# Holodepth: Programmable Depth-Varying Projection via Computer-Generated Holography

Dorian Chan[1], Matthew O'Toole[1], Sizhuo Ma[2], and Jian Wang[2]*

[1] Carnegie Mellon University
{dorianchan,mpotoole}@cmu.edu
[2] Snap Inc.
{sma,jwang4}@snap.com

**Abstract.** Typical projectors are designed to programmably display 2D content at a single depth. In this work, we explore how to engineer a depth-varying projector system that is capable of forming desired patterns at multiple depths. To this end, we leverage a holographic approach, but a naïve implementation of such a system is limited in its depth programmability. Inspired by recent work in near-eye displays, we add a lens array to a holographic projector to maximize the depth variation of the projected content, for which we propose an optimization-driven calibration method. We demonstrate a number of applications using this system, including novel 3D interfaces for future wearables, privacy-preserving projection, depth sensing, and light curtains.

## 1 Introduction

Projectors have found widespread use in the real world today. They are ubiquitous in entertainment and education, used to show movies and messages on large screens for big audiences [6]. In industrial tasks, they are widely used to capture extremely accurate depth measurements through structured light [18]. In the modern era, they are rapidly being integrated into wearable devices for augmented reality tasks like screenless interfaces [1], 3D scanning [42] and more.

A traditional projector is designed to display content on a 2D planar screen at a particular depth. To do so, a digital-micromirror device (DMD) or liquid crystal display (LCD) programmably masks light from a bulb, which is then imaged into the scene by a projection lens to the correct depth. The pattern at any another depth is simply a blurred version of this in-focus image.

While such a projection model is practical for many tasks, more explicit programmable control of the projected content at different depths, *e.g.*, the projected pattern morphs from one desired image to another with depth, could be very useful. Such a general purpose, programmable depth-varying projector would find immediate application in both industrial and artistic settings. For example, such a system could be used to create novel screenless 3D interfaces,

---

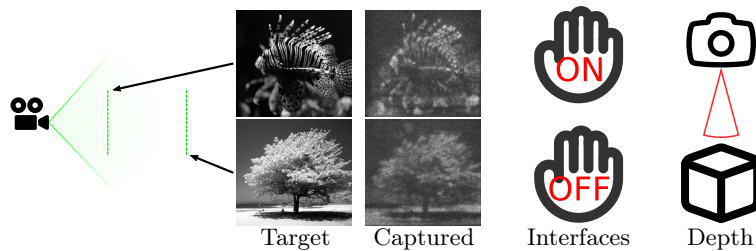* Work partially completed while D. Chan was a Snap Research intern. Sample code shared online.

**Fig. 1:** We leverage a holographic projector for forming depth-varying patterns. In other words, we can program unique content at multiple depths per pixel simultaneously. This capability could be useful for future interfaces, depth sensing, and more.

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 stage, a foreground image could be shown on moving actors, while a different background image automatically appears behind them. In industrial applications, an appropriate depth-varying pattern could be used as a depth cue, without the need for a stereo baseline. Such a device could additionally form multiple light curtains flexibly and simultaneously unlike past work [4, 13, 51].

However, existing projectors that could tackle these applications come with various tradeoffs. Coded aperture [20, 24, 25, 31, 54] and light field projectors [21] are limited in programmability, and struggle to form arbitrary content at different planes (Fig. 2(a)), *e.g.*, coded apertures are restricted to an intensity convolution between an in-focus image and a scaled version of the aperture pattern [20]. Temporal multiplexing can improve the degrees of freedom [21, 31], 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 [5, 37], 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 [52, 55] 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 (Fig. 2(b)). Simultaneously, producing content at multiple depths requires a high level of temporal multiplexing, again decreasing framerate and increasing bandwidth requirements.

Holographic projection provides an alternative that avoids many of these pitfalls. For one, a holographic approach can project depth-dependent content using just a single pattern on a single SLM [29,33,50,53,56,57], avoiding the need for time multiplexing. Additionally, holographic setups have far more degrees of freedom than incoherent approaches [38], increasing resolution and programmability (Fig. 2(d)). Moreover, such systems are light redistributive and can form

patterns without blocking light, increasing dynamic range [13–15, 47]. Due to these reasons, a holographic depth-varying projector is a promising solution.

In practice, however, 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 [56], 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. We derive that this relationship between depth variation and field-of-view is fundamentally connected to the *étendue* of a holographic projector — 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 [2, 8, 28, 30, 36], we introduce a lens array into the optical path of a typical holographic projector to increase depth variation, enabling far more complex patterns than past work [56]. 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, that is capable of achieving all of the aforementioned applications in a single setup.

The contributions of this work include:

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

## 2    Related work

Historically, holographic projectors have primarily been explored for their potential compact form factor [7, 34, 35, 46], as they can create a large field-of-view without the need for a lens. In recent work, holographic projectors have seen a resurgence in computer vision for their inherent light redistribution properties [13–15, 47], which allow for the projection of extremely bright patterns. This increased contrast enables fast structured light [13], eye-safe 3D sensing [47], and longer range continuous-wave time-of-flight (CWTOF) imaging [14]. Some past work has touched on using holographic projectors to account for projector defocus [46] by digitally refocusing content, but the level of improvement is small thanks to the already limited depth variation and therefore defocus of a typical system. In contrast, our work increases the depth variation of a holographic projector, which we use to engineer a programmable depth-varying projector.

