground truth depth, a maximum-likelihood estimation could be used to instead fuse the measurements. While this would require significantly more calibration data which could be prohibitively expensive to capture, it would produce a better final depth measurement. We leave this to future work.

3.1.3 Hardware and implementation

We show our real HoloTOF prototype in Figure 3.6. We used a Texas Instruments DLP6750Q1EVM for our phase SLM [13], which provides 4 bit phase control at a 1280×800 pixel resolution and supports framerates up to $1440\,\text{Hz}$. For our CWTOF sensor, we used a DME 660 sensor with a $f/1.6$ 4.4-11 mm lens along with a $650\,\text{nm}$ optical bandpass filter to reject ambient light. For a reference camera, we used a UI-3070CP-C-HQ R2 camera fitted with a $f/1.6$ 4-12 mm lens, fitted with a 605nm SWP filter to block out the projector light. For our light source, we used a Thorlabs L638P200 laser diode, operating at $638\,\text{nm}$ at $200\,\text{mW}$. It is

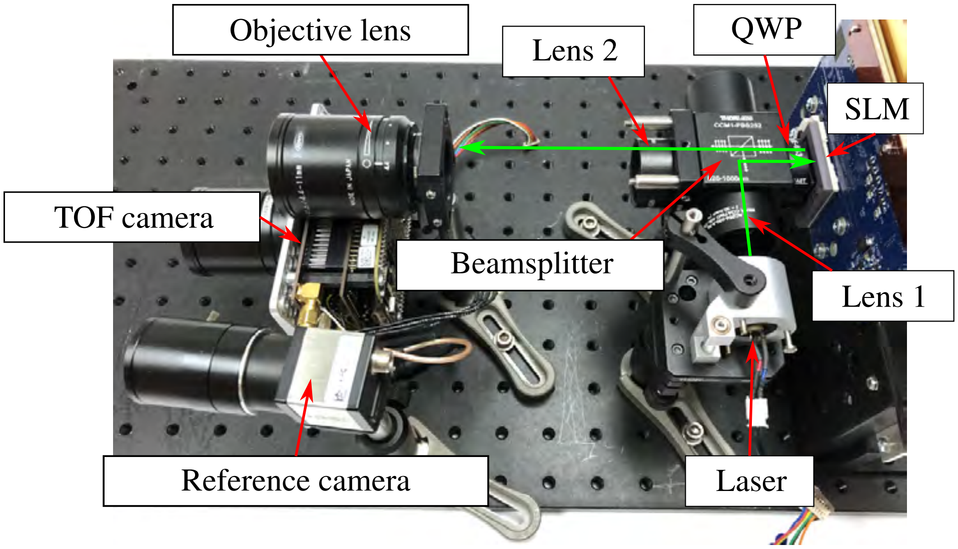


Figure 3.6: **Prototype hardware.** The green arrows indicate the propagation of light in the system.

modulated at $20\,\mathrm{MHz}$, and the exposure time for each quad is set between $24\,\mathrm{ms}$ to $48\,\mathrm{ms}$ for our experiments, resulting in a FPS of about $5\,\mathrm{Hz}$.

Because our SLM is reflective rather than transmissive, more complex optics are required than in Figure 2.3. For Lens 1, a $30\,\mathrm{mm}$ achromatic doublet (Thorlabs AC254-030-A-ML) collimates light from a laser diode. This polarized light then enters a polarizing beamsplitter (Thorlabs PBS252), which redirects it towards our SLM, passing through a $633\,\mathrm{mm}$ quarter-wave plate (QWP) (Thorlabs WPQ10ME-633). After reflecting off the SLM, it passes through the quarter-wave plate again, resulting in an overall rotation of the polarization state by $90°$, such that all of the light passes through the beamsplitter without any light loss. A $75\,\mathrm{mm}$ achromatic doublet (Thorlabs AC254-075-A-ML) forms the projected image (Lens 2). Finally, we used a $f/1.6$ 4.4-$11\,\mathrm{mm}$ lens for our projection lens.

On the software end, we used Fienup's hybrid input-output algorithm [41] to recover the appropriate phase to display for a given illumination pattern. In our PyTorch implementation, each projected pattern requires $0.5\,\mathrm{s}$ on a NVIDIA Titan V for 10 iterations on 1280×800 pixel images. For computing the actual illumination patterns, our implementation of the relighting schemes in Section 3.1.2 takes $2\,\mathrm{ms}$ on an Intel i7-9700k.

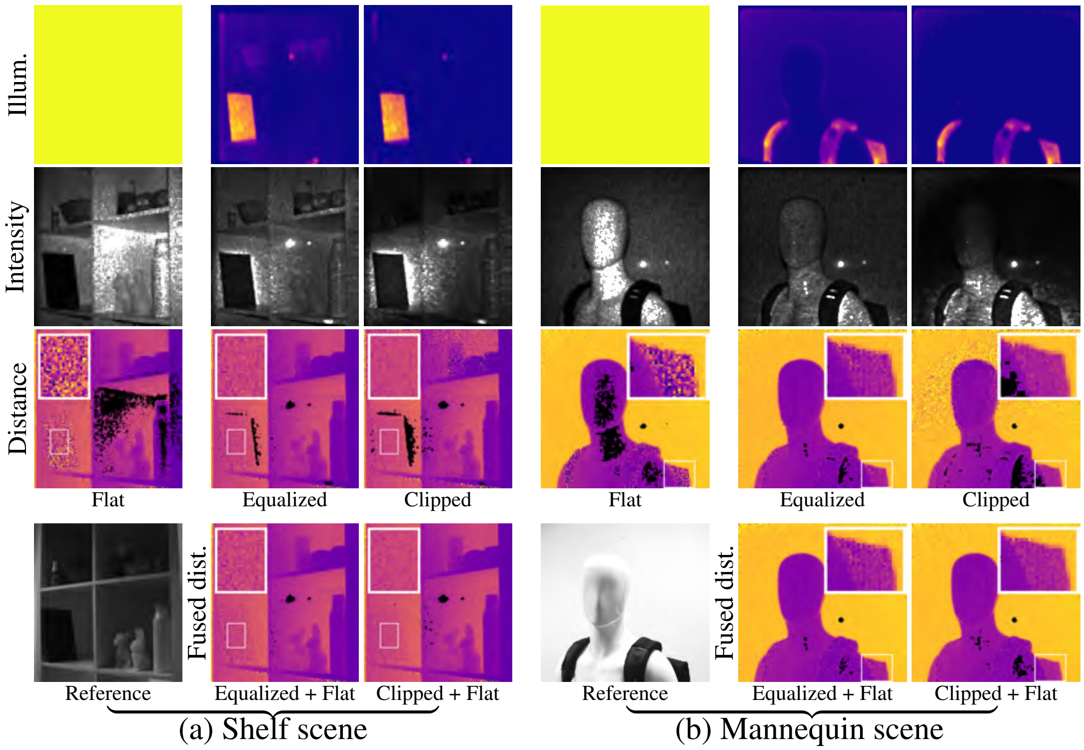


Figure 3.7: **Light redistribution.** Our system enables the redistribution of light from saturated to dark regions of the scene. This results in better depth measurements for both. **(a)** The dark board becomes much less noisy, while the central saturated region significantly decreases in size. **(b)** The specular reflection off the face and neck and mannequin becomes significantly dimmer, while the dark backpack becomes better ranged.

3.1.4 Results

Full frame relighting. In Figures 3.4, 3.7, and 3.12, we show examples of our relighting approach compared with the traditional flat illumination pattern typically used by modern devices. For these results, we assume a flat illumination pattern is used for a prior CWTOF capture, from which we estimate the scene throughput $\boldsymbol{\theta}$. Many of these scenes are saturated in lighter, closer areas when flat illumination is used, resulting in no depth information being captured. Simultaneously, darker, farther objects are underexposed, such that the recovered depths are very noisy. By relighting these scenes with the equalized illumination scheme proposed in Section 3.1.2, the entire scene is better lit, allowing for much better depth captures across the entire image. Alternatively, with the clipped illumination scheme, the intermediate captures may seem undesirable, but the final fused measurements are somewhat improved in the darker regions of the original scene when compared to

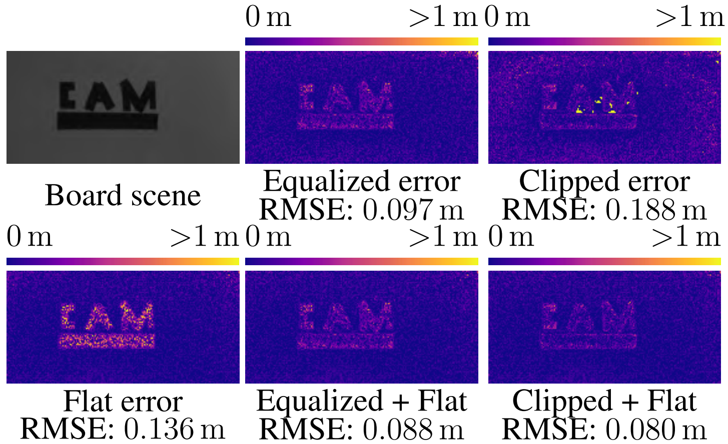


Figure 3.8: **Ground truth comparison.** We used our system to image a board with strips of dark tape, placed at a known position in the scene according to an attached calibration pattern. Light redirection quantitatively improves the reconstruction accuracy, especially in dark regions. We assumed depth 0 for saturated pixels. Note that the root mean-squared error (RMSE) increased for the clipped scheme thanks to extra saturated regions from system non-idealities, but when fused with the flat capture provides the best measurement.

the equalized case. We quantitatively show the improvement provided in Figure 3.8.

Another way to handle dark regions is to simply capture multiple frames and average them together. In Figure 3.11, we compare our relighting results with the averaging of 20 frames under flat illumination. Despite using significantly more frames, the quality of depth in dark regions is still much noisier under the averaging approach compared to our relit scenes, while still remaining saturated in bright regions.

These relighting schemes can also leverage other initial illumination patterns, as may be required by video applications. We show a scene in Figure 3.13 where the previous illumination was not flat. Both our equalized and clipped schemes still perform well for avoiding saturation and underexposure.

Relighting transport separation. Our relighting methodology can be easily integrated with existing approaches for separating light transport. To demonstrate, in Figure 3.9, we integrate our system with technique proposed by Achar et al. [1] for isolating epipolar light transport. Instead of projecting uniform laser lines, our system instead projects linear patterns according to the schemes from Section 3.1.2, and images the appropriate camera pixels like in Achar et al. [1]. This adjusted scheme readily rejects indirect light transport (Figure 3.9(a)) and ambient light (Figure 3.9(b)), while handling saturation and other exposure issues.

Relighting sparse points. We can use our system akin to a LiDAR scan and illuminate a sparse set of points across a scene (see Figure 3.10). This extends the effective range by concentrating energy into sparse measurements. A depth completion algorithm can then recover a dense depth map [82]. By applying our relighting schemes, we can again better handle dark objects.

Iterative relighting. We can apply the relighting processes described in Section 3.1.2 iteratively. More specifically, our methodology relies on an accurate measurement of the throughput of a scene. However, the estimates can potentially be noisy and inaccurate for underexposed or saturated regions of the scene, resulting in poor relighting. We can potentially address this issue by re-estimating the throughput and relighting over an iterative process.

A simple approach is to directly apply the equalized scheme for each new capture, recomputing the throughput, as shown in Figure 3.14(a). As can be seen by the illumination patterns, more and more energy is concentrated into the darkest regions over each iteration in an attempt in order to equalize the brightness. However, as can be seen by the hand in the center, other regions of the scene may become unintentionally saturated due to non-idealities in the holographic illumination hardware.

These effects can potentially be avoided using a more sophisticated approach, as shown in Figure 3.14(b). In short, we can apply the fusion process described in Section 3.1.2 to each iteration, and update the estimated throughput in addition to depth. This results in a much cleaner final depth map, with much smaller saturated regions. This same fusion-based procedure can also be applied to the clipped scheme, which we show in Figure 3.14(c).

3.1.5 Summary

In this work, we demonstrated HoloTOF: a holographic light source for CWTOF imaging. By leveraging the wave properties of laser light, we can build light-redistributive illumination for time-of-flight imaging, using which we can form bright arbitrary patterns. Using this device, not only can we implement any desired prior illumination scheme for TOF sensing, but also intelligently relight scenes. For instance, such a system can redirect light from saturated regions of a scene to darker objects such that the depth of an entire scene is well measured, addressing a key challenge of existing TOF solutions that struggle with brightness in the real world. We validate such an approach in practice, and show that it is compatible with previous illumination modalities.

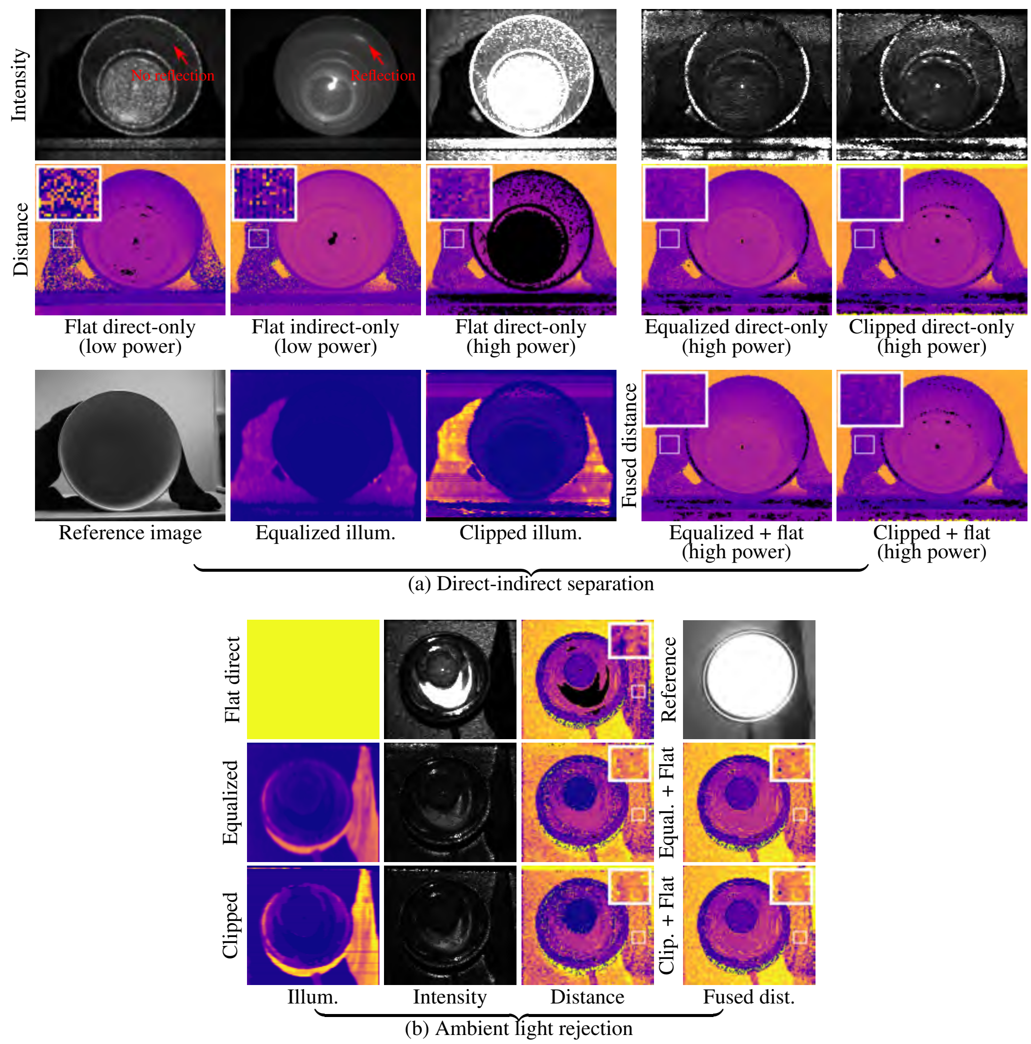


Figure 3.9: **Light redistribution with epipolar scanning. (a)** Following Achar et al. [1], our system can separate direct from indirect light transport, as shown by the reflection on the inside of the pot in the low power scans. However, the recovered depth for the black cloth draped around the pot is noisy. We can increase the laser power, but that saturates the inside of the pot. By applying our relighting schemes, we avoid the saturation while mildly improving the depth reconstruction for the dark cloth. **(b)** We show that our device, like Achar et al. [1], is able to image the surface of a light bulb even when it is turned on, demonstrating ambient light rejection. As an added benefit, our relighting schemes readily handle the saturated specular reflections off the metal lamp surface.

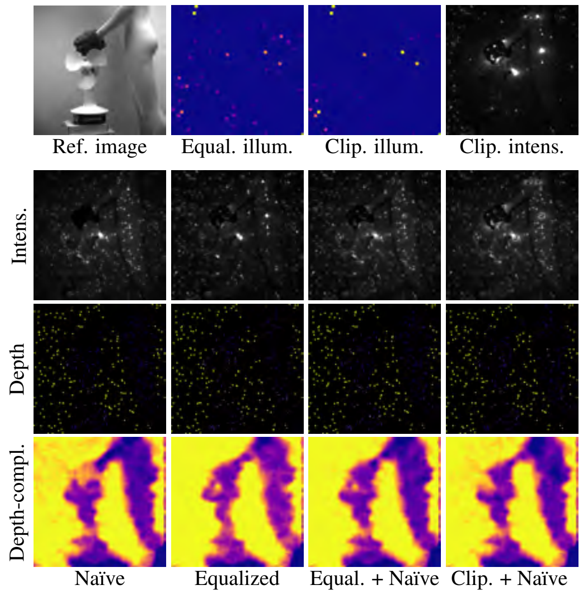


Figure 3.10: **Projecting sparse points with TOF information.** Our holographic device can project a number of sparse bright points like a LiDAR device. When combined with a normal intensity image, a full depth map can be extracted with depth completion postprocessing [82]. We can also relight the scene according to previous captures just like in Section 3.1.2. Notice that the dark-colored fist is not well ranged in the naïve point projection approach. Upon relighting, the fist is readily reconstructed.

3.2 Limitations of light-efficient illumination

In general, there are a couple of limitations associated with our approach for tackling insufficient brightness via light redistribution. First, the improvement afforded by our system depends on scene content. For example, consider the setting tackled by HoloTOF where the system adapts to the throughput of scene objects. If many pixels are not underexposed, then a large amount of light can be redirected to dark pixels, significantly improving the measurement. If only a small chunk is not underexposed, only a small amount of light can be redistributed, providing minor benefit. To explore the real world benefit provided by our method, we took the DIV2K [4] and CIFAR-100 [73] datasets, and used the images as approximate throughput values for real world scenes. With these throughputs, we calculated the potential brightness increase for each scene, as shown in Figure 3.15. We first calculated the brightness increase for the darkest pixel in every scene for our equalized scheme. In both datasets, the distribution of brightness increase is long-tailed, resulting in a large average brightness increase factor and a peak in the distribution

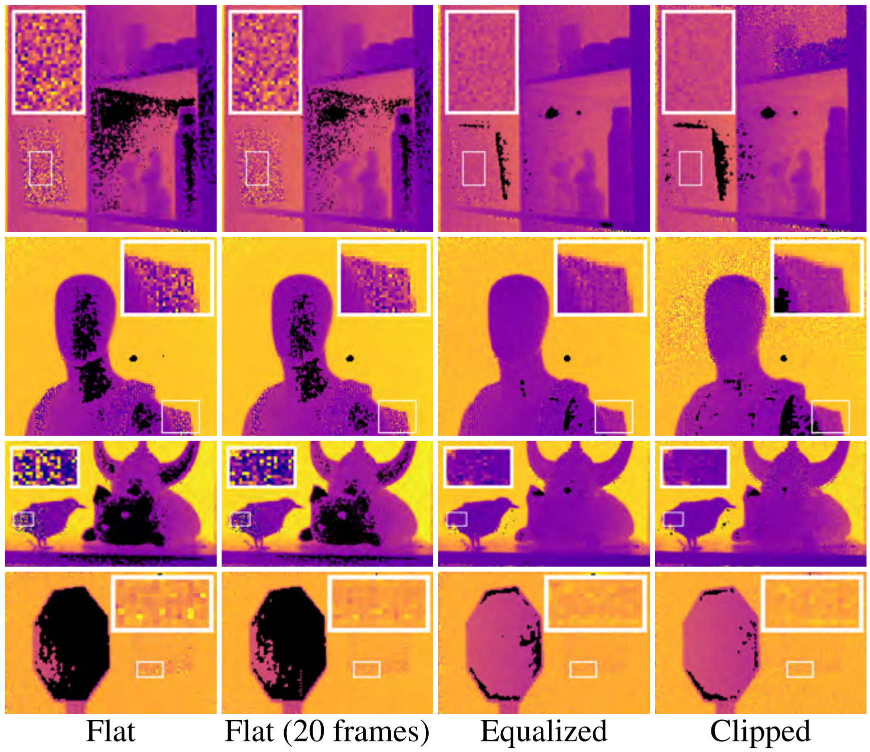


Figure 3.11: **Comparison with averaging multiple frames.** To handle low light scenes, a system could simply capture multiple frames and average them together. We show that even averaging 20 frames under flat illumination does not match our relit results for dark regions, while doing little for the saturated regions.

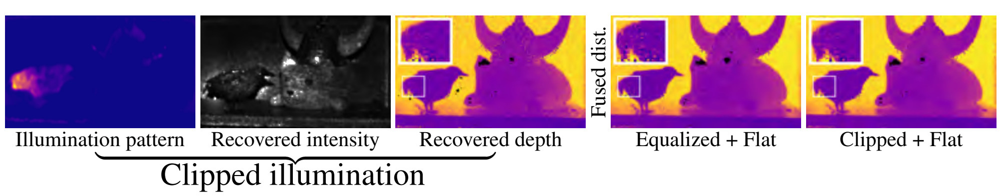


Figure 3.12: **Figure 3.4 continued.** The clipped illumination scheme further reduces the depth noise of the crow, at the cost of a noisier helmet and background. Fusing the relit captures with the original flat measurement provides the best results.

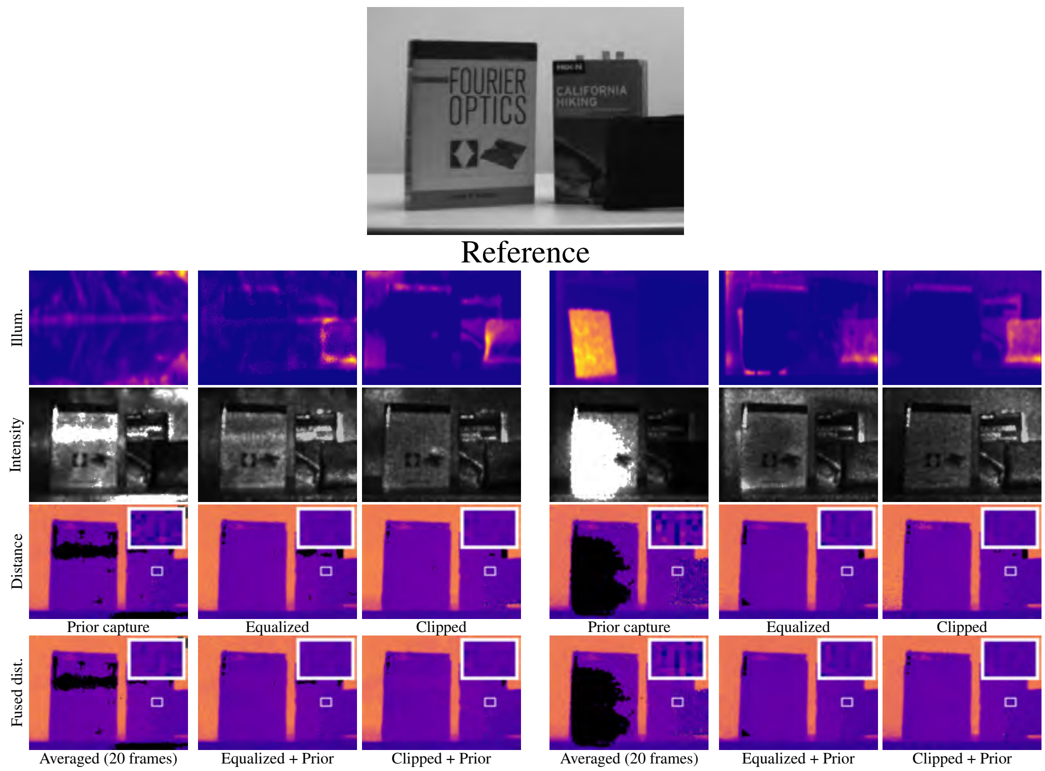


Figure 3.13: **Non-flat prior CWTOF capture.** Here, we show results where the previous capture was not measured under flat illumination. Both the equalized and clipped schemes readily handle general prior illumination patterns.

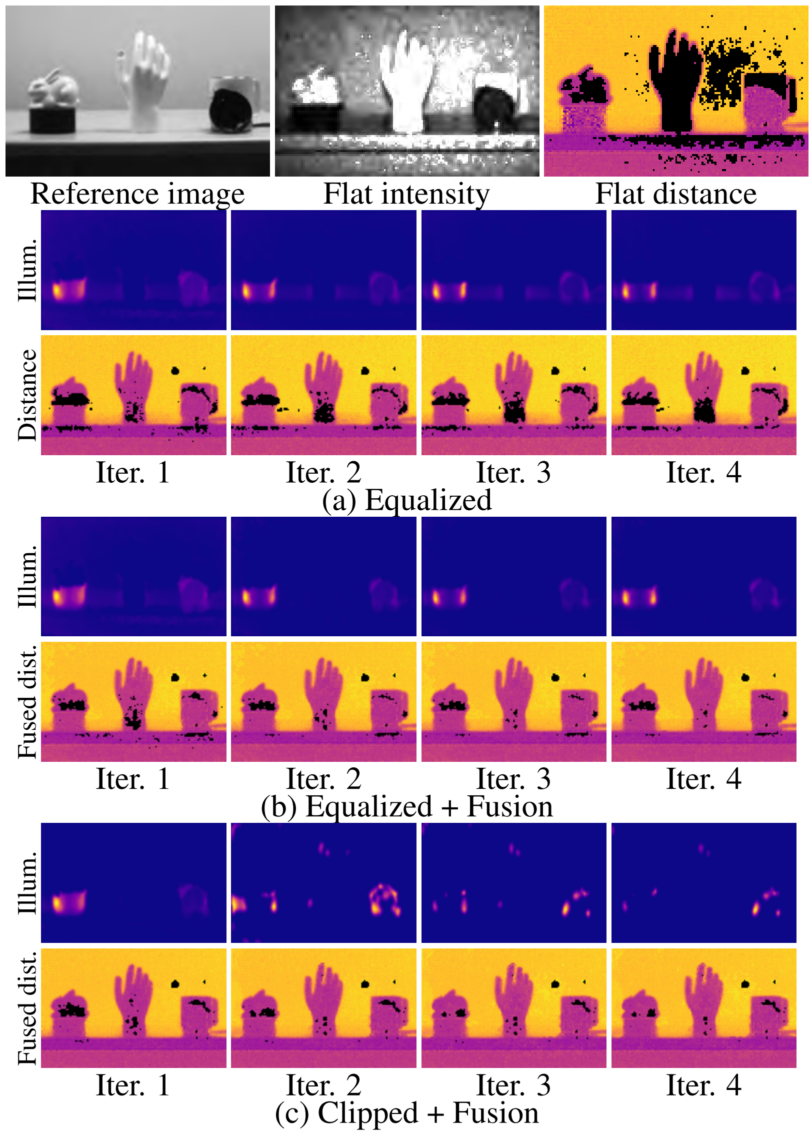
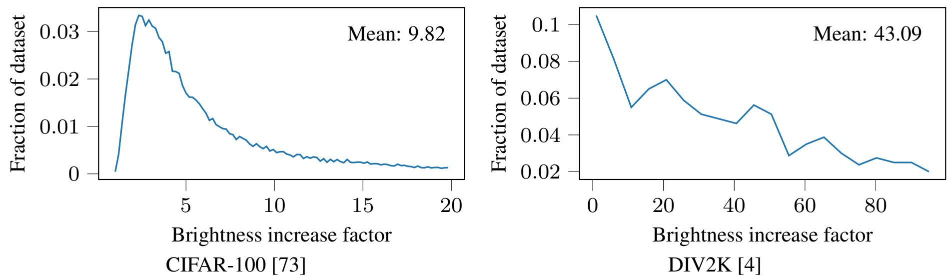


Figure 3.14: **Iterative relighting.** We can iteratively apply the relighting schemes given in Section 3.1.2. While a naive approach provides modest benefit **(a)**, fusing the results with each input image results in an improved depth map (**(b)**, **(c)**).

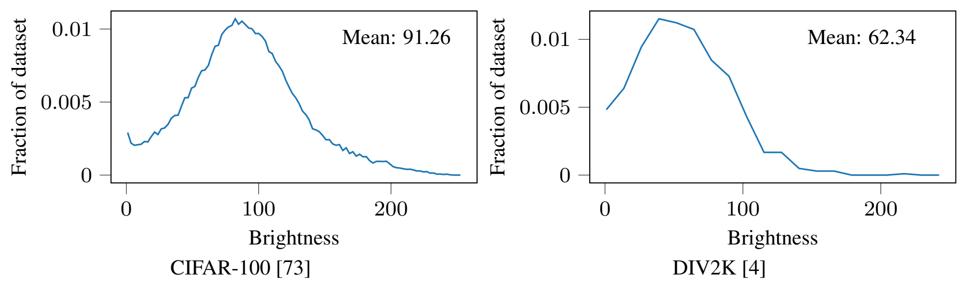
at a smaller factor as shown by Figure 3.15(a). We then plotted the distribution of brightness after equalization for both datasets in Figure 3.15(b). If the minimum intensity required for the desired depth accuracy is less than the equalized brightness, then using the equalized illumination scheme provides benefit. For the clipped scheme, using the constraint given by Equation (3.1), we computed the maximal target brightness for each image, and plotted the distribution over the datasets in Figure 3.15(c). As in Figure 3.15(b), if this target brightness is larger than the required intensity for the desired depth accuracy, the clipped scheme will yield improved measurements in dark regions.

Second, our approach requires prior knowledge of the scene, *e.g.*, where the bright and dark objects are in the scene, or where the sunlit and shadowed regions are. If this prior knowledge is erroneous, overall performance can potentially degrade. For instance, in our HoloTOF system, our illumination and camera are not colocated, and thus our system must estimate correspondences between the camera and illumination to perform relighting. A noisy prior estimate of depth can produce misaligned relighting — we sometimes observed new saturated pixels after relighting as a result.

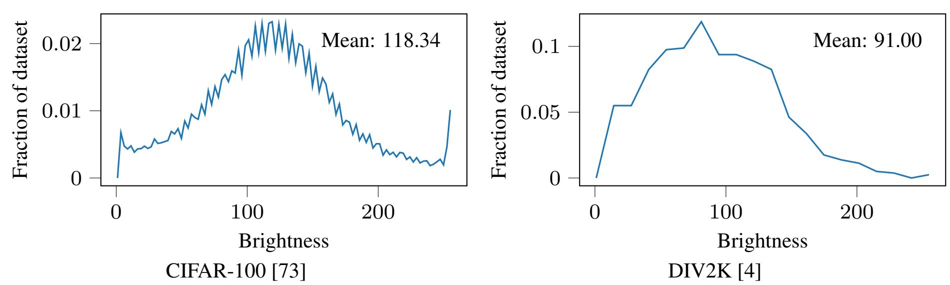
Because our approach requires projecting spatially-varying patterns, it can be affected by defocus when a traditional source would not. Like a standard projector, the objective lens may need to be tuned to appropriately focus projected patterns — however, such a problem could be tackled directly via holographic illumination, as we discuss in Chapter 5. Holographic illumination also comes with certain practical limitations, like a DC spot, speckle and complex computation/modeling. We discuss these in more detail in Chapter 6.



(a) Brightness increase factor for darkest pixel (Equalized scheme)



(b) Output equalized brightness (Equalized scheme)



(c) Max target brightness (Clipped scheme)

Figure 3.15: **Real-world brightness improvement.** For the CIFAR-100 [73] and DIV2K [4] datasets, we analyzed the potential brightness increase for our relighting schemes. In **(a)**, we calculated the increase in brightness for the darkest pixel in every dataset image using our equalized scheme, and plotted the distribution. Both distributions are long-tailed, with a peak at a small brightness increase factor. In **(b)**, we calculated the output equalized brightness level for every image. These intensity values can then be used to gauge the overall depth quality after relighting if a noise model is known for the output depth. In **(c)**, for our clipped scheme, we calculated the maximum intensity level κ that dark regions of the images could be relit to, and plotted the distribution.

Chapter 4

High-speed holographic illumination

Many types of active sensing rely on "structured light". In short, spatially-encoded patterns, typically provided by some sort of projector, are used to illuminate the scene. By analyzing how the response at a camera changes as a function of different patterns, information about the scene (often, depth) can potentially be estimated [45]. However, active sensing systems that leverage such a methodology are often considered slow and unsuitable for dynamic scenes. The reasons are twofold.

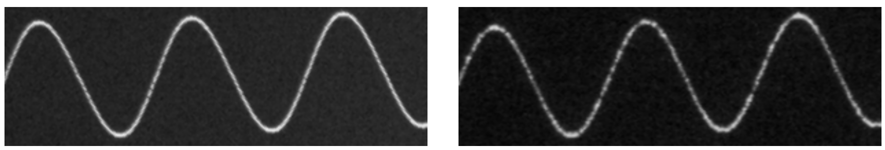
First, modern projector architectures can be very dim, and therefore require long exposure times and low corresponding framerates to resolve. Why is this the case? Most modern projectors either rely on DLP/LCD/LCoS-based mask-driven architectures, or laser raster scanning via a MEMS mirror. Mask-driven systems fundamentally block light to create desired patterns, *i.e.*, a programmable "transparency" reduces the magnitude of light in a spatially-varying fashion, which is then directly imaged onto the scene via a projector lens. Thus, such systems are extremely *light inefficient*, and can also be limited to low power thanks to the extreme heat produced by the blocked energy [100]. On the other hand, raster scanning systems are inherently not very eye safe as energy is concentrated into a very short duration per pixel, and thus are similarly restricted to low power. As a result, the output patterns of most projector systems are typically insuffiently bright for fast sensing.

Second, the underlying projector technologies are also often slow. LCD/LCoS architectures are limited by the modulation speed of the utilized liquid crystal. DLP systems must temporally multiplex many high speed low-bit depth patterns together, resulting in a slow overall framerate. The scanning mirrors used in MEMS systems are also mechanically limited [58]. Thus, most systems can only produce patterns at around $60\,\mathrm{Hz}$ [97], bottlenecking the kilohertz sensing systems available today (event cameras [90], SPADs [126], coded exposure sensors [138], etc.).

To avoid these problems, one option is to sacrifice programmability. For instance, if 1D patterns are sufficient, a scanning mirror can be used to move a con-

Technology	Light Eff.	Pattern Rate	Programmability
Static Pattern	High	N/A	None
Galvo Mirror	High	$\approx 250\,\mathrm{Hz}$	Arbitrary Scan Pattern
MEMS Mirror	High	$\geq 10\,\mathrm{kHz}$	Lissajous Pattern
LCD Projector	Low	$\approx 60\,\mathrm{Hz}$	Arbitrary Pattern
DLP Projector	Low	$\geq 10\,\mathrm{kHz}$	Binary Pattern
Binary Holography (Ours)	High	$\geq 10\,\mathrm{kHz}$	Arbitrary Pattern

Table 4.1: **Comparison of different structured lighting solutions.**



Regular projector   Holo. projector [Ours]

Figure 4.1: **Brightness of binary holographic illumination.** A LCD projector with a $135\,\mathrm{W}$ bulb produces an equally bright curve as our system at $50\,\mathrm{mW}$.

centrated laser line across the scene for each capture. This solution is used by commercial 3D scanners [5], light transport probing [1, 99] and triangulation light curtains [12, 130, 132, 134]. MEMS mirrors can potentially be used to trace out more complex Lissajous patterns at high framerates [58]. With sufficiently sensitive experimental sensors, the individual high-speed binary frames of a DMD-based system could also be used for structured light [125]. However, as illustrated in Table 4.1, no existing solution provides full programmability, while preserving light efficiency and speed.

To fill this gap, in our work, we propose *binary holographic illumination*. In short, we use a DMD as a binary amplitude SLM in a far-field holographic system, that multiplies the amplitude of the wavefront over a plane by 0 or 1. Unlike a traditional configuration where this would limit the output patterns to be binary, the inherent light steering modality of holographic illumination allows for arbitrary programmable output. Thus, such devices can produce arbitrary structured patterns at the native binary rate of the DMD ($\approx 30\,\mathrm{kHz}$), far faster than the $60\,\mathrm{Hz}$ rates previous projectors provided. Simultaneously, as discussed in the previous chapter, these systems produce extremely bright, light concentrated patterns that prevent the need for long exposure times, as shown in Figure 4.1. These capabilities make them a powerful tool for high-speed structured light.

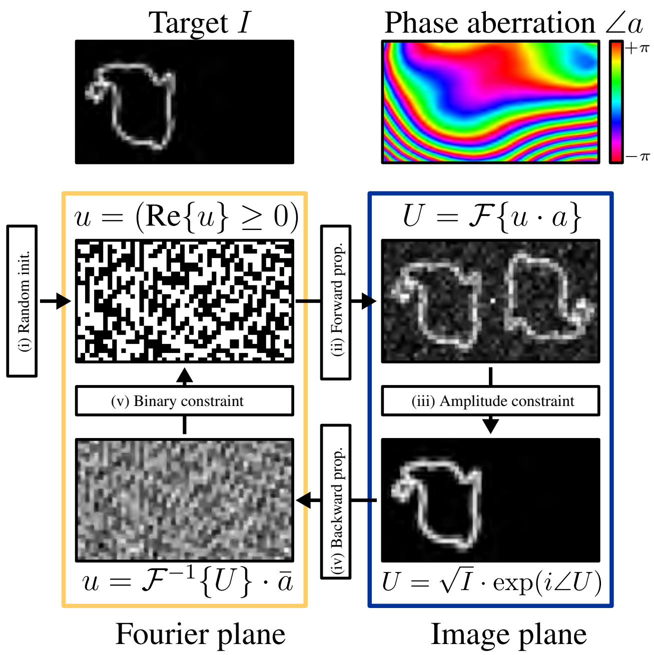
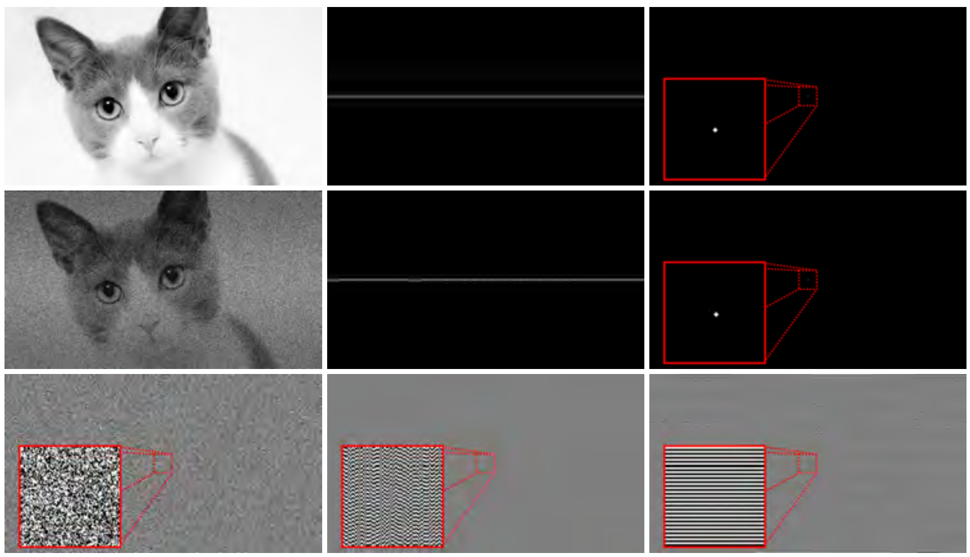


Figure 4.2: **Flowchart for our modified Gerchberg-Saxton (GS) algorithm.** Given a target image I and complex-valued phase aberration image a, the objective is to find a binary pattern that can be displayed on a DMD to reproduce the target image. GS starts by **(i)** initially starting with a random binary pattern, **(ii)** propagating the wavefront from the Fourier plane to the image plane, **(iii)** replacing the wavefront's amplitude with $\sqrt{I}$, **(iv)** propagating the wavefront from the image plane to the Fourier plane, and **(v)** binarizing the result. The GS algorithm quickly converges after a few iterations. An important attribute of our GS algorithm is that it accounts for large phase aberrations created by the DMD, producing sharper target image reconstructions as a result.

Solving for binary patterns in computer-generated holography

In 2D computer-generated holography, the objective is to compute a pattern $u(x, y)$ that produces a target image $I(s,t) = |U(s,t)|^2$, analogous to Equation (2.2). However, the binary nature of the desired pattern imposes an additional constraint that needs to be accounted for. To achieve this, existing work has used variants of soft rounding [77] or Gumbel-Softmax estimators [33] to model quantized SLMs as part of a first-order optimization approach. In our work, we use a modified Gerchberg-Saxton (GS) algorithm [46], similar to past work in binary holography [89, 106, 123]. The algorithm alternates between enforcing a constraint on the hologram's intensity at the image plane, and enforcing the binary constraint on



(a) A dense pattern (50.02% pixels on) (b) A line pattern (50.00% pixels on) (c) A point pattern (50.01% pixels on)

Figure 4.3: **Binary holograms generated through the modified Gerchberg-Saxton optimization procedure. Row 1**: The target patterns. **Row 2**: Simulated reconstructions. Scaled such that the white level is 25% of the maximum value, and the black level is 2.5% of the maximum value. **Row 3**: The recovered binary patterns. Note that approximately 50% of the DMD pixels are turned on in the recovered binary patterns, indicating that only half of the light is blocked by the DMD.

the DMD pattern at the Fourier plane.

After initializing the DMD pattern $u(x, y)$ with random binary values, the algorithm iteratively performs four simple operations to compute the hologram, as highlighted in Figure 4.2. First, we use Equation (3.2) to simulate the propagation of the wavefront from the Fourier plane to the image plane. This involves performing an element-wise multiplication with a pre-computed phase pattern $a(x, y)$ and computing the Fourier transform of the result, producing a conjugate-symmetric wavefront $U(s, t)$. Second, we keep the phase $\angle U(s, t)$ of this wavefront, but replace its amplitude to match the target intensity image. Third, we invert the propagation operator by using an inverse Fourier transform and performing an element-wise multiplication with the complex conjugate of the phase pattern $a(x, y)$. And fourth, we binarize the result, by setting all values with positive real components to 1 and setting all other values to 0. The GS algorithm repeats these four steps until convergence, which typically requires only a few iterations. Note that the target intensity image is designed to account for conjugate symmetry via additional padding, as visualized in Figure 4.2. Example outputs are shown in Figure 4.3.