In a different domain, near-eye display research [10–12, 16, 17, 23, 27, 32, 38, 39, 44, 45] has shown impressive results using holography to produce 3D accommodation cues, replicating depth maps and focal stacks with the target of realistic
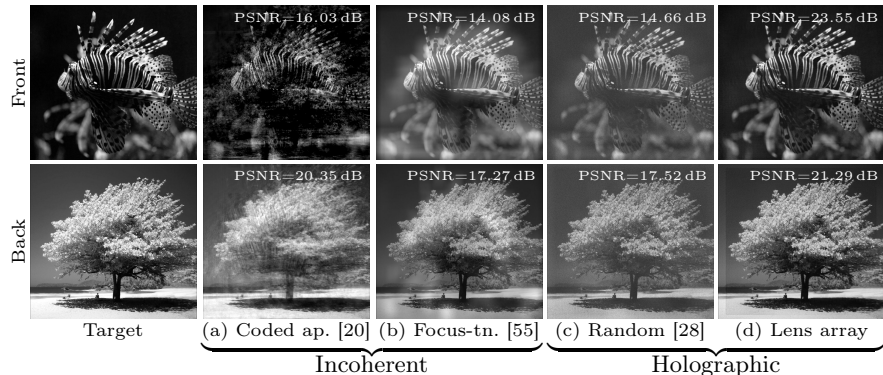
**Fig. 2: Simulated comparison of depth-varying projection solutions.** To compare with other potential configurations, we simulate the projection of unique content at two planes. A coded aperture setup with two SLMs [20] (implem. details in supp.), struggles to form the desired content as shown in **(a)**. A time-multiplexed focus-tunable lens and high-speed projector **(b)** does better visually, but low frequency errors occur thanks to crosstalk on top of the practical challenges associated with time-multiplexing (Sec. 1). A holographic system that uses a custom random binary phase mask [28] **(c)** results in loss of contrast. The off-the-shelf lens array used in this work **(d)** performs similarly to the time-multiplexed case with just a single SLM pattern.

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 systems 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 [33,50,57]. Time multiplexing can be applied to better disambiguate content at different depths [29,53], but it does not fundamentally improve depth variation. Most akin to our real system, Yu *et al.* [56] 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.

## 3    3D holographic projection

We aim to programmably project content at different depths in the scene using a holographic projector. We start by discussing the image formation model of a normal holographic projector, and how it can be extended to project depth-varying content. Then, we discuss depth variation and étendue-expanding optics.
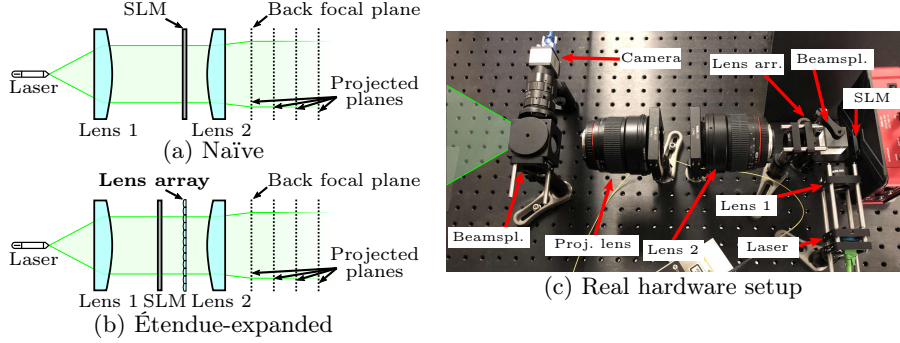
Fig. 3: **System diagram.** **(a)** shows a naïve approach, while **(b)** shows an étendue-expanded version with an additional lens array. **(c)** shows our real hardware prototype.

### 3.1   Projection model

Consider the setup given in Fig. 3(a). Laser light collimated by Lens 1 illuminates a phase SLM at the front focal plane of Lens 2. The Fourier transform $\mathcal{F}$ of the SLM modulation $U_{\mathrm{SLM}} = e^{j\phi_{\mathrm{SLM}}}$ appears at the back focal plane of Lens 2:

$$U_{\mathrm{back}} = \mathcal{F}\{U_{\mathrm{SLM}}\}. \tag{1}$$

This Fourier relationship effectively redirects light from dark to bright regions of output patterns, increasing dynamic range over a typical projector [13–15, 47].

In practice, different depths in the scene will receive slightly different content, which can be predicted using wave optics. For example, the angular spectrum method (ASM) can be used to simulate the light propagation between some plane $U$ and some other plane $z$ away [19], with $S = \mathcal{F}\{U\}$:

$$\mathcal{P}(U, z) = \iint_{\Omega} S(f_x, f_y) e^{j\frac{2\pi}{\lambda} z \sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}} e^{j2\pi(f_x x + f_y y)} \, df_x \, df_y, \tag{2}$$

$$g_{\mathrm{multiplane}}(U, z) = \mathcal{P}(\mathcal{F}(U), z), \tag{3}$$

where $\lambda$ is the laser wavelength, $f_x$ and $f_y$ denote Fourier transform frequencies, and $\Omega$ is the support of $S$, handling evanescent waves. This operator $\mathcal{P}$ can be expressed as a simple elementwise multiplication in the frequency domain, allowing for efficient computation. Eq. (3) determines the appearance of the projector pattern for different depths in the scene. We can also attempt to invert this model in order to project desired depth-varying content. In our work, we solve the following optimization problem using Adam [26]:

$$\min_{\phi_{\mathrm{SLM}}} \quad \sum_k \mathcal{L}\left(\mathcal{D}\left(|g(U_{\mathrm{SLM}}, z_k))|^2\right), I_k\right), \tag{4}$$

where $g(\cdot)$ is the desired forward model such as Eq. (3), $\mathcal{L}$ is some loss function, $\mathcal{D}$ is a resizing operator that matches the input and output degrees of freedom [36], and $I_k$ and $z_k$ denote the intensity and depth of the $k$th target.
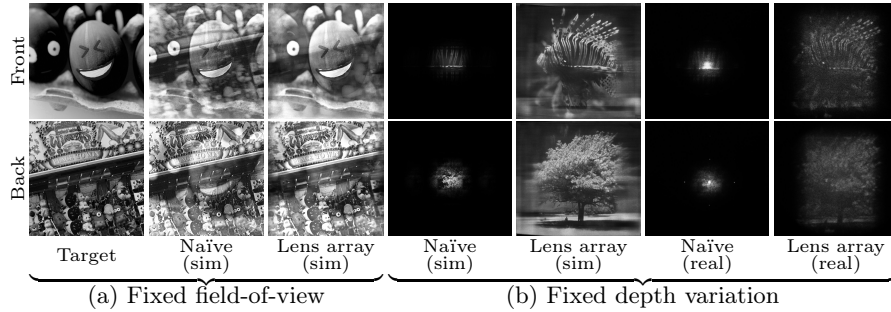
Fig. 4: **Étendue expansion with a lens array**. We visualize the importance of étendue for depth-varying projection. For a given field-of-view **(a)**, without étendue expansion, there is a significant amount of crosstalk between two target depth planes. With an étendue-expanding lens array, the quality of projection is much higher. For a given depth variation **(b)**, the effective field-of-view is very small with a typical SLM. With étendue-expansion, the field-of-view is much larger.

### 3.2  Étendue-expanded 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 $\delta$ can be calculated as [28]:

$$\sin\theta = \frac{\lambda}{2\delta}, A = \delta^2 N_x N_y, \mathcal{E} = 4A\sin^2\theta = \lambda^2 N_x N_y, \tag{5}$$

where $\theta$ 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 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 [30], but such an approach requires an extremely fast SLM which is not currently readily available [36]. 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 [2, 8, 28, 36]. Lens arrays serve as a low-cost alternative [8, 36] to these custom-fabricated phase masks [2, 28].

For holographic projectors, 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 $\partial\{|\mathcal{P}(U,z)|^2\}/\partial z$:

$$\iint_\Omega \iint_\Omega \left( \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2} - \sqrt{1 - (\lambda a_x)^2 - (\lambda a_y)^2} \right)$$
$$\cdot P^*(f_x, f_y, x, y, z) P(a_x, a_y, x, y, z) \mathrm{d}f_x \mathrm{d}f_y \mathrm{d}a_x \mathrm{d}a_y \tag{6}$$

where $P(\cdot)$ contains all terms inside the integral of Eq. (2). 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 Eq. (3), 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, a holographic projector cannot project patterns that significantly vary in depth without an excessively small FOV. For a 1920×1200 resolution SLM with 8 µm pixel pitch, we find that projecting unique content spread 15 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 [2, 28] or lens array [8, 36]. In our simulations (Fig. 2(c),(d)), we find that a lens array preserves more contrast than a random phase mask [28] 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 $z_{\mathrm{array}}$ as shown in Fig. 3(b). Our étendue-enhanced forward model can then be written as:

$$g_{\mathrm{enhanced}}(U, z) = \mathcal{P}(\mathcal{F}(\mathcal{M}(\mathcal{P}(U, z_{\mathrm{array}}))), z), \tag{7}$$

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 Eq. (4). With an array with lens pitch 1.0 mm×1.4 mm and focal length 4.7 mm, we can expand FOV to about 15°. We visualize the results in Fig. 4.

**Lens array effects** In the context of near-eye displays, Monin *et al.* [36] 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. These effects are visualized in the supplement.

**Calibrating the lens array** In practice, aligning étendue-expanding optics can be a challenging process, as tiny axial (<1 mm) and lateral misalignments (≈ 1 SLM pixel) can cause a drastic reduction in output quality (Fig. 5(a)). Existing approaches have typically reduced SLM pixel resolution and introduced precise alignment processes to mitigate these effects [2, 28]. Additionally, optical aberrations further reduce quality [8], but past work has primarily ignored these effects, assuming ideal lenses and masks [2, 8, 28]. Such artifacts are exacerbated in low cost off-the-shelf optics, like the lens array we use in Sec. 4.1.
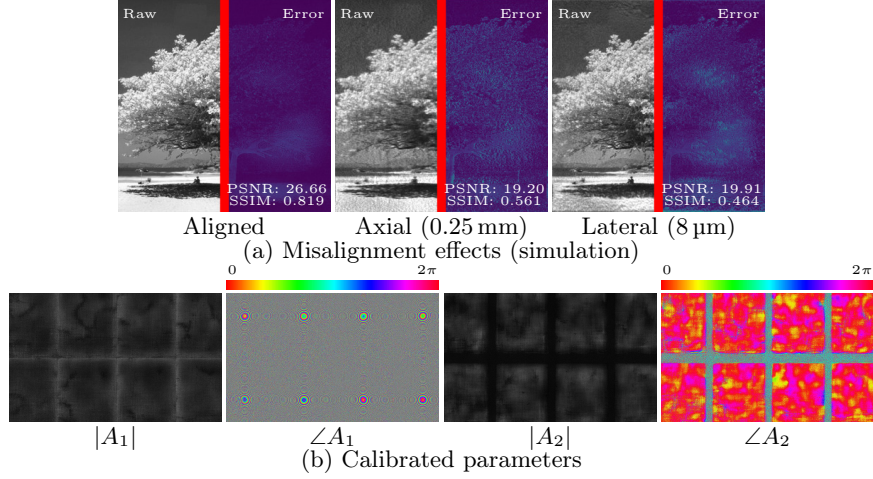
Aligned            Axial (0.25 mm)         Lateral (8 µm)
(a) Misalignment effects (simulation)