4.1 Holocurtains: programming light curtains via binary holography

To illustrate the power of this binary holographic system, we applied it to the context of light curtains. A light curtain is an optical barrier that detects the presence or absence of objects within regions of 3D space [68]. For example, light curtains are used in elevators and garage doors, in order to keep doors open when a person or object is in the doorway. Safety light curtains also are used in environments containing hazardous equipment (*e.g.*, machine tools, robotic arms) to protect personnel from injury, by automatically turning off dangerous machinery whenever a curtain is breached.

These light curtains involve two key components: emitters and receivers. Traditional light curtains position an emitter to directly illuminate a receiver through direct line of sight. If an obstacle blocks the light traveling from the emitter to the receiver, the drop in the detected light signal triggers an event. While extremely reliable and simple devices, light curtains must be physically configured for their specific environments, which is a laborious process.

Wang *et al.* [134] recently proposed a programmable approach to generate light curtains through triangulation. A scene is illuminated with a laser line, and the response is measured with a line scan camera [134] or a rolling-shutter camera [12]. The intersection of the illumination and sensing planes produces a 3D line (see

(a) An arm passes through a toy cow-shaped light curtain

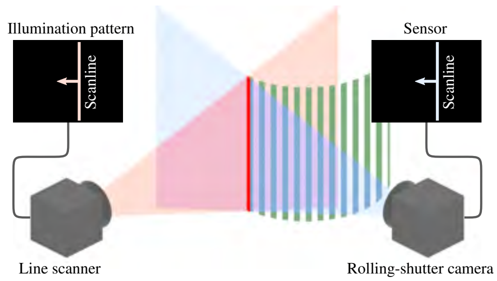
(b) A foam board sweeps through the light curtain

Figure 4.4: **Holocurtain visualization.** Our Holocurtain system turns a user-defined volume or surface (*e.g.*, a toy cow) into a 3D light curtain. **(a)** An arm passes through the light curtain. In the left image, blue pixels represent the light curtain and green pixels highlight areas intersecting the light curtain. The right image represents the raw data from the system used to detect the intersections. **(b)** The left image shows a foam board moving through the light curtain. The right image is a composition of raw measurements captured at different instances in time, where the colors represent different frames.

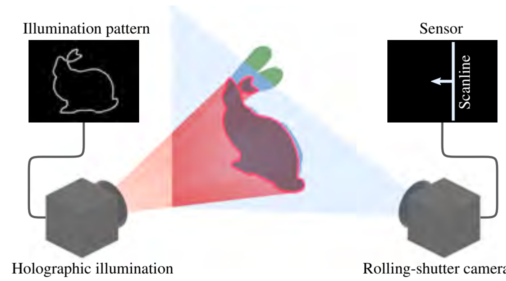
Figure 4.5(a)). If an object touches this line, light from the source reflects off the object and reaches the camera—triggering an event. Rapidly changing the position of the illumination and sensing planes (*e.g.*, with mirror galvanometers) creates ruled surfaces, *i.e.*, surfaces defined by unions of straight lines. Triangulation light curtains offer several benefits, including the ability to program the shape in real time and operate under strong ambient light, which can potentially be leveraged for new safety applications including assisted or autonomous navigation through unknown environments [8, 9, 109].

Despite these advantages, a key limitation is that these light curtains have been fundamentally restricted to being ruled surfaces. Moreover, prior systems offer only one degree of freedom over the positions of the laser line and scan line, limiting triangulation light curtains to an even smaller subset of ruled surfaces. Current prototype systems are only designed to produce either predominantly vertical curtains [12, 134] or horizontal curtains [130, 132].

In this work, we remove the constraints on the shape of light curtains by explor-



(a) Triangulation light curtains [12, 134]



(b) Holocurtains **[ours]**

Figure 4.5: **Illustration of light curtain systems. (a)** A triangulation light curtain combines a laser line with either a 1D line sensor [134] or a rolling-shutter camera [12]. The light curtain (green) is formed at the intersection of the illumination plane (red) and sensing plane (blue), which are synchronously scanned together to form a ruled surface. **(b)** We propose using a holographic illumination system that steers light to select regions of the scene at high speeds. The light curtain of a bunny is formed by synchronizing illumination patterns (cross-sections of the bunny) with the rolling-shutter sensor.

ing a novel approach to structured illumination that we call "Holocurtains". Specifically, we replace the laser line with a holographic source capable of generating arbitrary illumination patterns by redistributing light at high speeds (up to $10\,\text{kHz}$). When synchronized with a rolling-shutter camera, our system is capable of generating light curtains of arbitrary shape, as illustrated in Figure 4.5(b). Moreover, we demonstrate the ability to multiplex multiple light curtains into a single measurement. We leverage these new capabilities for a number of new applications, including optical disturbance detection and 3D optical touch sensing.

We start by discussing how light curtains work, and how they can potentially be formed via a fast, bright projector via binary holographic illumination. We then show results of our Holocurtain system and finish with discussion.

4.1.1 Generating an arbitrary light curtain

The objective of a light curtain is to detect if an object touches a user-defined virtual surface G. This can be done with a camera and projector, by following three simple steps: expose a single camera pixel on the sensor, computationally intersect the camera ray with virtual surface G, and project a pattern that selectively illuminates the intersection. Although this procedure creates the desired light curtain, this naive solution would need to be repeated for every camera pixel individually—a time consuming endeavor.

Instead, one can start by simultaneously exposing an entire row (or column) of pixels to produce a planar viewing frustum, or sensing plane (see Figure 4.5). The next step is to compute the intersection between this plane and the virtual surface G. Finally, the projector can then selectively illuminate the intersected regions.[1] This process would be repeated for every row of pixels on the sensor.

To produce light curtains of arbitrary shape, we need a projector with three key properties: high light efficiency, speed, and programmability. However, as discussed earlier, no existing systems meet all these criteria. Prior light curtain solutions therefore opt to use either a galvo mirror system [12, 134] or MEMS mirror [99, 130, 132] to scan a laser line across the scene, which maintain high light efficiency and speed. However, because these prior systems could only project line patterns, the shape of G was constrained to being a ruled surface.

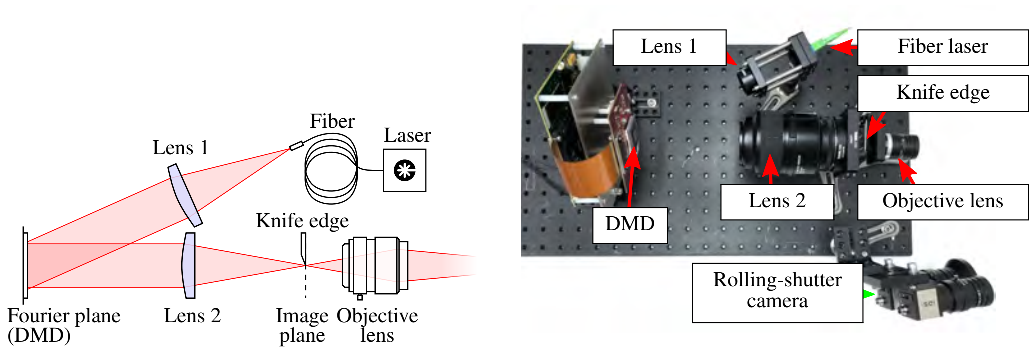
As a result, in our work, we leverage binary holographic illumination, which simultaneously provides light efficiency, speed and programmability. Our setup is shown in Figure 4.6. We use a rolling-shutter camera to image different sensing planes over time, for each of which our holographic system illuminates a different pattern. With such a system, we can create light curtains of arbitrary shape.

Hardware details

We show an image of our hardware setup in Figure 4.6(b). For our laser source, we use a Coherent Sapphire LPX 530-300 Laser, which emits $530\,\mathrm{nm}$ light anywhere from $10\,\mathrm{mW}$ to $330\,\mathrm{mW}$. We use $300\,\mathrm{mW}$ for our experiments. The laser light is collimated using lens 1, a $75\,\mathrm{mm}$ achromatic doublet (Thorlabs AC254-075-A-ML). A DLP Lightcrafter 6500 EVM from Texas Instruments controls the binary DMD, which has a resolution of 1920×1080, and operates up to $9523\,\mathrm{Hz}$; we use the Pycrafter 6500 library [107] to interface with the device. Lens 2 is a $105\,\mathrm{mm}$ $f/2.8$ DSLR lens (Nikon AF Micro-Nikkor) focused at infinity. We used a $9\,\mathrm{mm}$ $f/1.4$ lens (Fujinon HF9HA) for the objective lens. Our optics are angled to select the brightest mode that appears from the DMD and a knife edge at the image plane blocks the bright DC component and the conjugate-symmetric copy, which we discuss in more detail in Section 4.2.

For the rolling-shutter camera, we used a UI-3240CP-NIR camera, fitted with a $8\,\mathrm{mm}$ $f/1.4$ lens and operated in $2\times$ binning mode for an image resolution of 640×512. We also mounted a $531\,\mathrm{nm}$ bandpass filter with a $10\,\mathrm{nm}$ FWHM to reject ambient light. To match the temporal resolutions of the DMD and the rolling-shutter camera, we ran the DMD at about $7575\,\mathrm{Hz}$ with a $40\,\mathrm{MHz}$ pixel clock at the camera for a final framerate of $28.64\,\mathrm{Hz}$, with 256 projector patterns per frame. We set the exposure of the camera to the pattern exposure time of the DMD. This

[1] This step assumes that the sensing plane is not an epipolar plane of the projector-camera system.



(a) Illustration of illumination optics (b) Photo of light curtain setup

Figure 4.6: **Overview of prototype holographic illumination.** **(a)** Illustration of setup. The light from a $530\,\mathrm{nm}$ fiber-pigtailed laser is collimated by a lens, and illuminates a DMD (digital micromirror device) at a 24° angle. The DMD selectively reflects light back through a second lens, which forms a holographic image at the image plane. Since this image is conjugate-symmetric, we position a knife-edge aperture to block half of the image, along with the bright DC component. An objective lens projects the resulting image into the scene. **(b)** Photo of the prototype setup, which includes both the illumination optics and the rolling-shutter camera used for generating light curtains.

camera was placed about $18\,\mathrm{cm}$ from the projector's center of projection.

We also added an additional UI-3240CP-NIR camera mounted with a $537\,\mathrm{nm}$ notch filter with a $162\,\mathrm{nm}$ FWHM to aid with visualization. For this camera, we used $2\times$ binning mode with a $6\,\mathrm{mm}$ $f/1.2$ lens.

4.1.2 Results

As we showed in Section 4.1.1, our setup allows us to generate arbitrary light curtains. To illustrate the new capabilities of our system, we demonstrate four different categories of tasks that are difficult or inefficient for other programmable light curtain setups.

Light Curtains of Arbitrary Shape. To start, we use our system to simultaneously generate both a flat and vertical light curtain in Figure 4.7. Current prototype light curtain systems [12, 134] can only form one of the flat or vertical curtains, but not both at the same time. Our holographic illumination system has no such limitations, and supports placing bounding boxes around objects.

We generate a three-dimensional light curtain in Figure 4.4 of complex shape. In Figure 4.8, we use our system to selectively image objects in a scene, such that

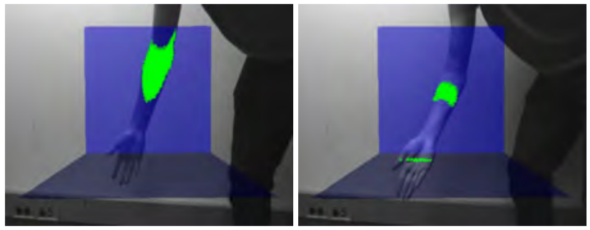


Figure 4.7: **Multi-orientation light curtains.** In this figure, we simultaneously place both a vertical and horizontal light curtain into the scene. Over the course of a single rolling-shutter exposure, our system is able to detect objects intersecting either of these curtains.

Figure 4.8: **Privacy-preserving imaging**. Similar to Ueda *et al*. [132], our system can selectively image user-specified 3D regions within the scene, like a teapot in this example. However, the advantage of our approach is that these regions can take on any 3D shape. The light from all objects not contained within this 3D region is optically filtered out, including the light reflecting off of a top secret document.

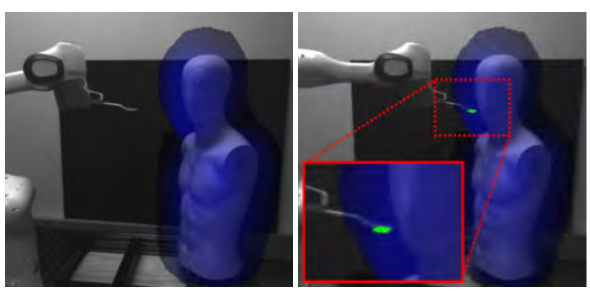


Figure 4.9: **Robot safety.** Our system can create form-fitting light curtains. Here, we create a light curtain $5\,\mathrm{cm}$ off the surface of a mannequin. Such a light curtain could be used in robot safety applications to detect if a robot gets too close to a person or delicate object. Feeding robots could also use these curtains as a cue for where to place a spoon, as shown above. Note that the visualization may be misleading - the handle of the spoon is in front, not behind the curtain.

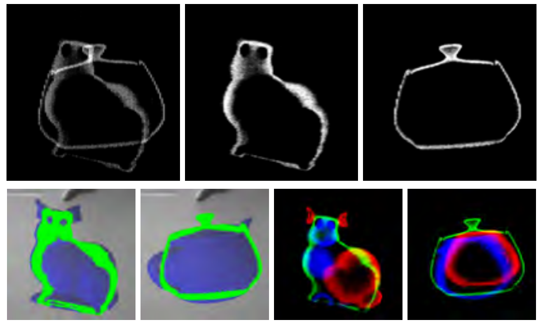


Figure 4.10: **Multiplexing light curtains.** Our system can simultaneously capture multiple light curtains by spatiotemporally multiplexing different curtains. **Row 1:** The demultiplexing process. In the leftmost raw image, even columns correspond to a bunny curtain while odd columns correspond to a teapot curtain. The separated curtains are shown in the middle and right images. **Row 2:** Sweeping a foam board through the curtains. Three positions of the board are represented by different colors in the two rightmost images.

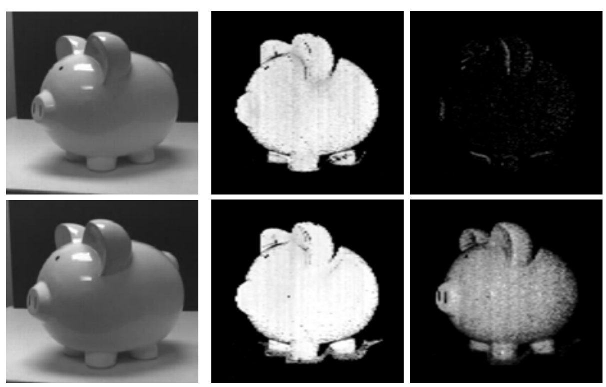
the confidentiality of a sensitive document is preserved. In particular, such light curtains could be useful for robot safety applications. For example, Figure 4.9 shows a form-fitting light curtain $5\,\mathrm{cm}$ off the body of a mannequin. A robot could use such a light curtain to detect whether it is too close to a person or object. For example, this could be used for assisted feeding, to aid those who cannot manage to feed themselves.

Disturbance Detection. A light curtain system can determine whether objects in a scene have been disturbed, by forming a tight light curtain over the surfaces of the scene. If no objects have been disturbed, our rolling-shutter camera records a bright image. In contrast, if objects have been moved, damaged, or dented, the measured intensity decreases in places where the object no longer lines up with the light curtain. Thus, we can extract a *disturbance map* by first imaging this light curtain when the scene is undisturbed, and subtracting the light curtain output after disturbance. In effect, this captures a difference image over a specific geometry of interest. We demonstrate an example of this disturbance detection in Figure 4.11. If any objects are disturbed, the recorded images registers a significant, dense change in contrast to the low-intensity, sparse output of a difference image. This idea may have important implications in manufacturing; for example, if mounted over an assembly line, our device could inspect whether objects passing through have defects.

Multiple Simultaneous Light Curtains. We show that our system can simultaneously generate and separate multiple light curtains within a single rolling-shutter frame. This is done by interleaving the patterns associated with two (or more) tar-

(a) Original scene (b) After the disturbance (c) Difference image (×2) (d) Disturbance map (×2)

Figure 4.11: **Disturbance detection.** Our system can optically detect if objects have been disturbed in a scene. In regions where little texture is present or the pixels are saturated or underexposed, a difference image provides low-intensity, sparse information on whether a scene is disturbed. However, a light curtain more accurately registers any change in surface geometry.



(a) Original scene (b) Thin curtain (c) Thick curtain

Figure 4.12: **Disturbance measurement.** Our system can optically detect the magnitude of a disturbance by multiplexing light curtains of different thicknesses. In this diagram, we show a thin curtain and a thick curtain illuminating the scene. **Row 1:** When the disturbance is small, while the thin curtain readily registers the disturbance, the thick curtain receives little signal. **Row 2:** In contrast, when the disturbance is large, both curtains easily detect the disturbance.

(a) Raw input (b) Visualization

Figure 4.13: **3D touch interface.** We form a light curtain about $2\,\mathrm{cm}$ off the surface of the object, which we can use to detect interactions with the object. By accumulating interactions over time, we can transform any object into a virtual drawing surface.

get light curtains. We show an example of such a setup applied to a toy scene in Figure 4.10.

To demonstrate why this might be useful, we show that the disturbance detection idea from the previous section can be combined with multiplexing to estimate the magnitude of a disturbance in Figure 4.12. Our system projects two curtains: a thin curtain and a thick curtain. If the thick curtain receives signal while the thin curtain receives no signal, the disturbance must be small. However, if both curtains receive no signal, the disturbance much be large.

Three-Dimensional Touch Interface. We also demonstrate an optical 3D touch interface in Figure 4.13(a). We form a light curtain about $2\,\mathrm{cm}$ above some desired surface. When a person's finger interacts with this surface, the light curtain detects its location. While the setup given in Tsuji *et al.* [130] was limited to planar surfaces, our system can turn any arbitrary geometry into a virtual touch interface. In the space of augmented reality, detecting where a user interacts with a scene could be used as a new input for art or entertainment applications. For example, as shown in Figure 4.13(b), our methodology can be used to turn any real-life object into a virtual drawing surface.

4.1.3 Summary

In this work, we demonstrated a binary holographic approach for structured lighting in the context of triangulation light curtains. Such a device based on a DMD is fast and light efficient enough to generate light curtains of arbitrary shape when combined with a rolling-shutter camera. We showed that such a system can be used to multiplex multiple curtains into a single image, optically detect disturbances and their magnitudes, and generate three dimensional touch interfaces. All of these applications are enabled by a holographic illumination system.

4.2 Limitations of binary holographic illumination

In practice, fast structured light systems via binary holography come with a few tradeoffs. First, the use of a binary SLM provides limited pattern quality, thanks to the requisite quantization. As we discuss in Chapter 6, artifacts manifest themselves in the form of speckle, which reduces the contrast of the generated pattern. This can result in noisy correspondences, causing a reduction in resolution.

Second, while such a binary holographic system is light efficient, the use of an amplitude SLM creates particular intricacies for such a system, that raise questions about the overall improvement in brightness. First, the underlying amplitude process must inherently block light to form desired pattern — in our experiments (Figure 4.3), we found that around 50% of the DMD pixels are turned off to form any desired pattern, resulting in 50% light loss. Second, a bright DC spot forms in the output, as the average amplitude modulation will always be strictly greater than 0; if about half of the DMD pixels are turned on, this point will use 25% of the total input light. Finally, because the modulation effectively produces a real-valued wavefront, the Fourier output pattern will always be conjugate symmetric. Thus, only one symmetric half of the output pattern is usually used, and the other half is blocked by an optical element resulting in one more 50% loss in the available light. In our work, we measured an overall loss of 92% in our real setup. Despite this, the light efficiency of a holographic system produced about a $2700\times$ improvement in brightness over a standard LCD projector for 1D patterns, visualized in Figure 4.1. In addition, many of the contributing factors for the light loss could be avoided by using a fast phase-only SLM, which are currently under development [13].

Chapter 5

Depth-programmable holographic illumination

The illumination systems used today for active sensing are not depth programmable. For instance, most standard projectors leverage lens systems, and therefore form desired content at only a single depth [144]. Content at other depths is defocused. While such systems are sufficient for simple tasks like projection onto a flat wall, this constraint can introduce various non-idealities for more complex settings. For one, if the target surface is not flat but curved, the projected pattern will never be totally sharp — in the best case, only part of the surface will be in focus, while the rest of the surface is out-of-focus [53]. When used as illumination for structured light, this effect reduces accuracy and performance because the resolution of output patterns is decreased in out-of-focus regions. For projection mapping configurations where dynamic content is projected onto non-flat surfaces for applications in art and entertainment, the visual quality of output patterns is similarly reduced.

How could one tackle these challenges? One powerful solution is a depth-programmable illumination system — a device that can program desired content as a function of depth. More broadly, not only could such a system tackle the depth-of-field challenges mentioned above, but it could also enable a number of new applications by projecting patterns that morph with distance. For instance, it could introduce new modes of dynamic projection mapping, where an object that moves with depth would automatically be illuminated with different patterns without requiring precise tracking [93]. In the context of human-computer interaction, such a system could be used to create novel screenless 3D interfaces, where the projector displays different buttons on objects at different depths. As part of a wearable gadget, such a device could project private content to just nearby objects like a user's arm, while farther objects only see a scrambled version. On the vision side, an appropriate depth-varying pattern could be used as a depth cue, without the

Technology	Programmability	HW complexity	Time-mult.	Light eff.
LCD/DLP [19, 101, 144]	Low (blur)	Low (1 SLM)	No	Low
Laser scanning [99, 130, 132]	Low (minimal blur)	Low (scanning mirror)	No	High
Focus-tunable lens [141]	Moderate (crosstalk)	Moderate (SLM+tun. lens)	Yes	Depends
Coded aperture [53, 65, 66]	Moderate (conv. with aper.)	Low (SLM+static ap.)	No	Low
Time-mult. coded ap. [57, 83]	High	High (2 SLMs)	Yes	Low
Holographic [21, 31, 86, 87, 118, 128]	Moderate (limited étendue)	Low (SLM+laser)	No	High
Holo. + lens array (ours, [25, 91] in NEDs)	High	Low (Holo.+lens arr.)	No	High

Table 5.1: **High-level comparison of different potential depth-varying illumination systems.** A holographic approach with a lens array étendue-expander allows for depth-dependent content to be programmed on a single SLM, without needing time multiplexing or light loss. Thus, our work uses such a system to engineer a depth-varying projector.

need for a stereo baseline.

However, existing devices that could tackle these applications come with various tradeoffs. Coded aperture [53, 65, 66, 83, 140] and light field projectors [57] are limited in programmability, and struggle to form arbitrary content at different planes (Figure 5.1(a)), *e.g.*, coded apertures are restricted to an intensity convolution between an in-focus image and a scaled version of the aperture pattern [53]. Temporal multiplexing can improve the degrees of freedom [57, 83], but at the cost of lower framerates, increased bandwidth requirements and the need for multiple spatial-light modulators (SLMs). Furthermore, light is inherently blocked to form desired patterns, reducing output brightness. Another possibility is multiple overlapping projectors focused at different depths [14, 92], but such systems again require multiple SLMs, increasing cost and form factor. Alternatively, a fast focus-tunable lens that is synchronized with a high-speed projector [135, 141] could be used to temporally multiplex patterns focused at different depths. While theoretically effective, current focus-tunable lenses are limited in aperture size, increasing crosstalk between different depths (Figure 5.1(b)). Simultaneously, producing content at multiple depths requires a high level of temporal multiplexing, again decreasing framerate and increasing bandwidth requirements. These architectures are summarized in Table 5.1.

To address these limitations of existing systems, we propose the use of holographic illumination for building a depth-varying projector. In theory, we can build a system that programs arbitrary patterns at arbitrary depths by simply solving a variant of Equation (2.4) with the right patterns:

$$\min_{\phi} \sum_{d \in D} \left\| T_{\text{target}}(d) - |\mathcal{P}_{\text{ASM}}(\mathcal{F}\{\mathcal{S}(\phi)\}, d)|^2 \right\|^2, \tag{5.1}$$

where the optimization instead aims to reproduce a stack of unique targets at different depths. A projection lens then images these outputs into the scene as before. Such an approach does not require time-multiplexing unlike traditional incoherent projector systems, as coherent holographic systems have more degrees of freedom [102]. While such configurations have been explored for applications in building true 3D holographic displays [77, 85, 133, 136, 142, 143], no past work has applied these ideas for building projector systems.

5.1 Holodepth: practical depth-varying holographic illumination

In practice, naïvely rescaling a holographic setup to the field-of-view expected of a projector results in limited depth variation. In fact, forming unique content at two depths requires that these depths are separated on the order of meters [142], making the desired applications impractical on holographic systems. In our experiments on a typical SLM, we find that achieving a useful amount of depth variation necessitates a large focal length projector lens, which results in a tiny field-of-view.

In this section, we explore one potential approach for tackling this problem, and its applications to real systems. First, we derive that this relationship between depth variation and field-of-view is fundamentally connected to the *étendue* of a holographic system — a measure of light spread over area and angle. Inspired by research that tries to tackle the field-of-view and eyebox tradeoff of near-eye displays [25, 74, 76, 91, 129], we introduce a lens array into the optical path of a standard holographic illumination system to increase depth variation, enabling far more complex patterns than past work in true 3D displays [142]. We develop a novel optimization-driven calibration for this optic, that tackles challenges like misalignment and aberration that past approaches ignore. With these modifications, we realize a practical, high-resolution programmable depth-varying projector which we term "Holodepth", that achieves all of the aforementioned applications in a single setup.

In summary, we will discuss:

- A proof-of-concept system for depth-varying holographic illumination, with étendue expanded by a lens array;

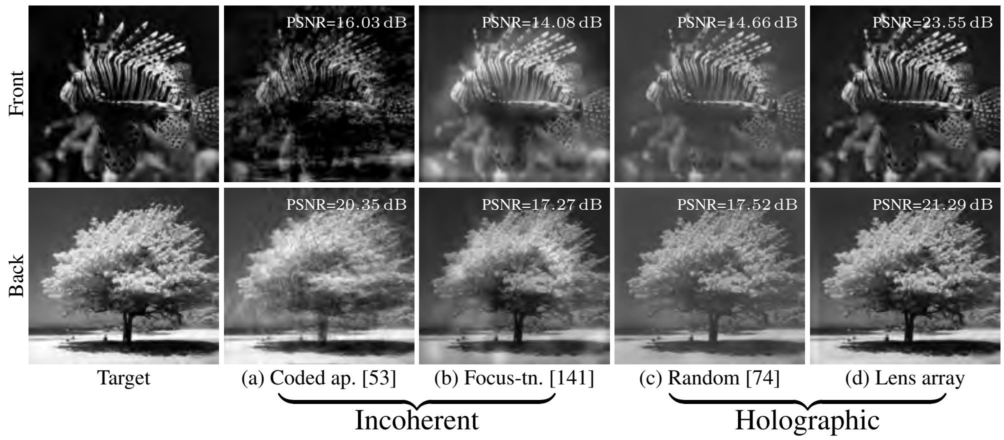
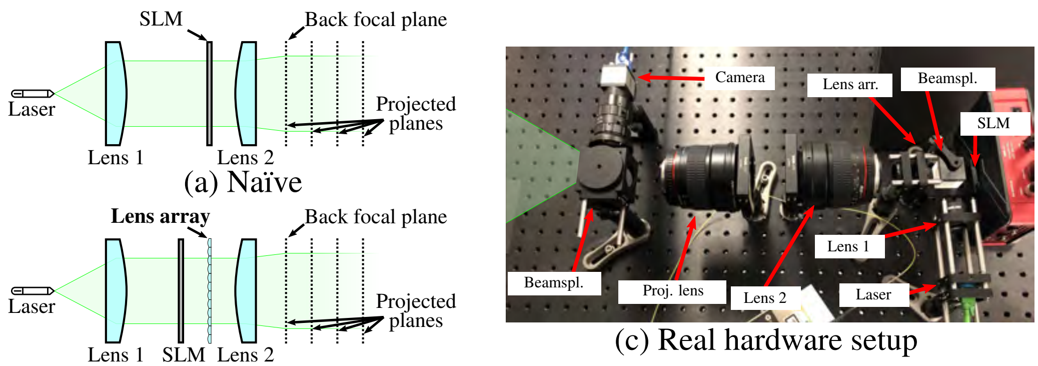


Figure 5.1: **Simulated comparison of depth-varying illumination solutions.** To compare with other potential configurations, we simulate the projection of unique content at two planes. **(a)** A coded aperture setup with two SLMs [53] struggles to form the desired content. **(b)** A time-multiplexed focus-tunable lens and high-speed projector does better visually, but low frequency errors occur thanks to crosstalk on top of the practical challenges associated with time-multiplexing. **(c)** A holographic system that uses a custom random binary phase mask [74] results in loss of contrast. **(d)** The off-the-shelf lens array used in this work performs similarly to the time-multiplexed case with just a single SLM pattern.

- An analysis of the effects of using such a lens array approach;
- An optimization-driven calibration process for this étendue expander;
- A demonstration of multi-plane projection for 3D interfaces, privacy, multi-layer displays, defocus compensation and artistic applications; and
- An exploration of holographic depth-varying projection as a depth cue, for which we show techniques for depth capture and light curtains.

5.1.1 Related work

Our work is most analogous to past research in building 3D holographic displays. Recent near-eye display research [26, 27, 29, 32, 33, 64, 75, 84, 102, 104, 116, 117] has shown impressive results using holography to produce 3D accommodation cues, replicating depth maps and focal stacks with the target of realistic defocus — although inspiring, realistic defocus is not our goal, and can be overly restrictive for our applications. Our work is most similar in spirit to true 3D holographic displays, where independent control of every 3D point is desired — however, most such sys-



(a) Naïve

(b) Étendue-expanded

(c) Real hardware setup

Figure 5.2: **System diagram.** **(a)** Optical diagram of a naïve approach. **(b)** Optical diagram of an étendue-expanded version with an additional lens array. **(c)** Our real hardware prototype.

tems are also limited in depth variation, making them unsuitable for a projector. Most work in this space focuses on computational methods to improve the quality of phase retrieval [85, 133, 143]. Time multiplexing can be applied to better disambiguate content at different depths [77, 136], but it does not fundamentally improve depth variation. Most akin to our real system, Yu *et al.* [142] introduce a thick scattering layer to increase axial resolution, but their approach requires custom optics, careful interferometric calibration of a large lookup table and low resolution simple patterns with < 100 sparse points. In contrast, our system uses simple optics that can be easily calibrated in-setup with our proposed approach, and we demonstrate it on much more complex megapixel patterns. More generally, our work shows that the 3D capabilities of holographic displays are extremely useful for novel projector systems with the right modifications, and expands the use cases of these holographic devices.

5.1.2 Étendue-expanded depth-varying projection

In practice, current holographic systems are limited in *étendue*. Defined as the product of the system's spatial area A and the solid angle of emitted light, the étendue $\mathcal{E}$ of an SLM with $N_x \times N_y$ pixels of pitch δ can be calculated as [74]:

$$\sin\theta = \frac{\lambda}{2\delta}, A = \delta^2 N_x N_y, \mathcal{E} = 4A\sin^2\theta = \lambda^2 N_x N_y, \tag{5.2}$$

where θ is the maximal tilt angle of the SLM. In recent work, étendue has been extensively studied for near-eye displays, where it results in a tradeoff between

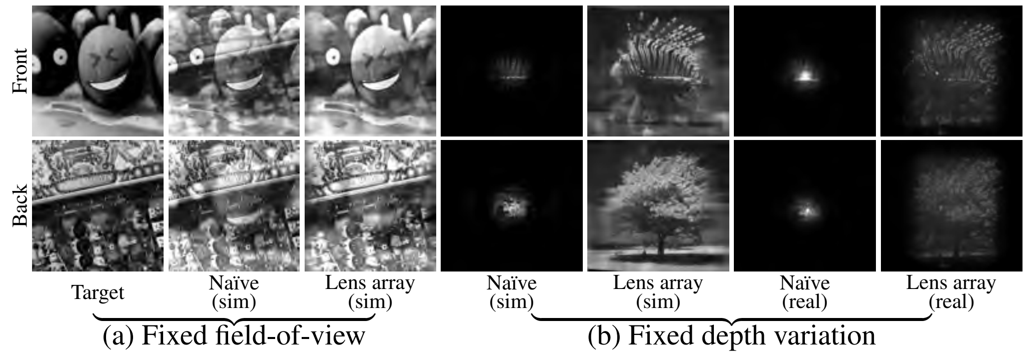


Figure 5.3: **Étendue expansion with a lens array.** We visualize the importance of étendue for depth-varying projection. **(a)** For a given field-of-view, without étendue expansion, there is a significant amount of crosstalk between two target depth planes. With an étendue-expanding lens array, the quality of projection is much higher. **(b)** For a given depth variation, the effective field-of-view is very small with a typical SLM. With étendue-expansion, the field-of-view is much larger.

field-of-view and eyebox size. In short, maximizing the area that the human eye can move and see an image from a holographic display minimizes the size of that image and vice versa, limiting the practicality of holographic near-eye displays. A number of approaches have been proposed to increase the étendue of near-eye displays. In one line of work, multiple laser sources are time-multiplexed to stitch together a larger FOV [76], but such an approach requires an extremely fast SLM which is not currently readily available [91]. Instead, recent work has focused on adding static high-resolution phase masks into the optical path, that effectively spatially multiplex the SLM over a larger FOV [25, 74, 91, 129]. Lens arrays serve as a low-cost alternative [25, 91] to these custom-fabricated phase masks [74, 129].

For holographic illumination, étendue effectively presents a similar tradeoff between depth variation and field-of-view. To quantify this, we can calculate the rate of change of a projected pattern with depth by calculating $\partial\{|\mathcal{P}(U,z)|^2\}/\partial z$:

$$\iint_\Omega \iint_\Omega \left(\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2} - \sqrt{1-(\lambda a_x)^2-(\lambda a_y)^2}\right) \cdot P^*(f_x, f_y, x, y, z) P(a_x, a_y, x, y, z) \mathrm{d}f_x \mathrm{d}f_y \mathrm{d}a_x \mathrm{d}a_y \tag{5.3}$$

where $P(\cdot)$ contains all terms inside the integral of Equation (2.3). Since large differences between $f_x^2 + f_y^2$ and $a_x^2 + a_y^2$ are weighed more heavily than small differences in this expression, more depth variation can be achieved by stretching the Fourier spectrum S of a wavefront, as this increases the maximum $f_x^2 + f_y^2$. In the setup given by Equation (5.1), this operation is directly equivalent to increasing SLM area. However, naively using a lens to increase the size of an SLM also increases pixel size, resulting in a smaller field-of-view. Thus, like the eyebox in near-eye displays, depth variation suffers from a fundamental étendue tradeoff with field-of-view. As a result, on modern SLMs, holographic illumination cannot project patterns that significantly vary in depth without an excessively small FOV. For a 1920×1200 resolution SLM with $8\,\mu\mathrm{m}$ pixel pitch, we find that projecting unique content spread $15\,\mathrm{cm}$ apart leads to a projector FOV of about $3.8°$.

Thus, we need to expand étendue for a practical depth-varying projector. For simplicity, we opt for a static element, like a custom phase mask [74, 129] or lens array [25, 91]. In our simulations (Figure 5.1(c),(d)), we find that a lens array preserves more contrast than a random phase mask [74] for depth-dependent content. Thus, we place a lens array into our optical system at the front focal plane of Lens 2, and move the SLM forward some distance z_{array} as shown in Figure 5.2(b). Our étendue-enhanced forward model can then be written as:

$$\mathcal{P}_{\text{enhanced}}(U, z) = \mathcal{P}_{\text{ASM}}(\mathcal{F}(\mathcal{M}(\mathcal{P}_{\text{ASM}}(U, z_{\text{array}}))), z), \tag{5.4}$$

where $\mathcal{M}$ denotes the transformation imposed by the lens array. We can then find the best-matching SLM pattern for some target depth-varying pattern using Equation (5.1). With an array with lens pitch $1.0\,\mathrm{mm}\times 1.4\,\mathrm{mm}$ and focal length $4.7\,\mathrm{mm}$, we can expand FOV to about $15°$. We visualize the results in Figure 5.3.

Lens array effects

In the context of near-eye displays, Monin *et al*. [91] showed that the spatial multiplexing performed by static étendue-expanding optics, like a lens array, reduces either output contrast or resolution. For depth-varying projectors, this spatial multiplexing additionally manifests itself in the form of a *structured defocus pattern*. Intuitively, a given output point only receives light from a subset of the SLM thanks to the multiplexing. Since the SLM controls the angular distribution of light, the defocus pattern will be structured according to the multiplexing of the SLM. While typically imperceptible in practice, the contrast of depth variation is reduced when compared to a larger SLM with equivalent étendue. This structured defocus is illustrated in Figure 5.4. Practically, when projecting natural images, this effect reduces the contrast of depth variation when compared to a larger SLM with equivalent étendue, as shown in Figure 5.5.

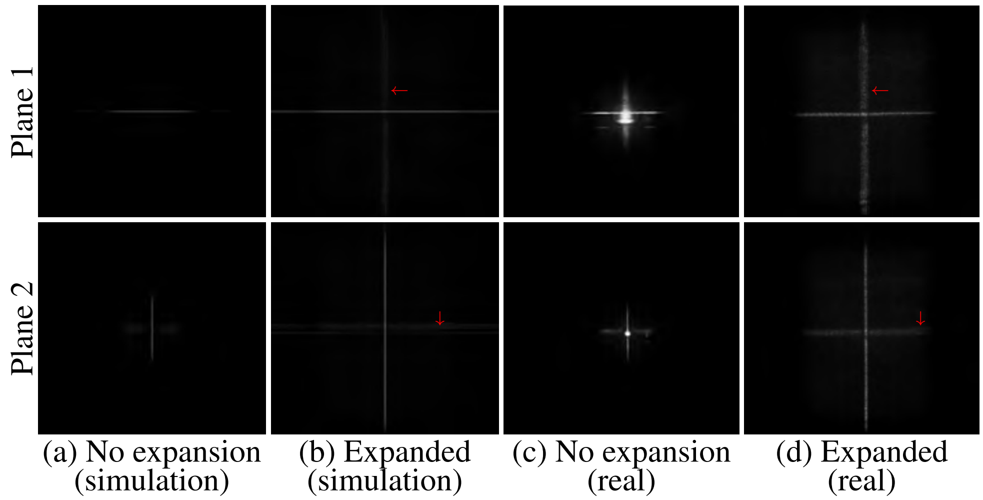


Figure 5.4: **Structured defocus from a static étendue expander.** We visualize the defocus pattern created by a system with étendue expanded by a static element, like a lens array [25, 91] or phase mask [74, 129]. While the defocus pattern stays the same size, it becomes more structured, as shown by the faint lines that appear **(b)**, **(d)**.

Calibrating the lens array

In practice, aligning étendue-expanding optics can be a challenging process, as tiny axial ($<1\,\mathrm{mm}$) and lateral misalignments ($\approx$ 1 SLM pixel) can cause a drastic reduction in output quality (Figure 5.6(a)). Existing approaches have typically reduced SLM pixel resolution and introduced precise alignment processes to mitigate these effects [74, 129]. Additionally, optical aberrations further reduce quality [25], but past work has primarily ignored these effects, assuming ideal lenses and masks [25, 74, 129]. Such artifacts are exacerbated in low cost off-the-shelf optics, like the lens array we use in Section 5.1.3.

In our work, we instead approximately align the lens array in the system, and then apply an optimization procedure to reconstruct the unknown lens array modulation $\mathcal{M}$ using a dataset of SLM patterns and real projections captured by a camera z_{calib} away. We can also simultaneously calibrate other non-idealities like SLM distortion and undiffracted light. Formally, this process is given by:

$$\min_{\substack{\mathcal{M}, A_{\text{cam}}, A_{\text{add}}, A_{\text{SLM}}, \\ \mathcal{P}^{\text{opt}}_{z_{\text{calib}}}, \mathcal{P}^{\text{opt}}_{z_{\text{array}}}}} \quad \sum_k \left\| I_k - \left|\mathcal{P}_{\text{calib}}(\mathcal{S}(\phi_k))\right|^2 \right\|^2, \tag{5.5}$$
$$\mathcal{P}_{\text{calib}}(U) = \mathcal{P}_{\text{ASM}}\left(A_{\text{add}} + A_{\text{cam}} \cdot \mathcal{F}\left(\mathcal{M}\left(\mathcal{P}^{\text{opt}}_{z_{\text{array}}}(A_{\text{SLM}} \cdot U)\right)\right), z_{\text{calib}}\right),$$

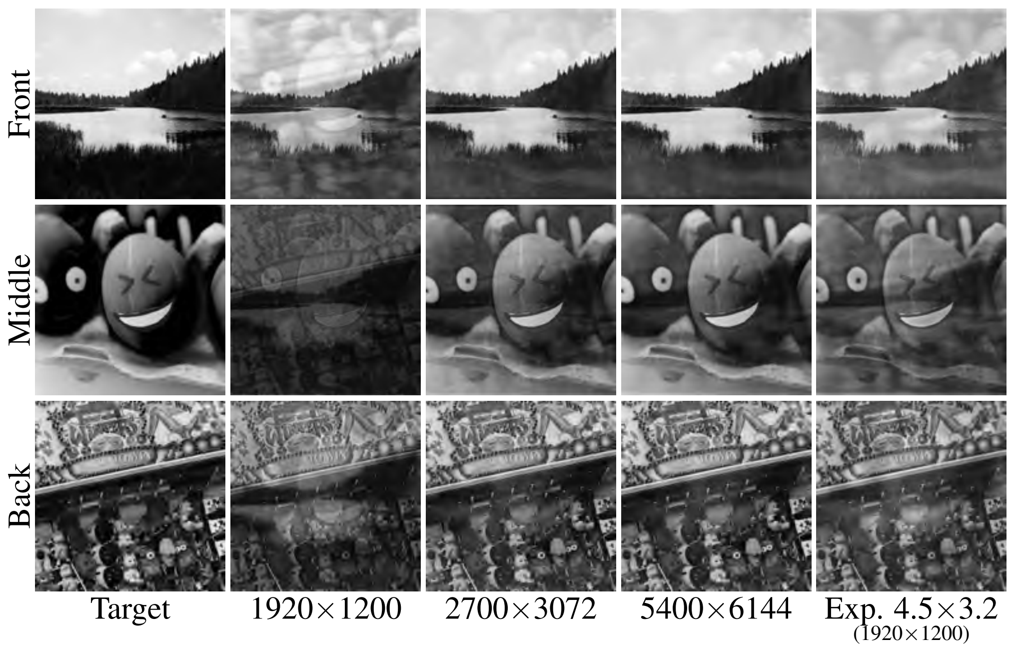
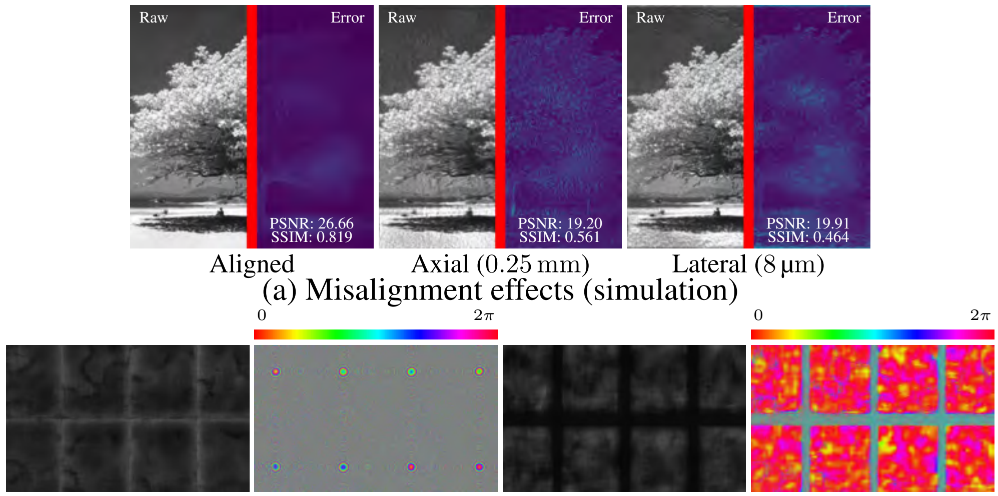


Figure 5.5: **Comparing depth variation of the étendue-expanded system with larger SLMs.** The structured defocus introduced by the static element reduces the quality of depth variation. In our system, a lens array expands the étendue of a $8\,\mu\mathrm{m}$ pixel pitch SLM by 4.5×3.2, creating the same field-of-view as a $1.78\,\mu\mathrm{m}\times 2.5\,\mu\mathrm{m}$ pitch SLM. Ideally, for the same fixed field-of-view, the depth variation of a system with étendue expanded by 4.5×3.2 should be similar to an SLM with 4.5×3.2 times the pixels, *e.g.*, 1920×1200 $\rightarrow$ 5400×6144. However, its depth variation is more similar to that of a SLM with 2700×3072 pixels.