$|A_1|$            $\angle A_1$            $|A_2|$            $\angle A_2$
(b) Calibrated parameters

**Fig. 5: 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 Eq. (8). $A_1$ learns the phase of the lenses, while $A_2$ seems to learn other distortions.

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 $z_{\mathrm{calib}}$ away. We can also simultaneously calibrate other non-idealities like SLM distortion and undiffracted light. Formally, this process is given by:

$$\min_{\substack{\mathcal{M}, A_{\mathrm{cam}}, A_{\mathrm{add}}, A_{\mathrm{SLM}}, \\ \mathcal{P}^{\mathrm{opt}}_{z_{\mathrm{calib}}}, \mathcal{P}^{\mathrm{opt}}_{z_{\mathrm{array}}}}} \quad \sum_k \mathcal{L}\left(\mathcal{D}\left(|g_{\mathrm{calib}}(U_k)|^2\right), I_k\right),$$

$$g_{\mathrm{calib}}(U) = \mathcal{P}\left(A_{\mathrm{add}} + A_{\mathrm{cam}} \cdot \mathcal{F}\left(\mathcal{M}\left(\mathcal{P}^{\mathrm{opt}}_{z_{\mathrm{array}}}(A_{\mathrm{SLM}} \cdot U)\right)\right), z_{\mathrm{calib}}\right),$$

$$(8)$$

where $U_k$ and $I_k$ are the corresponding SLM pattern and capture pair. $A_{\mathrm{cam}}$, $A_{\mathrm{SLM}}$ and $A_{\mathrm{add}}$ denote learnable complex modulations, and $\mathcal{P}^{\mathrm{opt}}_z$ denotes an optimizable propagation kernel initialized with the kernel for propagating $z$ [16,23]. Inspired by the ABCD matrix for thick lenses, we represent $\mathcal{M}$ as an optimizable multiplication, propagation and multiplication $\mathcal{M}(U) = A_2 \cdot \mathcal{P}^{\mathrm{opt}}_{z_{\mathrm{thickness}}}(A_1 \cdot U)$, where $z_{\mathrm{thickness}}$ is the estimated thickness of the lens array. We describe this multilayer model in more detail in the supplement.

Using Adam [26], 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 Fig. 5(b). Please refer to the supplement for more details.

(a) Defocus compensation                    (b) 3D interface

(c) Multi-layered display

Scene visualization

(d) Privacy-preserving projector

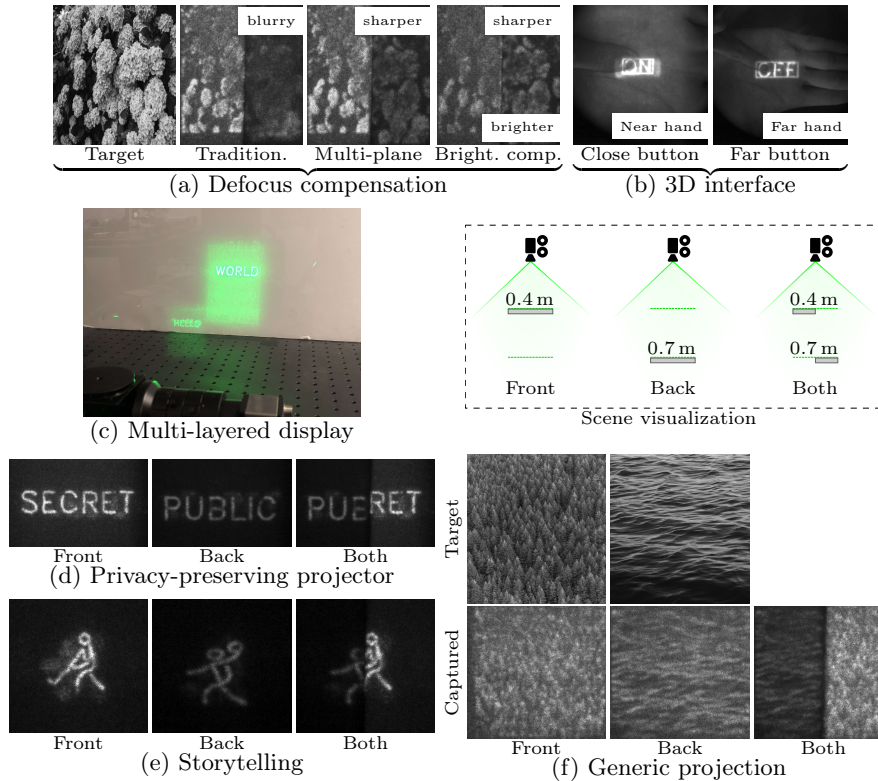(e) Storytelling                    (f) Generic projection