Aligned Axial (0.25 mm) Lateral (8 µm)

(a) Misalignment effects (simulation)

$|A_1|$ $\angle A_1$ $|A_2|$ $\angle A_2$

(b) Calibrated parameters

Figure 5.6: **Calibrated lens array modulation. (a)** Tiny array misalignments cause contrast loss and distortion, demonstrating the need for careful calibration. **(b)** We show a crop of the calibrated parameters from Equation (5.5). A_1 learns the phase of the lenses, while A_2 seems to learn other distortions.

where ϕ_k and I_k are the corresponding SLM pattern and capture pair. A_{cam}, A_{SLM} and A_{add} denote learnable complex modulations, and $\mathcal{P}_z^{\text{opt}}$ denotes an optimizable propagation kernel initialized with the kernel for propagating z [33, 64]. Inspired by the ABCD matrix for thick lenses, we represent $\mathcal{M}$ as an optimizable multiplication, propagation and multiplication $\mathcal{M}(U) = A_2 \cdot \mathcal{P}^{\text{opt}}_{z_{\text{thickness}}}(A_1 \cdot U)$, where $z_{\text{thickness}}$ is the estimated thickness of the lens array.

Using Adam [71], we run the above optimization over a dataset of 17340 SLM pattern/captured pairs of sparse and natural targets, and visualize the calibrated A_1 and A_2 in Figure 5.6(b).

Comparison with alternative architectures

Beyond Figure 5.3, more simulated comparisons between our proposed Holodepth system and traditional approaches that could create depth-varying illumination are shown in Figure 5.7. A holographic system with étendue expanded with a lens array outperforms a coded aperture projector[1], a high-speed projector focused at two

[1]Simulated with a 50mm aperture. Image at one plane is given by the in-focus pattern, while image at the other plane is given by its convolution with a pixel-resolution aperture pattern. Both the in-focus image and the aperture pattern are optimized.

Figure 5.7: **More simulated comparisons of depth-varying projectors.** Like Figure 5.3, unique content is projected to two planes. **(a)** Coded aperture projectors [53], **(b)** focus-tunable lens/high-speed projectors [141], and **(c)** naive holographic illumination systems struggle with crosstalk. **(d)** A holographic system with étendue expanded by a random phase mask [74] performs well visually, but low frequency errors significantly decrease PSNR. **(e)** Holographic illumination with étendue expanded by a lens array provides the best results, and we leverage such a setup for our real world prototype.

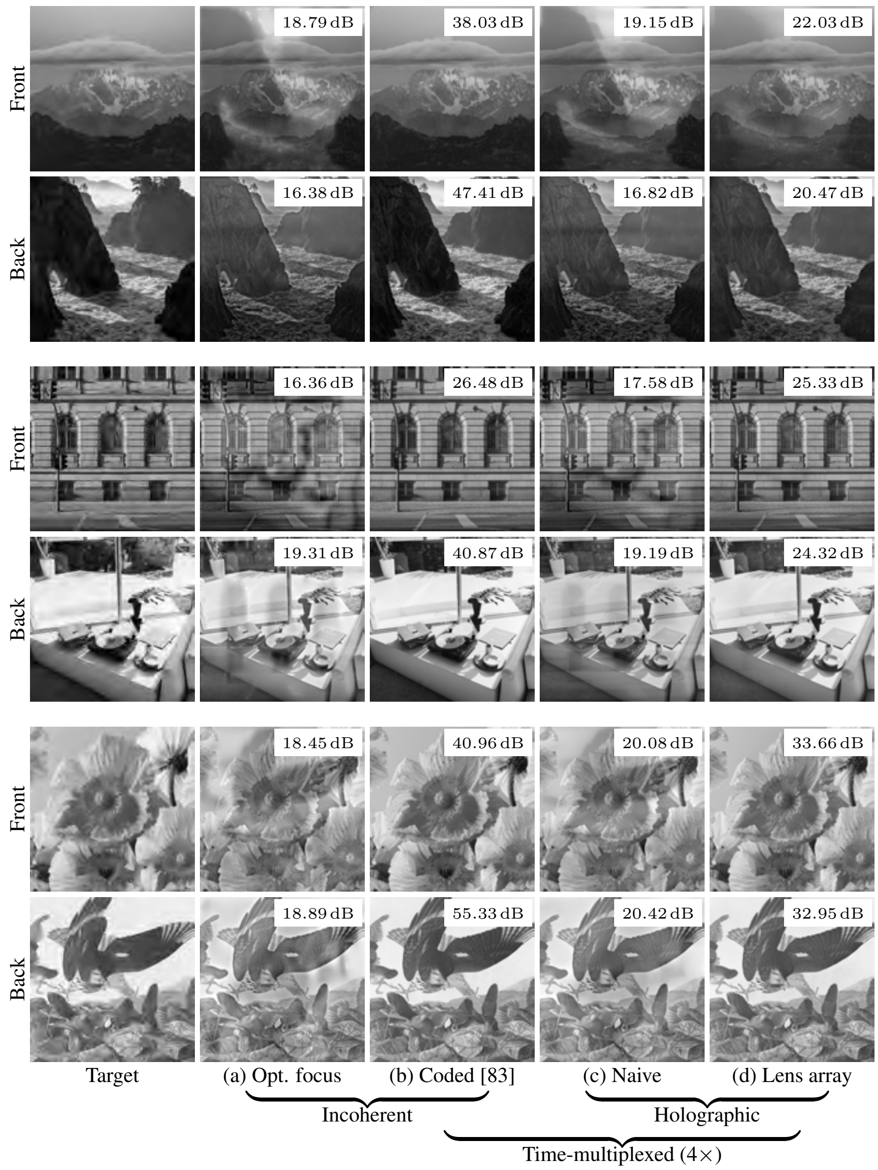


Figure 5.8: **More complex depth-varying projectors.** **(a)** We show an optimized version of a focus-tunable lens/high-speed projector, where the patterns projected for each distance are jointly optimized. This increases performance over Figure 5.7(b), but still does not match Figure 5.7(d). **(b)** We demonstrate a 4× time-multiplexed coded aperture projector [83], which significantly increases performance over the single aperture/display pattern case. **(c)**, **(d)** We demonstrate 4× time-multiplexed versions of a naive holographic system and an étendue-expanded version with the same field-of-view. Time multiplexing mildly improves performance for these holographic systems.

depths with a focus-tunable lens[2], a naive holographic illumination system rescaled to the same field-of-view, and a similar holographic system with a random phase étendue expander.

In Figure 5.8, we show more sophisticated projection systems. In Figure 5.8(a), we show a modified version of a focus-tunable lens and high-speed projector system, where the patterns projected for each focal length are jointly optimized to match the target images. This slightly improves the projection quality, but performance still does not match the case of a holographic system with étendue expanded by a lens array. In Figure 5.8(b), (c), and (d), we explore time-multiplexed versions of coded aperture projectors [83] and holographic systems. Time-multiplexing significantly improves the performance of a coded aperture approach as illustrated in Figure 5.8(b). The improvement is less pronounced for holographic systems as shown in Figure 5.8(c) and (d) — we attribute this to limited étendue even after expansion. While time multiplexing is a useful tool, it can be impractical in real applications thanks to increased bandwidth requirements and decreased framerates. Furthermore, a time-multiplexed coded aperture requires an additional SLM, increasing form factor and cost. Like standard coded aperture, projected patterns will also be much darker than a holographic approach thanks to the requisite blocking of light.

5.1.3 Hardware implementation

We show our hardware prototype in Figure 5.2(c). Our laser is a $530\,\mathrm{nm}$ Coherent Sapphire LPX. We use a Thorlabs Exulus HD-2 SLM, with 1920×1200 resolution and $8\,\mu\mathrm{m}$ pitch. For étendue expansion, we use a Thorlabs PMMA Microlens Array (MLA1M), where each microlens has pitch $1.0\,\mathrm{mm}\times1.4\,\mathrm{mm}$ and focal length $4.7\,\mathrm{mm}$. Simulating each microlens without aliasing requires a resolution of $2.5\,\mu\mathrm{m}\times1.78\,\mu\mathrm{m}$, translating to an expansion in étendue by 3.2×4.5. We therefore simulated our system at 9600×6000 with $1.6\,\mu\mathrm{m}$ pitch. We used a $75\,\mathrm{mm}$ achromatic doublet for Lens 1, and two $85\,\mathrm{mm}$ $f/1.4$ DSLR lenses for Lens 2 and a projection lens that magnifies the patterns into the scene. We use a beamsplitter to illuminate our reflective SLM along the optical axis. Using another beamsplitter, we colocate with the projector a UI-3240CP-NIR camera with a $16\,\mathrm{mm}$ $f/1.8$ lens. The captured scenes are between $0.4\,\mathrm{m}$ and $0.7\,\mathrm{m}$ away from the system. At $0.7\,\mathrm{m}$, the projected pattern spans roughly $0.1\,\mathrm{m}\times0.15\,\mathrm{m}$.

[2]Simulated with a $10\,\mathrm{mm}$ aperture following recent work [141].

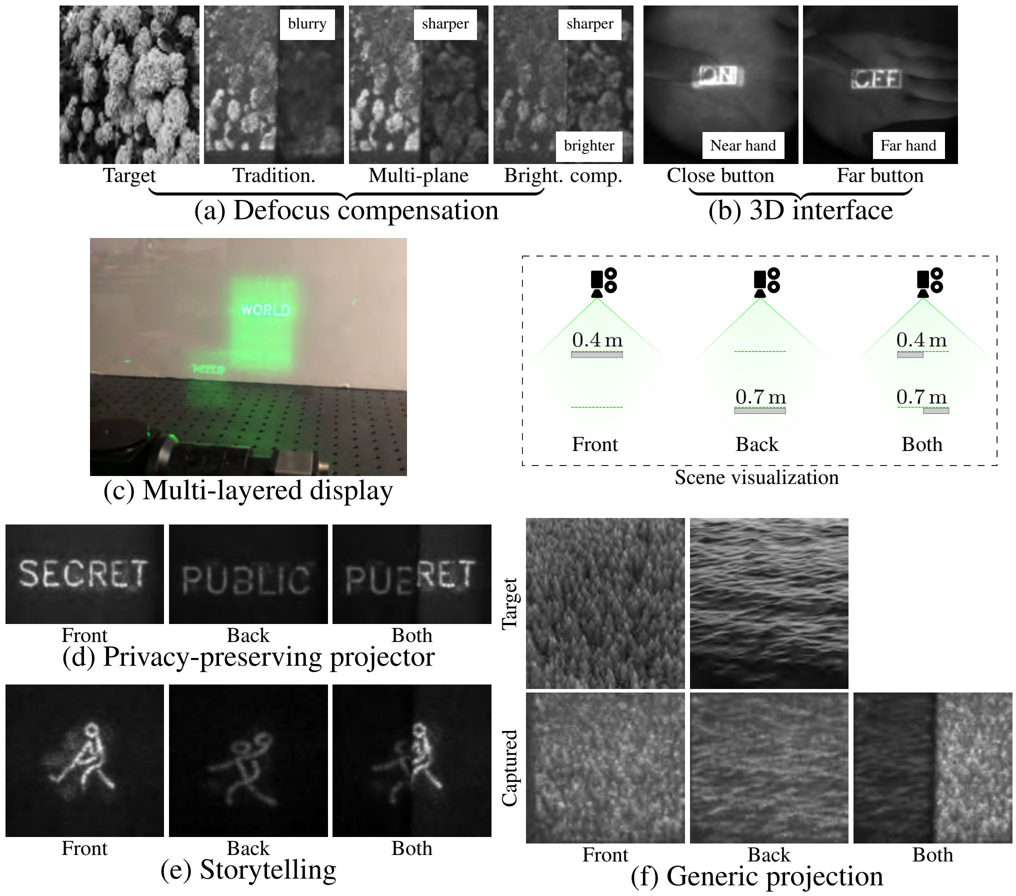


Figure 5.9: **Programmable depth-varying illumination.** Our depth-varying patterns can compensate for defocus, and they facilitate new modalities like 3D depth-varying interfaces, multi-layered displays, privacy-preserving projection, and other artistic avenues.

5.1.4 Applications

Multi-plane projection

Our Holodepth system can simultaneously project different content to multiple planes. As shown in Figure 5.9(a), such a system could be used to potentially address defocus in a scene with multiple depths. In addition, it can compensate for radiometric falloff as we discussed in Chapter 3, equalizing brightness without blocking light like a traditional projector.

Our system could also be used to enable novel user interfaces. For instance, projectors are being implemented into wearables for screenless computing [61], where users interact with the projection onto another object. As shown in Figure 5.9(b), our system could be used to create a form of 3D depth-varying interface for such a wearable, by projecting a different button to different depths. Such a device could also be applied to multi-layered displays [11] for the formation of pseudo-3D content (Figure 5.9(c)). We place a piece of translucent acrylic at one plane, and a white board at the other. Different content can be projected at each plane, both of which are visible to a viewer. Our device could also ensure that private content is only displayed on nearby objects, while different content is shown on farther objects that other people can see (Figure 5.9(d)).

We envision such a projector as a useful tool for artists and creatives (Figure 5.9(e)). An artist could use such a device to show different content depending on the location of a reflecting object, in order to interactively tell a story.

Depth estimation

Holographic depth-varying content can also be used as a depth cue. Consider two patterns formed at two planes. In between these planes, the intermediate pattern may provide enough of a cue to disambiguate the exact location between these two planes. We show an example in Figure 5.10(a), where an ‘X’ and ‘O’ are formed at two different planes. The pattern uniquely evolves between them with depth. We leverage this in the form of a simple depth recovery method, where we project different patterns at different planes, calibrate the intermediate patterns, and reconstruct depth from a camera image similar to [66, 115].

Let p_z be the projected pattern that appears at depth z, and i the captured image. Then, our goal is to find the p_z seen at every camera pixel. This is generally a hard problem, especially for sharp gradients in albedo and depth. If we assume spatial smoothness, we can use graph cuts to recover depth [115]:

$$E(z) = \sum_{\mathbf{x}} D\left(z_{\mathbf{x}}\right) + \lambda \sum_{\mathbf{x},\mathbf{x}'} V_{\mathbf{x},\mathbf{x}'}\left(z_{\mathbf{x}}, z_{\mathbf{x}'}\right) \tag{5.6}$$

where $z_{\mathbf{x}}$ is the depth estimate for pixel $\mathbf{x}$, $D(\cdot)$ penalizes texture mismatch between i and p_z, $V_{\mathbf{x},\mathbf{x}'}$ penalizes variations in depth between neighboring pixels, and λ is a weighting term. For $D(\cdot)$ between i and p_z, we run template matching on a small neighborhood around $\mathbf{x}$ using the normalized sum-of-squared differences. We can minimize $E(z)$ using any graph-cuts solver and estimate depth.

Armed with this simple reconstruction algorithm, we can run more sophisticated procedures. For instance, instead of projecting all of the ‘X’s at one plane and ‘O’s at another, we can instead project them at different planes in a spatially-varying fashion. For instance, if we have some prior that the left side of the scene is closer than the right side, we can place our ‘X’s and ‘O’s appropriately for each half,

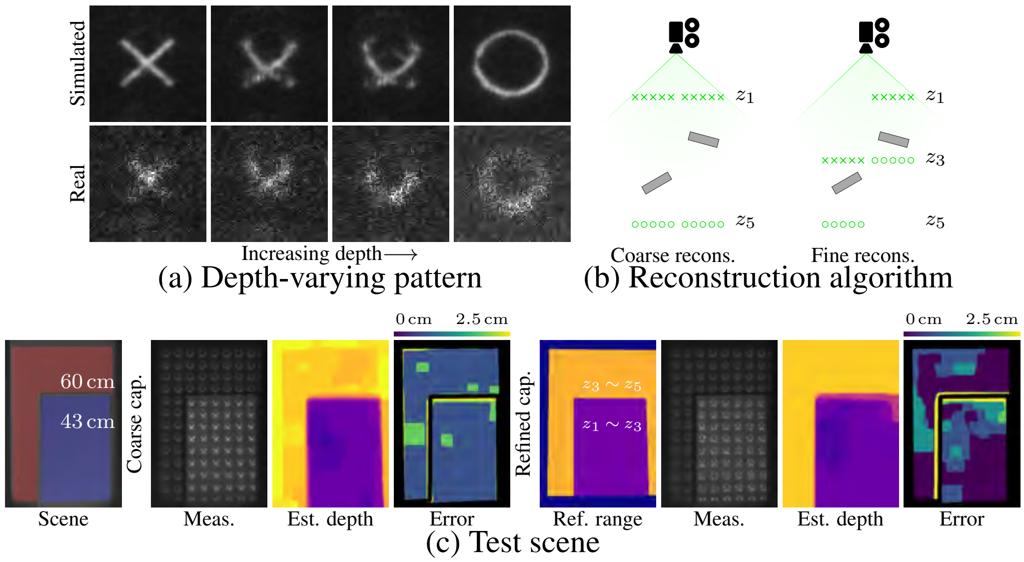


Figure 5.10: **Depth recovery from a depth-varying pattern.** The depth-varying pattern from our system can be used to recover depth when captured by a camera. **(a)** We plot the depth-varying pattern from our system when an 'X' and 'O' are projected at different planes. While there is a fair sim-to-real gap, the real-world pattern still significantly varies with depth. **(b), (c)** We show our simple depth reconstruction algorithm. We first project content at two planes spread far apart, to get a coarse estimate of depth. Once we have a rough estimate, we then move these two planes closer together, centered around the coarse depth, to get a finer depth measurement.

such that we get higher resolution for both sides. More concretely, consider the case where our projector can project content $I_1, \ldots I_n$ at a planes $z_1, \ldots z_n$. Our depth-varying pattern is defined by projecting o_1 at one plane and o_2 at another. For each pixel $\mathbf{x}$, we have some prior that its true depth z_t follows the constraint $z_i < z_t < z_j$, where i and j are two of the n planes our system can project to. Then, for that pixel, we set $I_i(\mathbf{x}) = o_1(\mathbf{x})$ and $I_j(\mathbf{x}) = o_2(\mathbf{x})$. We run this process for every pixel, and then plug the resulting $I_1, \ldots I_n$ in Equation (5.1) to determine the appropriate SLM pattern. We can then use the resulting capture in Equation (5.6) to get a refined depth measurement.

In Figure 5.10(b), we demonstrate the above algorithm with $n = 5$ planes. We first capture a coarse measurement by placing o_1 and o_2 at I_1 and I_5. With coarse depth, we determine z_i and z_j for each pixel, using which we generate a new pattern to get a more accurate depth measurement. We constrain $j = i + 2$ to handle the uncertainty of the previous measurement. In Figure 5.10(c), we

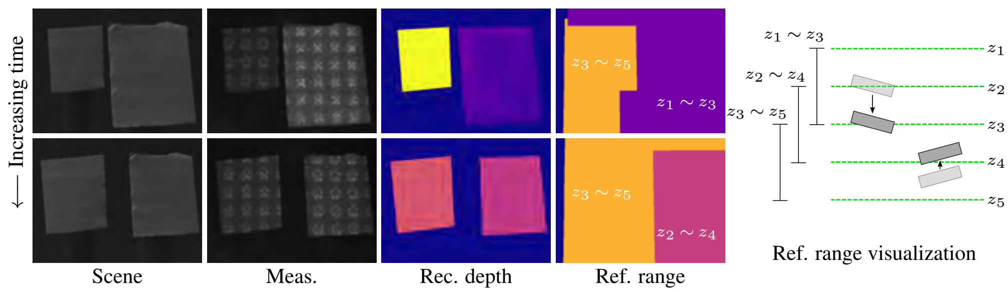


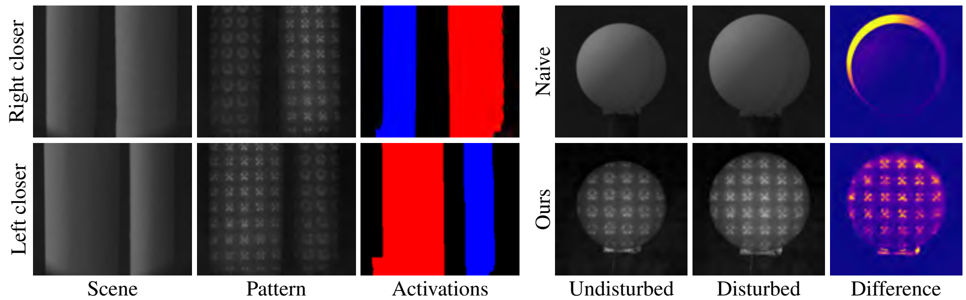
Figure 5.11: **Adapting the depth-varying pattern for moving objects.** Instead of using a coarse capture, we can use a previous depth estimate in the case of two moving objects. The left object moves towards the projector, while the right object moves away. Each object is captured with a different set of two planes, which is adjusted as the objects move. The rightmost column illustrates these planes.

apply this procedure to a test scene, where the refined measurement reduces error. In Figure 5.11, instead of relying on an initial coarse capture, we iteratively apply our methodology to a moving scene, where one object moves towards the system while the other object moves away. z_i and z_j shift as they move.

Programmable light curtains

Our depth-varying patterns can also be used to form programmable light curtains around objects [12, 130, 132, 134], as discussed in Chapter 4. To create a light curtain, we first place a calibration object in the scene, and capture an image of that object with a depth-varying pattern projected onto it. To determine if an object intersects the curtain, we take the difference between the calibrated capture and the current capture, and apply an activation threshold. We bilateral filter all images to reduce the effect of speckle. Note that unlike a traditional triangulation light curtain [12, 130, 132, 134], this procedure does not require stereo calibration or a baseline, and complex, multi-layer curtains can be easily formed with a single SLM pattern. In addition, this approach does not require the careful synchronization of complex camera optics that typical devices require. Furthermore, our light curtains are visible to the naked eye, unlike past approaches that rely on a scanning camera. This could potentially enable new applications like augmented sculpting, where an artist could directly identify if they carved an object correctly by looking at the texture formed on the object.

We show simple examples of our curtains in Figure 5.12. Our system can project two planar curtains simultaneously, and detect when two cylinders intersect these curtains. Our system can also be applied to disturbance detection, following Chapter 4. The change in projected patterns provides a strong cue as to when objects are



(a) Two simultaneous light curtains (b) Disturbance detection

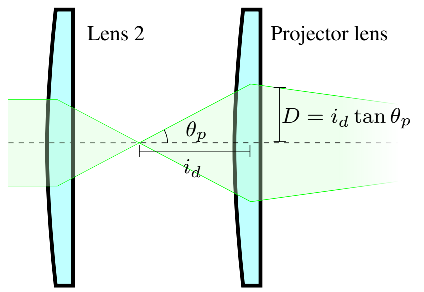
Figure 5.12: **Using depth-varying patterns to form coarse light curtains. (a)** We form a closer light curtain (red) and a farther planar curtain (blue). Our system is able to disambiguate which curtain is intersected with. **(b)** We use our system to detect disturbances in an object. When the object shifts axially, a traditional capture yields little information, but our depth-varying texture reveals a large change.

moved.

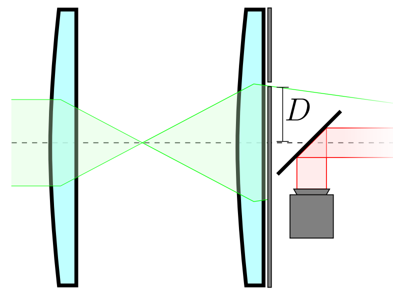
Depth cue comparison

Using system parameters and geometric optics approximations, we can approximately determine the resolution of our baseline-free holographic depth variation cue, as illustrated in Figure 5.13. First, we can compute the resolution of the wavefront after Lens 2 with focal length f_2 as $\frac{\lambda f_2}{\delta N}$, producing maximum diffraction angle $\sin\theta_p = \frac{\delta N}{2 f_2}$. This wavefront is then projected into the scene by a projection lens with focal length f_p. Then, for a point p_d that is a distance d from this projection lens, we can calculate the radius of the effective aperture from which this point receives light as $i_d \tan\theta_p$, where $\frac{1}{i_d} = \frac{1}{f_p} - \frac{1}{d}$. Now, following a similar argument to depth-from-defocus[112], consider the case where the depth-varying pattern is selected to be a single light ray that intersects the perimeter of this effective aperture at a_d and passes through p_d. This ray can be effectively replaced with a single collimated laser source at a_d (Figure 5.13(c)). From this point-of-view, the effective depth resolution of our system should be roughly equivalent to a stereo system with baseline $i_d \tan\theta_p$. For example, for a point that is $0.5\,\mathrm{m}$ away, the effective baseline should be $\approx 9\,\mathrm{mm}$.

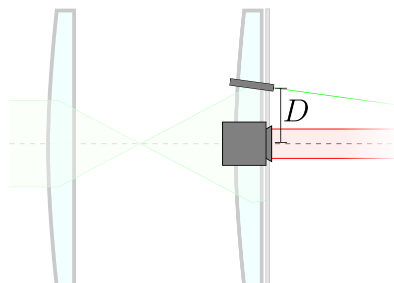
While our holographic system may have similar resolution to a small-baseline triangulation system, it has certain theoretical strengths when compared to other active sensing approaches. For one, analogous to depth-from-defocus, this cue is likely more robust to occlusions than a similar baseline stereo system [112]. Furthermore, structured light systems struggle with a tradeoff between depth-of-field



(a) Holographic illumination



(b) One ray



(c) Equivalent stereo baseline

Figure 5.13: **Holographic depth variation resolution visualization.** We can coarsely estimate the effective resolution of our system using geometric optics approximations, following a depth-from-defocus argument [112]. **(a)** Consider the geometry of a holographic illumination system. For some point d away from the system, we can calculate its effective aperture as $D = i_d \tan \theta_p$ with $\frac{1}{i_d} = \frac{1}{f_p} - \frac{1}{d}$ and $\sin \theta_p = \frac{\delta N}{2 f_2}$, where f_2 is the focal length of Lens 2, f_p the focal length of the projector lens, δ is the SLM pixel pitch and N is the number of SLM pixels along the largest dimension. **(b)** Now, consider the case where no light leaves the system except for the light ray that corresponds to this maximum diffraction angle θ_p. **(c)** This system is roughly equivalent to a stereo system with baseline D.

and brightness — a large aperture projector lens is desired so that output patterns are bright and legible under ambient light, but this simultaneously reduces projector depth-of-field and therefore resolution for scenes with a variety of depths. In contrast, our cue's resolution fundamentally increases with larger aperture, avoiding this tradeoff. Time-of-flight does not have depth-of-field challenges, but it lacks the depth resolution in settings like microscopy where the height profiles of tiny objects are required [60]. Our cue has no such limitation, as it will have high resolution for such close-by objects.

More generally, this holographic cue is *complementary* to other depth cues, and can directly be used alongside them. Given that holographic illumination is being applied to structured light (Chapter 4) and CWTOF (Chapter 3), our holographic depth cue could be potentially implemented in future holographic systems for these tasks as-is, and be combined with these cues *in the same capture* to create improved fused measurements. For instance, CWTOF suffers a tradeoff between ambiguous range and depth resolution. If the CWTOF sensor also emits a depth-dependent pattern, the holographic cues could be used to estimate a coarse depth without any ambiguous range, which CWTOF super-resolves using sinusoids with a larger ambiguous range but finer depth resolution. Thus, the ambiguity of CWTOF is decreased without impacting resolution or framerate. In the context of structured light, projecting depth-dependent structured light patterns could ensure meaningful depth can still be recovered under defocus.

Limits on étendue-expanded depth programmability

In general, some depth-varying patterns are easier to form than others on our holographic system. As shown in Fig. 5.14, arbitrary content at different planes can be more difficult than specially structured patterns, like a rotating plus. Additionally, the closer together these planes are, the more difficult it is to clearly disambiguate the content at each one, as visualized in Fig. 5.15. To explore these limits, in Fig. 5.16, we project unique content at two planes as we vary the distance between them. As we move these two planes closer together, it is harder to find an SLM pattern that will project the correct content at both planes, and instead an intermixing of the content arises. We note, however, that our simulated setup has a somewhat higher depth resolution than our real setup. This suggests that closing the simulation-to-real gap, perhaps using a neural-augmented model [27, 32, 33, 64, 104, 116], could potentially improve the real depth resolution.

5.1.5 Summary

In our work, we propose Holodepth: a holographic approach towards creating depth-varying patterns. To enhance this functionality, we place an étendue-expanding lens array into the system, introduce a novel calibration method for it, and explore its effects on the resultant patterns. With our prototype, we demonstrate novel applications that leverage depth-varying patterns. We show that it can be potentially useful for future wearables and entertainment as well as computer vision tasks like depth sensing and light curtains. We finished by analyzing the effective resolution of our system for these tasks, to better demonstrate the scenarios in which it may be useful.

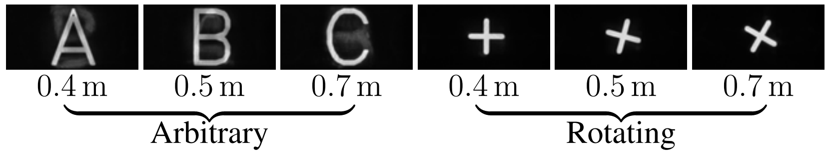


Figure 5.14: **Simulation of different depth-varying projections.** Some patterns can be easier to resolve than other ones, like a rotating pattern [51].

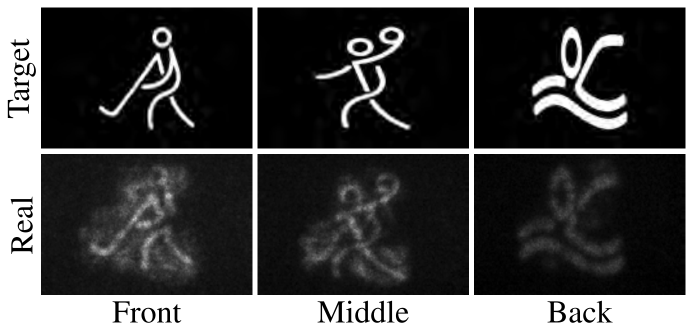


Figure 5.15: **Limits on depth variation**. We tried forming unique content at three planes separated $2.5\,\mathrm{mm}$ apart before the projector lens. There is much more blur than a similar scene in Figure 5.9(e).

5.2 Limitations of depth-varying holographic illumination

In practice, a number of drawbacks arise with a holographic approach for depth-varying illumination. For one, in order to accurately model coherent light transport in a system with sufficient étendue for depth-varying content, we need to simulate propagation at high resolution — for instance, we used 9600×6000 pixels in our Holodepth system. Operating at such high resolution has heavy memory and compute requirements — our NVIDIA RTX 3090 can recover just 5 depth targets at a time with Equation (5.1). One could also apply a non-iterative approach to reduce computation [84, 128], at the cost of pattern quality. We discuss further potential solutions in Chapter 6.

Furthermore, our proposed approach for depth sensing and light curtains performs best with a camera with little defocus. However, reducing the aperture of the camera lens to increase depth-of-field can increase the visibility of a phenomena called speckle. Properly accounting for this effect via pupil-aware modeling [28, 113] could potentially close this sim-to-real gap. Another potential cause is insufficient degrees of freedom for forming a desired depth-varying target. Possible solutions include temporally-multiplexing SLM patterns at the cost of frame

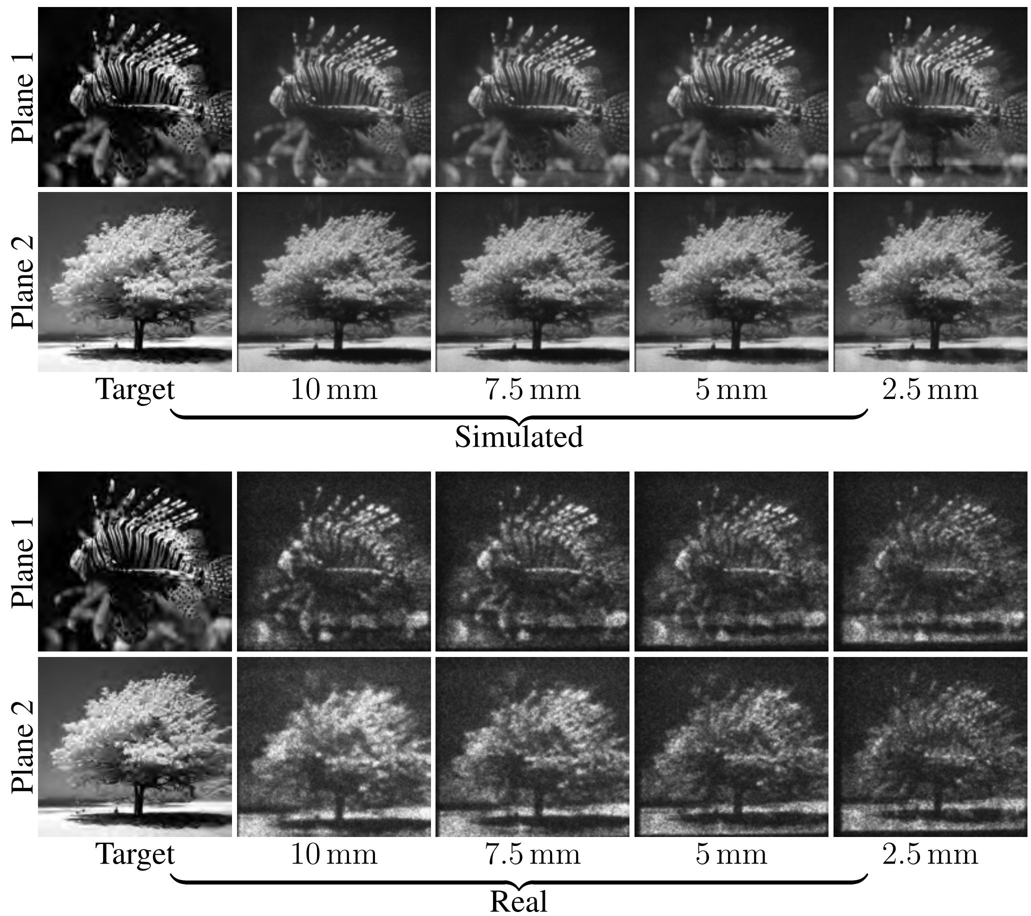


Figure 5.16: **Depth variation resolution visualization**. We visualize the depth variation of two planes receiving different content in simulation and on our real setup. As the distance between the two planes decreases, the more the content intermixes and the harder it is to visually separate. The distances shown are before the projection lens.

rate [33] or multiple SLMs and sources at the cost of extra system complexity [75]. We discuss speckle in more detail in Chapter 6.

Finally, as discussed in Section 5.1.4, the depth variation provided by current holographic systems is still limited. On our Holodepth prototype, two target planes need to be separated by $1/3$ of a diopter in order to meaningfully project unique content, which could still be too large of a range for depth-varying illumination applications. This suggests that further étendue expansion or a natively higher-resolution SLM may be required in practice.

Chapter 6

Discussion

In summary, we have proposed the use of holographic illumination in computer vision, where we control the coherent propagation of laser light via a programmable modulator. Such an approach tackles key challenges in active sensing. For one, we show that they can be used to implement light-efficient lighting, where energy can be concentrated as needed. We demonstrate the potential strength of such an approach in the context of time-of-flight sensors, where our modified system can sense farther and darker objects than before. This redistribution also allows for quantized fast modulators to produce arbitrarily programmable bright output patterns, enabling far faster and brighter projector systems. We apply such a system to new modes of triangulation light curtains. Finally, we show that a holographic approach can be used to implement depth-programmable illumination, where output patterns can be controlled as a function of distance. We apply such a system to mitigate projector defocus and unlock new kinds of 3D interfaces and projection mapping.

To conclude this thesis, we first start by explaining the limitations of holographic illumination, and possible ways to address them. We then finish by discussing possible future directions that build off our findings.

6.1 Limitations of holographic illumination

Despite their strengths as we have discussed, there are certain limitations associated with using a holographic illumination source. We discuss them in more detail below, and potential solutions for them.

6.1.1 Speckle

Perhaps the most significant limitation of holographic illumination is *speckle* — unavoidable noise-like artifacts that appear thanks to the coherence of laser light. These effects can dramatically decrease the visual quality of output patterns, and can potentially reduce the resolution of active sensing techniques like structured light.

To understand where these speckle effects come from, we must first consider how speckle is fundamentally formed. Consider the output pattern formed at a plane some arbitrary distance from the SLM — for now, let us assume a random phase pattern uniformly distributed between 0 and 2π is displayed on the SLM. At a given point x in the output pattern, it receives effectively a summation of random phasors from each pixel on the SLM. Thus, from the central limit theorem, it can be shown that the summed output phasor at x will have real and imaginary components distributed as zero mean, identically-distributed Gaussian random variables, and thus the intensity as this point will be distributed according to the negative exponential distribution. Thanks to the propagation of laser light from Chapter 2, changing this point x will produce a different set of random phasors, and thus produce a different sample from the negative exponential distribution [47]. This variation between neighboring points creates the noise-like effect of speckle. More generally, speckle forms at a high level due to the random interference of different paths taken by light to reach the output point.

With these ideas in mind, it may seem unintuitive that speckle appears in the output of holographic illumination, as we fundamentally use designed patterns via Equation (2.2) instead of the aforementioned random phase patterns. Why, then, does speckle appear in the outputs of holographic illumination?

Modeling

One important reason is inaccurate modeling. In practice, various non-idealities can occur in real systems. For instance, lenses are not ideal in the real world, breaking the assumed Fourier model in Equation (2.7). Not all light in the system follows the expected light path — some light is not properly modulated by the SLM thanks to the SLM's finite fill factor (producing a "DC" spot), while other light is diffracted according to a higher diffraction order [49], or perhaps the output plane is misaligned by a small amount. These types of effects introduce unexpected "random" behavior in the output patterns, that can result in speckle.

These problems can potentially be remedied via calibrating more sophisticated forward models. For one, one could use a more complete forward model that explicitly captures various potential physical effects, *e.g.*, models that account for non-idealities in the optics or stray light as we utilize in Chapter 5. However, such approaches require foreknowledge of various phenomena that could influence the output images such that they can be mathematically modeled, which is typically an infeasible challenge. Thus, recent work has proposed the use of neural-augmented modeling, where a neural network is used to close the sim-to-real gap in the forward model [27, 31, 32, 33, 104]. Typically, these networks are used to manipulate the input and output of a physics-informed propagation, for example in the far-field case:

$$\mathcal{P}_{\text{neural lens}}(\phi) = \mathcal{N}_{\text{output}}\left(\mathcal{F}(\mathcal{N}_{\text{input}}(S(\phi)))\right), \tag{6.1}$$

where $\mathcal{N}$ denotes a neural network. While effective, these networks typically introduce significant computational burden, as solving Equation (2.2) with such a neural-augmented model requires numerous backpropagation steps through a large neural network. Additionally, as they are not constrained to a physically-plausible model, they can behave uninterpretably, and require a significant number of input/output pairs to properly train. For example, we found that it can be hard to create bright, light-concentrated patterns as output from such a neural model, which would be necessary to apply the ideas in this thesis for light-efficient and high-speed holographic illumination. Addressing these limits could enable more robust, higher resolution holographic illumination systems.

Degrees of freedom

A second cause of speckle is insufficient degrees of freedom. For instance, in Chapter 5, we used a holographic system to form desired images at multiple depth planes — given that we only have programmability over the wavefront at a single plane via the SLM, there is a mismatch in the degrees of freedom of the input and desired outputs. Another case where insufficient degrees of freedom appear is the quantization of the SLM — in practice, current fast SLMs come with a tradeoff of lower bit depth, *e.g.*, the kilohertz SLM we use in Chapter 4 is limited to binary amplitude. Thus, solving Equation (5.1) necessarily results in some kind of tradeoff in the quality of the output signal, that typically manifests as visible speckle.

One solution that is often leveraged in the 3D display literature is to algorithmically encourage speckle-free holograms by enforcing smooth phase, *e.g.*, neighboring output points are constrained to have similar phase values [32]. However, such an approach produces unwanted side effects. First, recent display research has found that such approaches produce unnatural sharp defocus patterns [75, 113], because most of the energy is concentrated into a very small region of the eyebox. In the context of holographic illumination, this translates to only a small amount of light redistribution and reduced effective étendue, making such approaches impractical.