**Fig. 6: Programmable depth-varying projection**. 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.

## 4 Results

### 4.1 Hardware implementation

We show our hardware prototype in Fig. 3(c). Our laser is a 530 nm Coherent Sapphire LPX. We use a Thorlabs Exulus HD-2 SLM, with 1920×1200 resolution and 8 µm pitch. For étendue expansion, we use a Thorlabs PMMA Microlens Array (MLA1M), where each microlens has pitch 1.0 mm×1.4 mm and focal length 4.7 mm. Simulating each microlens without aliasing requires a resolution of 2.5 µm×1.78 µm, translating to an expansion in étendue by 3.2×4.5. We therefore simulated our system at 9600×6000 with 1.6 µm pitch. We used a 75 mm achromatic doublet for Lens 1, and two 85 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 mm $f/1.8$ lens. The captured scenes are between 0.4 m and 0.7 m away from the system. At 0.7 m, the projected pattern spans roughly 0.1 m×0.15 m.

## 4.2   Applications

**Multi-plane projection** Our projector can simultaneously project different content to multiple planes. As shown in Fig. 6(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 [14], 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 [1], where users interact with the projection onto another object. As shown in Fig. 6(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 [3] for the formation of pseudo-3D content (Fig. 6(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 (Fig. 6(d)).

We envision such a projector as a useful tool for artists and creatives (Fig. 6(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 Fig. 7(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 [25, 43].

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 [43]:

$$E(z) = \sum_{\mathbf{x}} D\left(z_{\mathbf{x}}\right) + \lambda \sum_{\mathbf{x}, \mathbf{x}'} V_{\mathbf{x}, \mathbf{x}'}\left(z_{\mathbf{x}}, z_{\mathbf{x}'}\right) \tag{9}$$

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 $\lambda$ 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

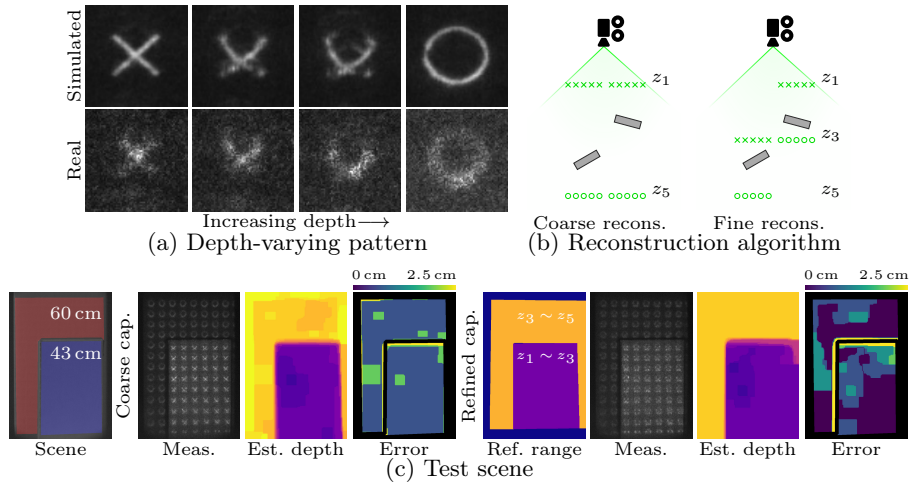(a) Depth-varying pattern    (b) Reconstruction algorithm

(c) Test scene

Fig. 7: **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.

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, 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 Eq. (4) to determine the appropriate SLM pattern. We can then use the resulting capture in Eq. (9) to get a refined depth measurement.

In Fig. 7(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 Fig. 7(c), we apply this procedure to a test scene, where the refined measurement reduces error. In Fig. 8, 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.
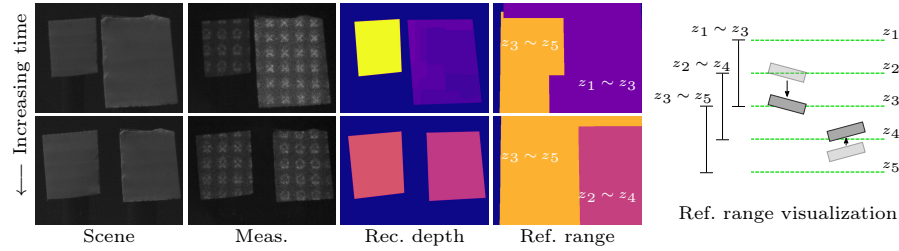
Fig. 8: **Adapting the depth-varying pattern for moving objects**. Instead of using a coarse capture, we can use a previous depth estimate in the case of two moving objects. The left object moves towards the projector, while the right object moves away. Each object is captured with a different set of two planes, which is adjusted as the objects move. The rightmost column illustrates these planes.
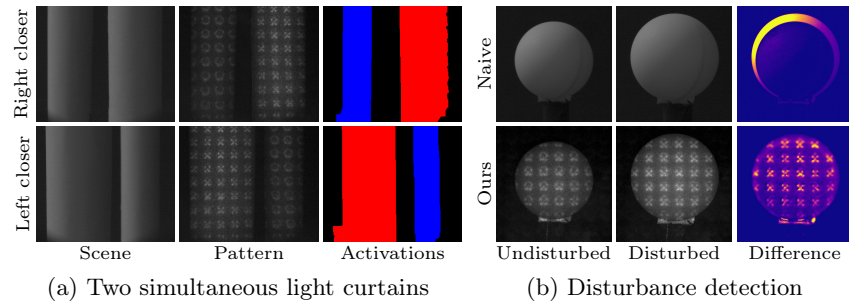


(a) Two simultaneous light curtains      (b) Disturbance detection

Fig. 9: **Using depth-varying patterns to form coarse light curtains**. In **(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. In **(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.

**Programmable light curtains** Our depth-varying patterns can also be used to form programmable light curtains around objects [4, 13, 48, 49, 51]. 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 [4, 13, 48, 49, 51], 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 Fig. 9. 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 [13]. The change in projected patterns provides a strong cue as to when objects are moved.

**Depth cue comparison**  Using geometric optics approximations (derivation in the supplement), we can estimate the resolution of our holographic depth variation cue to be roughly equivalent to a stereo baseline of $i_d \tan \theta_p$, where $\frac{1}{i_d} = \frac{1}{f_p} - \frac{1}{d}$ and $\sin \theta_p = \frac{\delta N}{2f_2}$. $f_2$ and $f_p$ are the focal lengths of Lens 2 and the projection lens, and $d$ is the target distance. For example, for a point $0.5\,\mathrm{m}$ away, the effective baseline of our prototype is $\approx 9\,\mathrm{mm}$.

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 [40]. Furthermore, structured light systems struggle with a tradeoff between depth-of-field 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 [22]. 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 projectors are being applied in structured light [13, 47] and CWTOF [14], 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.

## 5    Limitations

In practice, a number of drawbacks arise when using holography. For one, in order to accurately model coherent light transport through our lens array, we need to simulate propagation at $9600{\times}6000$ pixels. 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 Eq. (4). One could also apply a non-iterative approach to reduce computation [32, 47], at the cost of pattern quality.

Furthermore, unavoidable noise-like artifacts called speckle appear in the projections, reducing contrast and sharpness. Exacerbating these effects, 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 speckle. We posit that one cause is because the actual light propagation in the system is much more complicated than our simple model in Eq. (4). Properly accounting for more complex effects via pupil-aware [9, 41] and neural-enhanced [12, 16, 17, 23, 39] modeling 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 rate [16] or multiple SLMs and sources at the cost of extra system complexity [27].

In general, even with the lens array, the depth variation produced by our projector is still limited. In our experiments, two target planes need to be separated by 5 mm before the projector lens in order to reliably project unique content. Blur begins to appear with less separation and more target planes. However, temporal multiplexing [29, 53] could help distinguish patterns at different depths, specific target patterns more amenable to propagation could be easier to resolve, and better expansion or future high-resolution SLMs could greatly increase étendue.

## 6   Conclusion and future work

In our work, we leverage a holographic projector to create depth-varying patterns. To enhance this functionality, we place an étendue-expanding lens array into the system, and introduce a novel calibration method for it. 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.

There are many future directions. For one, there is significant room for improvement in closing the simulation-to-real gap, as mentioned in Sec. 5. Properly integrating neural models that can capture the non-idealities that a simpler theoretical model cannot could significantly improve pattern quality [12, 16, 17, 23, 39]. Alternatively, time multiplexing could be integrated to reduce speckle effects and crosstalk between different depths [16, 29, 53].

The étendue-expanding lens array is also a general purpose optic. It is likely that a custom scattering mask tailored to the depth-varying task of interest could improve performance [2], *e.g.* a phase mask optimized for depth sensing. Furthermore, the depth recovery pattern and algorithm discussed in Sec. 4.2 are extremely simple, and do not make full usage of the programmability of our proposed setup. It is highly probable that with further engineering, such as a deep learning-based end-to-end approach [54], the quality of depth measurements will be significantly improved with better illumination and reconstruction.