To tackle this problem, a number of approaches have been proposed in the 3D display literature that introduce some degree of incoherence to effectively "blur" out the speckle. Perhaps the simplest instantiation of this idea is the use of a partially coherent source like an LED [37, 105], but such a method also simultaneously blurs the output pattern, reducing resolution. In recent work, the most popular is time multiplexing, where multiple SLM patterns are shown for each individual target image [23, 33, 77, 80]. Such an approach could be applied for our light-efficient and depth-programmable holographic illumination configurations. However, it would be impractical for our high-speed illumination system, as we would need far faster SLMs than are currently available to be able to show multiple SLM frames for each high-speed frame. Other potential approaches involve using multiple lasers or wavelengths with multiple SLMs [75, 114], but such approaches increase cost, complexity and form factor. Building simple-but-compact optical setups for effective despeckling would help enable both higher resolution active sensors and displays.

Subjective speckle

The discussion above has primarily focused on the case of "objective" speckle — speckle that forms thanks to the propagation of light before it interacts with any scene objects. However, "subjective" speckle that forms thanks to the object-to-camera propagation can also reduce measurement quality. In general, using a larger aperture on the lens can minimize the effects of subjective speckle, as the size of the average speckle typically scales inversely with the aperture. However, this is not necessarily an option in computer vision — we often assume all-in-focus images, which can necessitate a small camera aperture for a large depth-of-field (this caused problems for our work in Chapter 5). If most of the subjective speckle is caused not by surface microgeometry but the imaging system itself, these effects could potentially be remedied via some form of pupil-aware modeling [28, 113], that accounts for the aperture of the camera.

Importance of speckle in computer vision

In general, speckle may be a bigger challenge for holographic displays and projectors than in our context of holographic illumination, as the images we capture can actually be post-processed to handle the speckle. For instance, a bilateral filter could potentially be applied to try and remedy any visible speckle. Furthermore, if subjective speckle is not significant, well-calibrated objective speckle can potentially be used as an additional cue for vision tasks, *e.g.*, it could be used as a form of feature for correspondences [62]. However, any speckle visible from the camera can potentially induce decreased signal-to-noise ratio for any lows of the speckle pattern, resulting in reduced spatial resolution.

6.1.2 Computational cost

As discussed in Section 2.3, we typically rely on iterative algorithms to solve for the right SLM pattern to display. While such approaches typically result in good image quality, this process can unfortunately be expensive — each iteration typically requires Fourier transforms calculated at high resolution. This computational expense restricts many of the applications we have discussed for holographic illumination to precomputed patterns, limiting their practicality.

To reduce this computational cost, there are a number of potential solutions. One line of research has proposed the use of an extra optical aperture to ensure that Equation (2.2) can be solved extremely accurately with a single propagation step [84]. These approaches assign a phase value ϕ_T for every point in T, and then apply a transpose propagation operator to estimate a wavefront at the SLM, which is then converted to amplitude/phase via some operator $\Gamma(\cdot)$ [20, 84]:

$$\phi^* \approx \Gamma\left(\mathcal{P}^H(Te^{j\phi_T})\right). \tag{6.2}$$

For instance, in double-phase amplitude coding [84], $\Gamma(\cdot)$ creates a high-frequency grating that implicitly encodes amplitude on a phase SLM, but requires a high-pass filter to block unwanted light. While effective, these approaches unfortunately come with limitations that make them undesirable for the context of holographic illumination. First, the use of an aperture filter reduces field-of-view in the far-field configurations we leverage for holographic illumination, and can cause challenges with étendue and depth variation (Chapter 5). Second, such techniques are currently incompatible with quantized SLMs, precluding them from high-speed holographic illumination (Chapter 4).

Another possible solution leveraged in the display literature is to train a neural network to produce a good SLM pattern in a single step [104, 116]. These ideas could be extended to the context of holographic illumination — however, given the nature of light redistribution, special care needs to be taken to ensure that the network architecture, currently typically limited in receptive field due to the use of convolutional layers [116], is capable of mapping a single output point to the wavefront over the entire SLM. Inspired by the GAN inversion literature [81], another potential approach is to use a neural network to learn some space in which fewer iterations are needed to converge when solving Equation (2.2).

One alternative direction we propose is to manipulate previously-computed SLM patterns. For instance, in the case of a far-field configuration (Equation (2.7)), if we simply need a translated version of a previous pattern, it suffices to simply multiply the previous SLM pattern by an appropriate phase ramp. If we need to change the depth at which the output pattern is focused at, we can multiply the SLM pattern by a quadratic function in phase. If we need to rotate the output pattern, we can simply rotate the SLM pattern. More broadly, for two very similar output patterns, a powerful research direction would be how to produce a good SLM pattern for one output pattern given a SLM pattern for the other. Given that sequential output patterns are likely similar, such an approach could potentially dramatically reduce computation.

6.1.3 Form factor

In practice, our holographic illumination prototypes possess significant form factor, compared to the typical illumination systems used today in active sensing. In theory, this is primarily due to the requisite lenses in the system — we use a lens to collimate the laser light incident on the SLM, a second lens to perform a Fourier transform, and a final objective lens to magnify the output image into the scene. However, in theory, none of these lens are required — in fact, a simple diverging source without any lenses could be used to magnify the holographic content onto the scene [84]. But, the effective eyebox is reduced for each output point [84], potentially reducing the amount of light redistribution for Chapters 3 and 4 and level of defocus for Chapter 5. Addressing this limitation via ideas from étendue expansion could be a powerful push towards the practicality of real-world holographic illumination.

6.1.4 Is étendue a challenge?

In general, étendue is a significant challenge for holographic displays — as discussed in Chapter 5, the limited pixel count of modern SLMs presents a tradeoff between field-of-view and eyebox size, which could preclude integration of a holographic display into a real head-mounted display. In the context of holographic illumination, limited étendue presents more of a mixed bag. For one, as discussed in Chapter 5, increased étendue could allow for more meaningful depth-programmable projection. However, small étendue can also be a strength as it results in less visible defocus, making current holographic illumination systems well-suited for scenes with significant depth variation where traditional projectors struggle with defocus. However, increasing output pattern resolution also inherently requires a higher-resolution modulator, which would increase étendue and therefore also produce more defocus — however, this defocus could be compensated following the methodology described in Chapter 5. Thus, practically speaking, increased étendue will likely also be helpful for holographic illumination.

6.2 Broader future directions

Beyond addressing the above limitations, there are many avenues of future work. We discuss two potential directions in more detail below.

6.2.1 Adaptive light concentration

First, our proposed holographic illumination unlocks a new modality of general purpose light-redistributive projectors. As we demonstrated in Chapter 3, such a system fits very naturally into the context of adaptive sensors, that adjust their measurements according to scene priors and/or previous captures. Our light-efficient modality provides many opportunities for such a paradigm. For one, light could be concentrated according to many other priors. For example, energy could be concentrated in regions of strong multi-path interference or bright ambient light. Energy could also be concentrated in tandem with other depth modalities. For instance, followup work to our original paper showed that such systems can be used to concentrate light in featureless regions for energy-efficient active stereo [128]. Light could be concentrated in regions of uncertainty from a generative monocular depth model [67].

Light could also be concentrated under other modalities. For instance, our work described in Chapter 3 focused primarily on intensity concentration, where light is redistributed across different spatial regions. Alternatively, our system could also be used to concentrate light with spatially-varying patterns. More specifically, dot patterns are a traditional approach to extend the range of active sensors by locally concentrating light from nearby pixels into points. Such an approach therefore trades off spatial resolution with depth accuracy. Our holographic illumination could be used to adaptively program the concentration pattern in a spatially-varying fashion. For example, regions of a scene where higher spatial resolution is needed could be illuminated with a uniform pattern, where other darker regions that reflect little light could be illuminated with a sparse pattern.

Light could also be concentrated along the temporal dimension, for some given measurement time. In general, concentrating light temporally can often be beneficial. For one, temporally concentrated light is often used in the form of flash photography to minimize motion blur. Temporally concentrating light can also effectively reduce the contribution of ambient light if the exposure time is set to match the duration of the flash [99]. However, temporally spreading light can also be potentially helpful. For one, in the context of holographic illumination, more fast SLM frames could be used for despeckling via time multiplexing. Second, more distinct patterns could be used to illuminate the scene — while each pattern would use less light, the use of more patterns could be used to create, *e.g.*, denser measurements in the context of structured light. Third, using temporally spread illumination typically improves eye safety. Thus, spatially-varying the temporal concentration depending on sensing needs could be a very powerful direction, *e.g.*, light could be temporally concentrated on moving objects, while spread for time multiplexing on static objects.

These ideas underly the fundamental question of light concentration — given a measurement time, an energy budget, a sensor and some prior information, what is the optimal lighting pattern that produces the desired information about a target scene? A method for directly estimating both the optimal lighting pattern and reconstruction algorithm could be potentially tremendously useful for the automatic design of active sensing systems.

6.2.2 Multi-element optical systems

A recent trend in the display literature has shown that extra optical elements and modulators can potentially produce desirable capabilities in our optical systems. For instance, as we showed in Chapter 5 among other work in holographic displays [25, 74, 91], an extra high resolution element can potentially expand the étendue of an optical system. Kuo et al. and Schiffers et al. showed that additional SLMs and laser sources/wavelengths can potentially help mitigate speckle [75, 114]. Qin et al. showed that an extra SLM and two extra cubic phase plates can be used to produce a solid-state multifocal display [108]. This poses a powerful research direction — if we can arbitrarily compose extra elements, light sources, and modulators, what kinds of cameras and displays can we build when combined with computation? Answering questions on theoretical limits, how to sample appropriate designs for specific tasks, and more could result in a new space of more powerful cameras and displays, and create an underlying theory for their design.

Bibliography

[1] Supreeth Achar, Joseph R. Bartels, William L. 'Red' Whittaker, Kiriakos N. Kutulakos, and Srinivasa G. Narasimhan. Epipolar time-of-flight imaging. *ACM Transactions on Graphics (TOG)*, 36(4), July 2017. xiii, 17, 28, 30, 38

[2] Jens Ackermann, Michael Goesele, et al. A survey of photometric stereo techniques. *Foundations and Trends® in Computer Graphics and Vision*, 9 (3-4):149–254, 2015. 1

[3] Advanced Illumination. Photometric stereo. `https://advancedillumination.com/photometric-stereo/`, 2025. 2

[4] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, July 2017. xiv, 31, 36

[5] Alexandrea P. The complete guide to 3d scanners using laser triangulation. `https://www.3dnatives.com/en/3d-scanner-laser-triangulation080920174-99/`, 2017. 38

[6] Toshiyuki Amano and Hirokazu Kato. Real world dynamic appearance enhancement with procam feedback. In *the International Workshop on Projector-Camera Systems (PROCAMS)*, pages 1–2. ACM/IEEE, 2008. 17

[7] Toshiyuki Amano and Ryo Suzuki. Virtual recovery of the deteriorated art object based on ar technology. In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–2. IEEE, 2007. 17

[8] Siddharth Ancha, Yaadhav Raaj, Peiyun Hu, Srinivasa G. Narasimhan, and David Held. Active perception using light curtains for autonomous driving. In *the European Conference on Computer Vision (ECCV)*, pages 751–766. Springer, 2020. 42

[9] Siddharth Ancha, Gaurav Pathak, Srinivasa G. Narasimhan, and David Held. Active safety envelopes using light curtains with probabilistic guarantees. In *Robotics: Science and Systems*, 2021. 42

[10] Apple Inc. ARKit 6 - Augmented Reality. `https://developer.apple.com/augmented-reality/arkit/`, 2025. 1

[11] Peter C. Barnum, Srinivasa G. Narasimhan, and Takeo Kanade. A multi-layered display with water drops. In *ACM SIGGRAPH*, pages 1–7. Association for Computing Machinery, New York, NY, USA, 2010. 65

[12] Joseph R. Bartels, Jian Wang, William Whittaker, Srinivasa G. Narasimhan, et al. Agile depth sensing using triangulation light curtains. In *the International Conference on Computer Vision (ICCV)*, pages 7899–7907, 2019. xv, 38, 41, 42, 43, 44, 45, 67

[13] Terry A. Bartlett, William C. McDonald, and James N. Hall. Adapting texas instruments dlp technology to demonstrate a phase spatial light modulator. In *Emerging Digital Micromirror Device Based Systems and Applications XI*, volume 10932, page 109320S. International Society for Optics and Photonics, 2019. 25, 50

[14] Oliver Bimber and Andreas Emmerling. Multifocal projection: A multiprojector technique for increasing focal depth. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 12(4):658–667, 2006. 52

[15] Oliver Bimber and Daisuke Iwai. Superimposing dynamic range. *ACM Transactions on Graphics (TOG)*, 27(5):1–8, 2008. 17

[16] Oliver Bimber, Daniel Klöck, Toshiyuki Amano, Anselm Grundhöfer, and Daniel Kurz. Closed-loop feedback illumination for optical inverse tone-mapping in light microscopy. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 17(6):857–870, 2010. 17

[17] Bryon R. Brown and Adolf W. Lohmann. Complex spatial filtering with binary masks. *Applied Optics*, 5(6):967–969, 1966. 6

[18] Bryon R. Brown and Adolf W. Lohmann. Computer-generated binary holograms. *IBM Journal of Research and Development*, 13(2):160–168, 1969. 6

[19] Michael S. Brown, Peng Song, and Tat-Jen Cham. Image pre-conditioning for out-of-focus projector blur. In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 1956–1963. IEEE, 2006. 52

[20] Olof Bryngdahl and Adolf W. Lohmann. Single-sideband holography. *Journal of the Optical Society of America*, 58(5):620–624, 1968. 79

[21] Edward Buckley. Holographic laser projection. *Journal of Display Technology*, 7(3):135–140, 2011. 7, 9, 10, 52

[22] Edward Buckley, Dominik Stindt, and Robert Isele. 14.4: Novel human-machine interface (hmi) design enabled by holographic laser projection. In *SID Symposium Digest of Technical Papers*, volume 40, pages 172–177. Wiley Online Library, 2009. 7, 9, 10

[23] AJ Cable, E Buckley, P Mash, NA Lawrence, TD Wilkinson, and WA Crossland. Real-time binary hologram generation for high-quality video projection applications. In *SID Symposium Digest of Technical Papers*, volume 35, pages 1431–1433. Wiley Online Library, 2004. 78

[24] Yunus A. Cengel and Michael A. Boles. Thermodynamics: an engineering approach. *Sea*, 1000(8862):287–93, 2002. 2, 15

[25] Minseok Chae, Kiseung Bang, Dongheon Yoo, and Yoonchan Jeong. Étendue expansion in holographic near eye displays through sparse eye-box generation using lens array eyepiece. *ACM Transactions on Graphics (TOG)*, 42(4), July 2023. xvii, 52, 53, 56, 57, 58, 82

[26] Praneeth Chakravarthula, Yifan Peng, Joel Kollin, Henry Fuchs, and Felix Heide. Wirtinger holography for near-eye displays. *ACM Transactions on Graphics (TOG)*, 38(6), November 2019. 13, 54

[27] Praneeth Chakravarthula, Ethan Tseng, Tarun Srivastava, Henry Fuchs, and Felix Heide. Learned hardware-in-the-loop phase retrieval for holographic near-eye displays. *ACM Transactions on Graphics (TOG)*, 39(6), November 2020. 11, 54, 70, 77

[28] Praneeth Chakravarthula, Seung-Hwan Baek, Florian Schiffers, Ethan Tseng, Grace Kuo, Andrew Maimone, Nathan Matsuda, Oliver Cossairt, Douglas Lanman, and Felix Heide. Pupil-aware holography. *ACM Transactions on Graphics (TOG)*, 41(6), November 2022. 71, 78

[29] Praneeth Chakravarthula, Ethan Tseng, Henry Fuchs, and Felix Heide. Hogel-free holography. *ACM Transactions on Graphics (TOG)*, 41(5), October 2022. 54

[30] Chenliang Chang, Jun Xia, and Yanqiao Jiang. Holographic image projection on tilted planes by phase-only computer generated hologram using fractional fourier transformation. *Journal of Display Technology*, 10(2):107–113, 2013. 7, 9

[31] Brian Chao, Manu Gopakumar, Suyeon Choi, and Gordon Wetzstein. High-brightness holographic projection. *Optics Letters*, 48(15):4041–4044, 2023. 10, 11, 52, 77

[32] Suyeon Choi, Manu Gopakumar, Yifan Peng, Jonghyun Kim, and Gordon Wetzstein. Neural 3d holography: learning accurate wave propagation mod-

els for 3d holographic virtual and augmented reality displays. *ACM Transactions on Graphics (TOG)*, 40(6), December 2021. 7, 8, 11, 54, 70, 77

[33] Suyeon Choi, Manu Gopakumar, Yifan Peng, Jonghyun Kim, Matthew O'Toole, and Gordon Wetzstein. Time-multiplexed neural holography: a flexible framework for holographic near-eye displays with fast heavily-quantized spatial light modulators. In *ACM SIGGRAPH*, New York, NY, USA, 2022. Association for Computing Machinery. 11, 39, 54, 60, 70, 73, 77, 78

[34] Pina Colarusso and Craig Brideau. Speed running your sample: Faster laser scanning in confocal and multiphoton microscopes. `https://bliqphotonics.com/faster-laser-scanning-confocal-multiphoton-microscopes/`, 2023. 18

[35] Miguel Heredia Conde, Klaus Hartmann, and Otmar Loffeld. Adaptive high dynamic range for time-of-flight cameras. *IEEE Transactions on Instrumentation and Measurement (TIM)*, 64(7):1885–1906, 2014. 23

[36] Gerwin Damberg, James Gregson, and Wolfgang Heidrich. High brightness hdr projection using dynamic freeform lensing. *ACM Transactions on Graphics (TOG)*, 35(3), May 2016. 11, 18

[37] Yuanbo Deng and Daping Chu. Coherence properties of different light sources and their effect on the image sharpness and speckle of holographic displays. *Scientific Reports*, 7(1):5893, 2017. 78

[38] Dragos Falie and Vasile Buzuloiu. Noise characteristics of 3d time-of-flight cameras. In *the International Symposium on Signals, Circuits and Systems (ISSCS)*, volume 1, pages 1–4. IEEE, 2007. 22

[39] Fandom contributors. TrueDepth. `https://apple.fandom.com/wiki/TrueDepth`, 2025. 1

[40] David Ferstl, Christian Reinbacher, Gernot Riegler, Matthias Rüther, and Horst Bischof. Learning depth calibration of time-of-flight cameras. In *British Machine Vision Conference (BMVC)*, pages 102.1–102.12. BMVA Press, September 2015. 23

[41] James R Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21(15):2758–2769, 1982. 13, 26

[42] Kate Flint. *Flash!: Photography, Writing, and Surprising Illumination*. Oxford University Press, 2017. 1

[43] Dennis Gabor. A new microscopic principle. *Nature*, 161:777–778, 1948. 6

[44] Chunyu Gao, Narendra Ahuja, and Hong Hua. Active aperture control and sensor modulation for flexible imaging. In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2007. 17

[45] Jason Geng. Structured-light 3d surface imaging: a tutorial. *Advances in Optics and Photonics*, 3(2):128–160, 2011. 1, 37

[46] R. W. Gerchberg and W. O. Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237–246, 1972. 13, 39

[47] Joseph W. Goodman. Some fundamental properties of speckle. *Journal of the Optical Society of America*, 66(11):1145–1150, Nov 1976. 6, 76

[48] Joseph W. Goodman. *Introduction to Fourier Optics*. McGraw-Hill physical and quantum electronics series. W. H. Freeman, 2005. ISBN 9780974707723. 5, 7, 9

[49] Manu Gopakumar, Jonghyun Kim, Suyeon Choi, Yifan Peng, and Gordon Wetzstein. Unfiltered holography: optimizing high diffraction orders without optical filtering for compact holographic displays. *Optics Letters*, 46(23): 5822–5825, 2021. 76

[50] Manu Gopakumar, Gun-Yeal Lee, Suyeon Choi, Brian Chao, Yifan Peng, Jonghyun Kim, and Gordon Wetzstein. Full-colour 3d holographic augmented-reality displays with metasurface waveguides. *Nature*, 629 (8013):791–797, 2024. 7, 8

[51] Adam Greengard, Yoav Y Schechner, and Rafael Piestun. Depth from diffracted rotation. *Optics letters*, 31(2):181–183, 2006. xix, 71

[52] Michael D Grossberg, Harish Peri, Shree K Nayar, and Peter N Belhumeur. Making one object look like another: Controlling appearance using a projector-camera system. In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages I–I. IEEE, 2004. 17

[53] Max Grosse, Gordon Wetzstein, Anselm Grundhöfer, and Oliver Bimber. Coded aperture projection. *ACM Transactions on Graphics (TOG)*, 29(3), July 2010. xvii, xviii, 3, 51, 52, 54, 61

[54] Mohit Gupta, Amit Agrawal, Ashok Veeraraghavan, and Srinivasa G. Narasimhan. Structured light 3d scanning in the presence of global illumination. In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 713–720. IEEE, 2011. 18

[55] Mohit Gupta, Qi Yin, and Shree K Nayar. Structured light in sunlight. In *the International Conference on Computer Vision (ICCV)*, pages 545–552, 2013. 2, 3, 15, 16, 17

[56] Samuel W Hasinoff, Frédo Durand, and William T Freeman. Noise-optimal capture for high dynamic range photography. In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 553–560. IEEE, 2010. 2