## Acknowledgements

## References

1. Humane previews first product in ted talk: https://hu.ma.ne/media/humane-previews-first-product-in-ted-talk, `https://hu.ma.ne/media/humane-previews-first-product-in-ted-talk`
2. Baek, S.H., Tseng, E., Maimone, A., Matsuda, N., Kuo, G., Fu, Q., Heidrich, W., Lanman, D., Heide, F.: Neural etendue expander for ultra-wide-angle high-fidelity holographic display. arXiv preprint arXiv:2109.08123 (2021)
3. Barnum, P.C., Narasimhan, S.G., Kanade, T.: A multi-layered display with water drops. In: ACM SIGGRAPH 2010 papers, pp. 1–7 (2010)
4. Bartels, J.R., Wang, J., Whittaker, W., Narasimhan, S.G.: Agile depth sensing using triangulation light curtains. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 7900–7908 (2019)
5. Bimber, O., Emmerling, A.: Multifocal projection: A multiprojector technique for increasing focal depth. IEEE Transactions on Visualization and Computer Graphics (TVCG) **12**(4), 658–667 (2006)
6. Brennesholtz, M.S., Stupp, E.H.: Projection displays (2008)
7. Buckley, E.: Holographic laser projection. Journal of display technology **7**(3), 135–140 (2011)
8. Chae, M., Bang, K., Yoo, D., Jeong, Y.: Étendue expansion in holographic near eye displays through sparse eye-box generation using lens array eyepiece. ACM Transactions on Graphics (TOG) **42**(4), 1–13 (2023)
9. Chakravarthula, P., Baek, S.H., Schiffers, F., Tseng, E., Kuo, G., Maimone, A., Matsuda, N., Cossairt, O., Lanman, D., Heide, F.: Pupil-aware holography. arXiv preprint arXiv:2203.14939 (2022)
10. Chakravarthula, P., Peng, Y., Kollin, J., Fuchs, H., Heide, F.: Wirtinger holography for near-eye displays. ACM Transactions on Graphics (TOG) **38**(6), 1–13 (2019)
11. Chakravarthula, P., Tseng, E., Fuchs, H., Heide, F.: Hogel-free holography. ACM Transactions on Graphics (TOG) **41**(5), 1–16 (2022)
12. Chakravarthula, P., Tseng, E., Srivastava, T., Fuchs, H., Heide, F.: Learned hardware-in-the-loop phase retrieval for holographic near-eye displays. ACM Transactions on Graphics (TOG) **39**(6), 1–18 (2020)
13. Chan, D., Narasimhan, S.G., O'Toole, M.: Holocurtains: Programming light curtains via binary holography. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 17886–17895 (2022)
14. Chan, D., O'Toole, M.: Light-efficient holographic illumination for continuous-wave time-of-flight imaging. In: ACM SIGGRAPH ASIA 2023 (2023)
15. Chao, B., Gopakumar, M., Choi, S., Wetzstein, G.: High-brightness holographic projection. Optics Letters **48**(15), 4041–4044 (2023)
16. Choi, S., Gopakumar, M., Peng, Y., Kim, J., O'Toole, M., Wetzstein, G.: Time-multiplexed neural holography: a flexible framework for holographic near-eye displays with fast heavily-quantized spatial light modulators. In: ACM SIGGRAPH 2022 Conference Proceedings. pp. 1–9 (2022)

17. Choi, S., Gopakumar, M., Peng, Y., Kim, J., Wetzstein, G.: Neural 3d holography: learning accurate wave propagation models for 3d holographic virtual and augmented reality displays. ACM Transactions on Graphics (TOG) **40**(6), 1–12 (2021)
18. Geng, J.: Structured-light 3d surface imaging: a tutorial. Advances in optics and photonics **3**(2), 128–160 (2011)
19. Goodman, J.W.: Introduction to Fourier Optics. McGraw-Hill physical and quantum electronics series, W. H. Freeman (2005)
20. Grosse, M., Wetzstein, G., Grundhöfer, A., Bimber, O.: Coded aperture projection. ACM Transactions on Graphics (TOG) **29**(3), 1–12 (2010)
21. Hirsch, M., Wetzstein, G., Raskar, R.: A compressive light field projection system. ACM Transactions on Graphics (TOG) **33**(4), 1–12 (2014)
22. Hu, Y., Chen, Q., Feng, S., Zuo, C.: Microscopic fringe projection profilometry: A review. Optics and Lasers in Engineering **135**, 106192 (2020)
23. Kavaklı, K., Urey, H., Akşit, K.: Learned holographic light transport. Applied Optics **61**(5), B50–B55 (2022)
24. Kawasaki, H., Horita, Y., Masuyama, H., Ono, S., Kimura, M., Takane, Y.: Optimized aperture for estimating depth from projector's defocus. In: 2013 International Conference on 3D Vision (3DV). pp. 135–142. IEEE (2013)
25. Kawasaki, H., Ono, S., Horita, Y., Shiba, Y., Furukawa, R., Hiura, S.: Active one-shot scan for wide depth range using a light field projector based on coded aperture. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 3568–3576 (2015)
26. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
27. Kuo, G., Schiffers, F., Lanman, D., Cossairt, O., Matsuda, N.: Multisource holography. ACM Transactions on Graphics (TOG) **42**(6), 1–14 (2023)
28. Kuo, G., Waller, L., Ng, R., Maimone, A.: High resolution étendue expansion for holographic displays. ACM Transactions on Graphics (TOG) **39**(4), 66–1 (2020)
29. Lee, B., Kim, D., Lee, S., Chen, C., Lee, B.: High-contrast, speckle-free, true 3d holography via binary cgh optimization. Scientific Reports **12**(1), 2811 (2022)
30. Lee, B., Yoo, D., Jeong, J., Lee, S., Lee, D., Lee, B.: Wide-angle speckleless dmd holographic display using structured illumination with temporal multiplexing. Optics Letters **45**(8), 2148–2151 (2020)
31. Ma, C., Suo, J., Dai, Q., Raskar, R., Wetzstein, G.: High-rank coded aperture projection for extended depth of field. In: IEEE International Conference on Computational Photography (ICCP). pp. 1–9. IEEE (2013)
32. Maimone, A., Georgiou, A., Kollin, J.S.: Holographic near-eye displays for virtual and augmented reality. ACM Transactions on Graphics (TOG) **36**(4), 1–16 (2017)
33. Makey, G., Yavuz, Ö., Kesim, D.K., Turnalı, A., Elahi, P., Ilday, S., Tokel, O., Ilday, F.Ö.: Breaking crosstalk limits to dynamic holography using orthogonality of high-dimensional random vectors. Nature Photonics **13**(4), 251–256 (2019)
34. Makowski, M., Ducin, I., Kakarenko, K., Suszek, J., Sypek, M., Kolodziejczyk, A.: Simple holographic projection in color. Optics Express **20**(22), 25130–25136 (2012)
35. Makowski, M., Ducin, I., Sypek, M., Siemion, A., Siemion, A., Suszek, J., Kolodziejczyk, A.: Color image projection based on fourier holograms. Optics Letters **35**(8), 1227–1229 (2010)
36. Monin, S., Sankaranarayanan, A.C., Levin, A.: Analyzing phase masks for wide étendue holographic displays. In: 2022 IEEE International Conference on Computational Photography (ICCP). pp. 1–12. IEEE (2022)