[57] Matthew Hirsch, Gordon Wetzstein, and Ramesh Raskar. A compressive light field projection system. *ACM Transactions on Graphics (TOG)*, 33(4), July 2014. 52

[58] Sven TS Holmström, Utku Baran, and Hakan Urey. Mems laser scanners: a review. *Journal of Microelectromechanical Systems*, 23(2):259–275, 2014. 37, 38

[59] Reynald Hoskinson, Boris Stoeber, Wolfgang Heidrich, and Sidney Fels. Light reallocation for high contrast projection using an analog micromirror array. *ACM Transactions on Graphics (TOG)*, 29(6), December 2010. 11, 18

[60] Yan Hu, Qian Chen, Shijie Feng, and Chao Zuo. Microscopic fringe projection profilometry: A review. *Optics and Lasers in Engineering*, 135:106192, 2020. 69

[61] Humane. Humane previews first product in ted talk. `https://hu.ma.ne/media/humane-previews-first-product-in-ted-talk`, 2023. 64

[62] Pranay Jain and Sanjay E. Sarma. Measuring light transport properties using speckle patterns as structured illumination. *Scientific Reports*, 9(1), 2019. 78

[63] Leyla Kabuli, Oliver Cossairt, Florian Schiffers, Nathan Matsuda, and Grace Kuo. Practical high-contrast holography. *arXiv preprint arXiv:2410.19347*, 2024. 10

[64] Koray Kavaklı, Hakan Urey, and Kaan Akşit. Learned holographic light transport. *Applied Optics*, 61(5):B50–B55, 2022. 54, 60, 70

[65] Hiroshi Kawasaki, Yuki Horita, Hitoshi Masuyama, Satoshi Ono, Makoto Kimura, and Yasuo Takane. Optimized aperture for estimating depth from projector's defocus. In *the International Conference on 3D Vision (3DV)*, pages 135–142. IEEE, 2013. 52

[66] Hiroshi Kawasaki, Satoshi Ono, Yuki Horita, Yuki Shiba, Ryo Furukawa, and Shinsaku Hiura. Active one-shot scan for wide depth range using a light field projector based on coded aperture. In *the International Conference on Computer Vision (ICCV)*, pages 3568–3576, 2015. 52, 65

[67] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Ro-

drigo Caye Daudt, and Konrad Schindler. Repurposing diffusion-based image generators for monocular depth estimation. In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9492–9502, 2024. 81

[68] Keyence Corporation of America. Safety light curtains. `https://www.keyence.com/products/safety/light-curtain/`, 2021. 41

[69] Dongyeon Kim, Seung-Woo Nam, Suyeon Choi, Jong-Mo Seo, Gordon Wetzstein, and Yoonchan Jeong. Holographic parallax improves 3d perceptual realism. *ACM Transactions on Graphics (TOG)*, 43(4), July 2024. 7, 8

[70] Jonghyun Kim, Manu Gopakumar, Suyeon Choi, Yifan Peng, Ward Lopes, and Gordon Wetzstein. Holographic glasses for virtual reality. In *ACM SIGGRAPH*, New York, NY, USA, 2022. Association for Computing Machinery. 7, 9

[71] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *the International Conference on Learning Representations (ICLR)*, 2015. 13, 60

[72] Andreas Kolb, Erhardt Barth, Reinhard Koch, and Rasmus Larsen. Time-of-flight cameras in computer graphics. In *Computer Graphics Forum*, volume 29, pages 141–159. Wiley Online Library, 2010. 1

[73] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009. xiv, 31, 36

[74] Grace Kuo, Laura Waller, Ren Ng, and Andrew Maimone. High resolution étendue expansion for holographic displays. *ACM Transactions on Graphics (TOG)*, 39(4), August 2020. xvii, xviii, 53, 54, 55, 56, 57, 58, 61, 82

[75] Grace Kuo, Florian Schiffers, Douglas Lanman, Oliver Cossairt, and Nathan Matsuda. Multisource holography. *ACM Transactions on Graphics (TOG)*, 42(6), December 2023. 54, 73, 77, 78, 82

[76] Byounghyo Lee, Dongheon Yoo, Jinsoo Jeong, Seungjae Lee, Dukho Lee, and Byoungho Lee. Wide-angle speckleless dmd holographic display using structured illumination with temporal multiplexing. *Optics Letters*, 45(8): 2148–2151, 2020. 53, 56

[77] Byounghyo Lee, Dongyeon Kim, Seungjae Lee, Chun Chen, and Byoungho Lee. High-contrast, speckle-free, true 3d holography via binary cgh optimization. *Scientific Reports*, 12(1):2811, 2022. 39, 53, 55, 78

[78] Marvin Lindner and Andreas Kolb. Calibration of the intensity-related distance error of the pmd tof-camera. In *Intelligent Robots and Computer Vision XXV: Algorithms, Techniques, and Active Vision*, volume 6764, pages 338–

345. SPIE, 2007. 23

[79] Fu Liu, Alexandros Pitilakis, Mohammad Sajjad Mirmoosa, Odysseas Tsilipakos, Xuchen Wang, Anna C. Tasolamprou, Sergi Abadal, Albert Cabellos-Aparicio, Eduard Alarcón, Christos Liaskos, Nikolaos V. Kantartzis, Maria Kafesaki, Eleftherios N. Economou, Costas M. Soukoulis, and Sergei Tretyakov. Programmable metasurfaces: State of the art and prospects. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5, 2018. 7

[80] Jung-Ping Liu, Ming-Hsuan Wu, and Peter WM Tsang. 3d display by binary computer-generated holograms with localized random down-sampling and adaptive intensity accumulation. *Optics Express*, 28(17):24526–24537, 2020. 78

[81] Ruoshi Liu, Chengzhi Mao, Purva Tendulkar, Hao Wang, and Carl Vondrick. Landscape learning for neural network inversion. In *the International Conference on Computer Vision (ICCV)*, pages 2239–2250, 2023. 79

[82] Tian Yu Liu, Parth Agrawal, Allison Chen, Byung-Woo Hong, and Alex Wong. Monitored distillation for positive congruent depth completion. In *European Conference on Computer Vision (ECCV)*, pages 35–53. Springer, 2022. xiii, 29, 31

[83] Chenguang Ma, Jinli Suo, Qionghai Dai, Ramesh Raskar, and Gordon Wetzstein. High-rank coded aperture projection for extended depth of field. In *the IEEE International Conference on Computational Photography (ICCP)*, pages 1–9. IEEE, 2013. xviii, 52, 62, 63

[84] Andrew Maimone, Andreas Georgiou, and Joel S Kollin. Holographic near-eye displays for virtual and augmented reality. *ACM Transactions on Graphics (TOG)*, 36(4), 2017. 7, 9, 54, 71, 79, 80

[85] Ghaith Makey, Özgün Yavuz, Denizhan K Kesim, Ahmet Turnalı, Parviz Elahi, Serim Ilday, Onur Tokel, and F Ömer Ilday. Breaking crosstalk limits to dynamic holography using orthogonality of high-dimensional random vectors. *Nature Photonics*, 13(4):251–256, 2019. 53, 55

[86] Michal Makowski, Izabela Ducin, Maciej Sypek, Agnieszka Siemion, Andrzej Siemion, Jaroslaw Suszek, and Andrzej Kolodziejczyk. Color image projection based on fourier holograms. *Optics Letters*, 35(8):1227–1229, 2010. 7, 9, 52

[87] Michal Makowski, Izabela Ducin, Karol Kakarenko, Jaroslaw Suszek, Maciej Sypek, and Andrzej Kolodziejczyk. Simple holographic projection in color. *Optics Express*, 20(22):25130–25136, 2012. 7, 9, 52

[88] Hidetoshi Mannami, Ryusuke Sagawa, Yasuhiro Mukaigawa, Tomio Echigo, and Yasushi Yagi. High dynamic range camera using reflective liquid crystal. In *the International Conference on Computer Vision (ICCV)*, pages 1–8. IEEE, 2007. 17

[89] Kazunobu Masuda, Yusuke Saita, Ryusuke Toritani, Peng Xia, Kouichi Nitta, and Osamu Matoba. Improvement of image quality of 3d display by using optimized binary phase modulation and intensity accumulation. *Journal of Display Technology*, 12(5):472–477, 2016. 39

[90] Nathan Matsuda, Oliver Cossairt, and Mohit Gupta. Mc3d: Motion contrast 3d scanning. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2015. 3, 37

[91] Sagi Monin, Aswin C Sankaranarayanan, and Anat Levin. Analyzing phase masks for wide étendue holographic displays. In *the IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2022. xvii, 52, 53, 56, 57, 58, 82

[92] Momoyo Nagase, Daisuke Iwai, and Kosuke Sato. Dynamic defocus and occlusion compensation of projected imagery by model-based optimal projector selection in multi-projection environment. *Virtual Reality*, 15:119–132, 2011. 52

[93] Gaku Narita, Yoshihiro Watanabe, and Masatoshi Ishikawa. Dynamic projection mapping onto deforming non-rigid surface using deformable dot cluster marker. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 23(3):1235–1248, 2016. 51

[94] Shree K Nayar and Vlad Branzoi. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *the International Conference on Computer Vision (ICCV)*, volume 2, pages 1168–1175, 2003. 17

[95] Shree K Nayar, Harish Peri, Michael D Grossberg, and Peter N Belhumeur. A projection system with radiometric compensation for screen imperfections. In *the International Workshop on Projector-Camera Systems (PROCAMS)*, volume 3. IEEE, 2003. 17

[96] Shree K Nayar, Vlad Branzoi, and Terrance E Boult. Programmable imaging using a digital micromirror array. In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages I–I. IEEE, 2004. 17

[97] Matthew O'Toole, Ramesh Raskar, and Kiriakos N Kutulakos. Primal-dual coding to probe light transport. *ACM Transactions on Graphics (TOG)*, 31

(4):39–1, 2012. 3, 37

[98] Matthew O'Toole, Felix Heide, Lei Xiao, Matthias B Hullin, Wolfgang Heidrich, and Kiriakos N Kutulakos. Temporal frequency probing for 5d transient analysis of global light transport. *ACM Transactions on Graphics (TOG)*, 33(4):1–11, 2014. 22

[99] Matthew O'Toole, Supreeth Achar, Srinivasa G. Narasimhan, and Kiriakos N Kutulakos. Homogeneous codes for energy-efficient illumination and imaging. *ACM Transactions on Graphics (TOG)*, 34(4):1–13, 2015. 17, 38, 44, 52, 81

[100] Scott P Overmann. Thermal design considerations for portable dlp tm projectors. *High-Density Interconnect and Systems Packaging, Santa Clara, CA, USA*, pages 17–20, 2001. 2, 15, 37

[101] Yuji Oyamada and Hideo Saito. Focal pre-correction of projected image for deblurring screen image. In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2007. 52

[102] Nitish Padmanaban, Yifan Peng, and Gordon Wetzstein. Holographic near-eye displays based on overlap-add stereograms. *ACM Transactions on Graphics (TOG)*, 38(6):1–13, 2019. 8, 53, 54

[103] Andrew D. Payne, Adrian A. Dorrington, Michael J. Cree, and Dale A. Carnegie. Improved measurement linearity and precision for amcw time-of-flight range imaging cameras. *Applied Optics*, 49(23):4392–4403, 2010. 22

[104] Yifan Peng, Suyeon Choi, Nitish Padmanaban, and Gordon Wetzstein. Neural holography with camera-in-the-loop training. *ACM Transactions on Graphics (TOG)*, 39(6):1–14, 2020. 7, 9, 11, 13, 54, 70, 77, 79

[105] Yifan Peng, Suyeon Choi, Jonghyun Kim, and Gordon Wetzstein. Speckle-free holography with partially coherent light sources and camera-in-the-loop calibration. *Science Advances*, 7(46), 2021. 78

[106] Rafael Piestun, Boris Spektor, and Joseph Shamir. On-axis binary-amplitude computer generated holograms. *Optics Communications*, 136(1-2):85–92, 1997. 39

[107] Paolo Pozzi, Dean Wilding, Oleg Soloviev, Hans Verstraete, Laurens Bliek, Gleb Vdovin, and Michel Verhaegen. High speed wavefront sensorless aberration correction in digital micromirror based confocal microscopy. *Optics Express*, 25(2):949–959, 2017. 44

[108] Yingsi Qin, Wei-Yu Chen, Matthew O'Toole, and Aswin C Sankaranarayanan. Split-lohmann multifocal displays. *ACM Transactions on Graph-*

ics (TOG), 42(4):1–18, 2023. 82

[109] Yaadhav Raaj, Siddharth Ancha, Robert Tamburo, David Held, and Srinivasa G. Narasimhan. Exploiting & refining depth distributions with triangulation light curtains. In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7434–7442, 2021. 42

[110] Jochen Radmer, Pol Moser Fusté, Henning Schmidt, and Jorg Kruger. Incident light related distance error study and calibration of the pmd-range imaging camera. In *the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1–6. IEEE, 2008. 23

[111] Malcolm Reynolds, Jozef Doboš, Leto Peel, Tim Weyrich, and Gabriel J Brostow. Capturing time-of-flight data with confidence. In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 945–952. IEEE, 2011. 23

[112] Yoav Y Schechner and Nahum Kiryati. Depth from defocus vs. stereo: How different really are they? *International Journal of Computer Vision (IJCV)*, 39:141–162, 2000. xix, 68, 69

[113] Florian Schiffers, Praneeth Chakravarthula, Nathan Matsuda, Grace Kuo, Ethan Tseng, Douglas Lanman, Felix Heide, and Oliver Cossairt. Stochastic light field holography. In *the IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2023. 8, 71, 77, 78

[114] Florian Schiffers, Grace Kuo, Nathan Matsuda, Douglas Lanman, and Oliver Cossairt. Holochrome: Polychromatic illumination for speckle reduction in holographic near-eye displays. *ACM Transactions on Graphics (TOG)*, 44 (3), May 2025. 78, 82

[115] Mark Sheinin and Yoav Y Schechner. Depth from texture integration. In *the IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2019. 65

[116] Liang Shi, Beichen Li, Changil Kim, Petr Kellnhofer, and Wojciech Matusik. Towards real-time photorealistic 3d holography with deep neural networks. *Nature*, 591(7849):234–239, 2021. 54, 70, 79

[117] Liang Shi, Donghun Ryu, and Wojciech Matusik. Ergonomic-centric holography: Optimizing realism, immersion, and comfort for holographic display. *Laser & Photonics Reviews*, 18(4):2300651, 2024. 54

[118] Tomoyoshi Shimobaba, Takashi Kakue, and Tomoyoshi Ito. Real-time and low speckle holographic projection. In *the IEEE International Conference on Industrial Informatics (INDIN)*, pages 732–741. IEEE, 2015. 7, 9, 52

[119] Gal Shtendel and Ayush Bhandari. Hdr-tof: Hdr time-of-flight imaging via

modulo acquisition. In *the IEEE International Conference on Image Processing (ICIP)*, pages 3808–3812. IEEE, 2022. 23

[120] Tal Simchony, Rama Chellappa, and Min Shao. Direct analytical methods for solving poisson equations in computer vision problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 12(5):435–446, 1990. 13

[121] SME. Structured Vision in Manufacturing. `https://www.sme.org/structured-vision-manufacturing`, 2025. 2

[122] Marija Strojnik and Michelle K Scholl. Radiometry. In *Advanced Optical Instruments and Techniques*, pages 459–516. CRC Press, 2017. 2

[123] Dustin Stuart, Oliver Barter, and Axel Kuhn. Fast algorithms for generating binary holograms. *arXiv preprint arXiv:1409.1841*, 2014. 13, 39

[124] Zhanghao Sun, Jian Wang, Yicheng Wu, and Shree Nayar. Seeing far in the dark with patterned flash. In *the European Conference on Computer Vision (ECCV)*, pages 709–727. Springer, 2022. 16, 17

[125] Varun Sundar, Sizhuo Ma, Aswin C. Sankaranarayanan, and Mohit Gupta. Single-photon structured light. In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17865–17875, June 2022. 38

[126] Varun Sundar, Sizhuo Ma, Aswin C Sankaranarayanan, and Mohit Gupta. Single-photon structured light. In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17865–17875, 2022. 3, 37

[127] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Nature, 2022. 1, 5

[128] Brevin Tilmon, Zhanghao Sun, Sanjeev J Koppal, Yicheng Wu, Georgios Evangelidis, Ramzi Zahreddine, Gurunandan Krishnan, Sizhuo Ma, and Jian Wang. Energy-efficient adaptive 3d sensing. In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5054–5063, 2023. 16, 52, 71, 81

[129] Ethan Tseng, Grace Kuo, Seung-Hwan Baek, Nathan Matsuda, Andrew Maimone, Florian Schiffers, Praneeth Chakravarthula, Qiang Fu, Wolfgang Heidrich, Douglas Lanman, et al. Neural étendue expander for ultra-wide-angle high-fidelity holographic display. *Nature Communications*, 15(1): 2907, 2024. xvii, 53, 56, 57, 58

[130] Mayuka Tsuji, Hiroyuki Kubo, Suren Jayasuriya, Takuya Funatomi, and Yasuhiro Mukaigawa. Touch sensing for a projected screen using slope disparity gating. *IEEE Access*, 9:106005–106013, 2021. 38, 42, 44, 49, 52, 67

[131] Sergey Turtaev, Ivo T Leite, Kevin J Mitchell, Miles J Padgett, David B Phillips, and Tomáš Čižmár. Comparison of nematic liquid-crystal and dmd based spatial light modulation in complex photonics. *Optics Express*, 25(24): 29874–29884, 2017. 13

[132] Tomoki Ueda, Hiroyuki Kubo, Suren Jayasuriya, Takuya Funatomi, and Yasuhiro Mukaigawa. Slope disparity gating using a synchronized projector-camera system. In *the IEEE International Conference on Computational Photography (ICCP)*, pages 1–9. IEEE, 2019. xvi, 38, 42, 44, 46, 52, 67

[133] Jiabao Wang, Jun Wang, Jie Zhou, Yuqi Zhang, and Yang Wu. Crosstalk-free for multi-plane holographic display using double-constraint stochastic gradient descent. *Optics Express*, 31(19):31142–31157, 2023. 53, 55

[134] Jian Wang, Joseph Bartels, William Whittaker, Aswin C. Sankaranarayanan, and Srinivasa G. Narasimhan. Programmable triangulation light curtains. In *the European Conference on Computer Vision (ECCV)*, pages 19–34, 2018. xv, 38, 41, 42, 43, 44, 45, 67

[135] Lihui Wang, Hongjin Xu, Satoshi Tabata, Yunpu Hu, Yoshihiro Watanabe, and Masatoshi Ishikawa. High-speed focal tracking projection based on liquid lens. In *ACM SIGGRAPH Emerging Technologies*, pages 1–2. Association for Computing Machinery, New York, NY, USA, 2020. 52

[136] Zi Wang, Tao Chen, Qiyang Chen, Kefeng Tu, Qibin Feng, Guoqiang Lv, Anting Wang, and Hai Ming. Reducing crosstalk of a multi-plane holographic display by the time-multiplexing stochastic gradient descent. *Optics Express*, 31(5):7413–7424, 2023. 53, 55

[137] Michael Watts, Christopher Poulton, Matthew Byrd, and Greg Smolka. Lidar on a Chip Puts Self-Driving Cars in the Fast Lane. `https://spectrum.ieee.org/lidar-on-a-chip`, 2025. 2

[138] Mian Wei, Navid Sarhangnejad, Zhengfan Xia, Nikita Gusev, Nikola Katic, Roman Genov, and Kiriakos N Kutulakos. Coded two-bucket cameras for computer vision. In *the European Conference on Computer Vision (ECCV)*, pages 54–71, 2018. 3, 37

[139] Taeseong Woo, Hye Yun Kim, Su Yeon Kim, Byungjae Hwang, Cheolwoo Ahn, Seok-Kyu Kwon, Jae-Ick Kim, and Jung-Hoon Park. High-throughput high-dynamic range imaging by spatiotemporally structured illumination. *APL Photonics*, 7(10):106106, 2022. 17

[140] Yicheng Wu, Vivek Boominathan, Xuan Zhao, Jacob T Robinson, Hiroshi Kawasaki, Aswin Sankaranarayanan, and Ashok Veeraraghavan. Freecam3d: Snapshot structured light 3d with freely-moving cameras. In

the European Conference on Computer Vision (ECCV), pages 309–325. Springer, 2020. 52

[141] Hongjin Xu, Lihui Wang, Satoshi Tabata, Yoshihiro Watanabe, and Masatoshi Ishikawa. Extended depth-of-field projection method using a high-speed projector with a synchronized oscillating variable-focus lens. *Applied Optics*, 60(13):3917–3924, 2021. xviii, 52, 54, 61, 63

[142] Panpan Yu, Yifan Liu, Ziqiang Wang, Jinyang Liang, Xingsi Liu, Yinmei Li, Chengwei Qiu, and Lei Gong. Ultrahigh-density 3d holographic projection by scattering-assisted dynamic holography. *Optica*, 10(4):481–490, 2023. 53, 55

[143] Jingzhao Zhang, Nicolas Pégard, Jingshan Zhong, Hillel Adesnik, and Laura Waller. 3d computer-generated holography by non-convex optimization. *Optica*, 4(10):1306–1313, 2017. 53, 55

[144] Li Zhang and Shree Nayar. Projection defocus analysis for scene capture and image display. In *ACM SIGGRAPH*, pages 907–915. Association for Computing Machinery, New York, NY, USA, 2006. 3, 51, 52

[145] Zhengyun Zhang and Marc Levoy. Wigner distributions and how they relate to the light field. In *the IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2009. 8