37. Nagase, M., Iwai, D., Sato, K.: Dynamic defocus and occlusion compensation of projected imagery by model-based optimal projector selection in multi-projection environment. Virtual Reality **15**, 119–132 (2011)

38. Padmanaban, N., Peng, Y., Wetzstein, G.: Holographic near-eye displays based on overlap-add stereograms. ACM Transactions on Graphics (TOG) **38**(6), 1–13 (2019)

39. Peng, Y., Choi, S., Padmanaban, N., Wetzstein, G.: Neural holography with camera-in-the-loop training. ACM Transactions on Graphics (TOG) **39**(6), 1–14 (2020)

40. Schechner, Y.Y., Kiryati, N.: Depth from defocus vs. stereo: How different really are they? International Journal of Computer Vision (IJCV) **39**, 141–162 (2000)

41. Schiffers, F., Chakravarthula, P., Matsuda, N., Kuo, G., Tseng, E., Lanman, D., Heide, F., Cossairt, O.: Stochastic light field holography. In: 2023 IEEE International Conference on Computational Photography (ICCP). pp. 1–12. IEEE (2023)

42. Sensing-Aiot: Apple vision pro perception analysis (Jun 2023), `https://4da.tech/?p=1138`

43. Sheinin, M., Schechner, Y.Y.: Depth from texture integration. In: 2019 IEEE International Conference on Computational Photography (ICCP). pp. 1–10. IEEE (2019)

44. Shi, L., Li, B., Kim, C., Kellnhofer, P., Matusik, W.: Towards real-time photorealistic 3d holography with deep neural networks. Nature **591**(7849), 234–239 (2021)

45. Shi, L., Ryu, D., Matusik, W.: Ergonomic-centric holography: Optimizing realism, immersion, and comfort for holographic display. arXiv preprint arXiv:2306.08138 (2023)

46. Shimobaba, T., Kakue, T., Ito, T.: Real-time and low speckle holographic projection. In: 2015 IEEE 13th International Conference on Industrial Informatics (INDIN). pp. 732–741. IEEE (2015)

47. Tilmon, B., Sun, Z., Koppal, S.J., Wu, Y., Evangelidis, G., Zahreddine, R., Krishnan, G., Ma, S., Wang, J.: Energy-efficient adaptive 3d sensing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5054–5063 (2023)

48. Tsuji, M., Kubo, H., Jayasuriya, S., Funatomi, T., Mukaigawa, Y.: Touch sensing for a projected screen using slope disparity gating. IEEE Access **9**, 106005–106013 (2021)

49. Ueda, T., Kubo, H., Jayasuriya, S., Funatomi, T., Mukaigawa, Y.: Slope disparity gating using a synchronized projector-camera system. In: 2019 IEEE International Conference on Computational Photography (ICCP). pp. 1–9. IEEE (2019)

50. Wang, J., Wang, J., Zhou, J., Zhang, Y., Wu, Y.: Crosstalk-free for multi-plane holographic display using double-constraint stochastic gradient descent. Optics Express **31**(19), 31142–31157 (2023)

51. Wang, J., Bartels, J., Whittaker, W., Sankaranarayanan, A.C., Narasimhan, S.G.: Programmable triangulation light curtains. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 19–34 (2018)

52. Wang, L., Xu, H., Tabata, S., Hu, Y., Watanabe, Y., Ishikawa, M.: High-speed focal tracking projection based on liquid lens. In: ACM SIGGRAPH 2020 Emerging Technologies, pp. 1–2 (2020)

53. Wang, Z., Chen, T., Chen, Q., Tu, K., Feng, Q., Lv, G., Wang, A., Ming, H.: Reducing crosstalk of a multi-plane holographic display by the time-multiplexing stochastic gradient descent. Optics Express **31**(5), 7413–7424 (2023)

54. Wu, Y., Boominathan, V., Zhao, X., Robinson, J.T., Kawasaki, H., Sankara-narayanan, A., Veeraraghavan, A.: Freecam3d: Snapshot structured light 3d with freely-moving cameras. In: 16th European Conference on Computer Vision (ECCV). pp. 309–325. Springer (2020)
55. Xu, H., Wang, L., Tabata, S., Watanabe, Y., Ishikawa, M.: 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)
56. Yu, P., Liu, Y., Wang, Z., Liang, J., Liu, X., Li, Y., Qiu, C., Gong, L.: Ultrahigh-density 3d holographic projection by scattering-assisted dynamic holography. Optica **10**(4), 481–490 (2023)
57. Zhang, J., Pégard, N., Zhong, J., Adesnik, H., Waller, L.: 3d computer-generated holography by non-convex optimization. Optica **4**(10), 1306–1313 (2017)