High Resolution Étendue Expansion for Holographic Displays

GRACE KUO, Facebook Reality Labs and University of California, Berkeley
LAURA WALLER, University of California, Berkeley
REN NG, University of California, Berkeley
ANDREW MAIMONE, Facebook Reality Labs

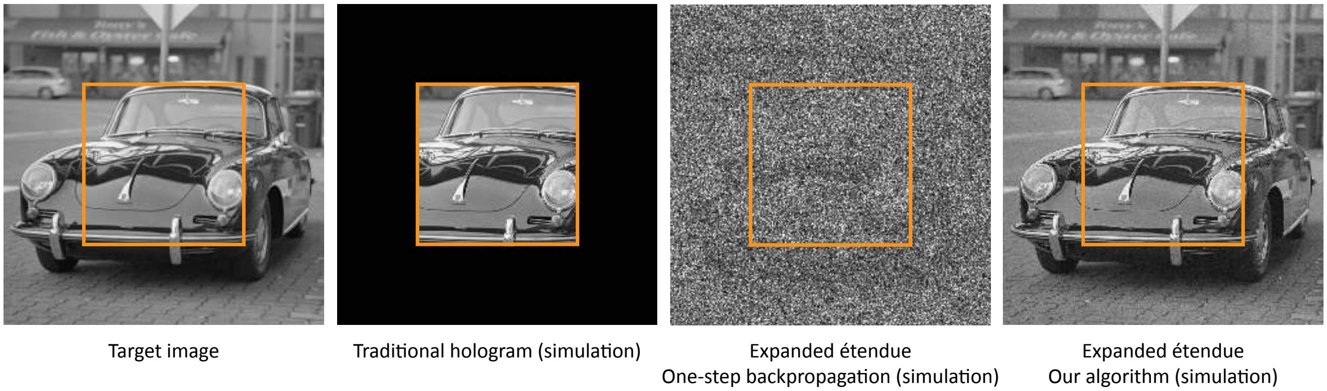
Target image
Traditional hologram (simulation)
Expanded étendue
One-step backpropagation (simulation)
Expanded étendue
Our algorithm (simulation)

Fig. 1. Traditional holographic displays have limited étendue resulting in a tradeoff between field-of-view (FoV) and eyebox size. If the eyebox is held constant, unique imagery cannot be displayed outside of the native FoV (orange box). The addition of a thin scattering mask into the system increases the diffraction angles, and thus the FoV, without sacrificing the eyebox. The scattering mask is taken into account during computation of the hologram through an iterative algorithm that outperforms the one-step backpropagation approach used in prior work. Car source image by Bill Newton (CC BY 2.0).

Holographic displays can create high quality 3D images while maintaining a small form factor suitable for head-mounted virtual and augmented reality systems. However, holographic displays have limited étendue based on the number of pixels in their spatial light modulators, creating a tradeoff between the eyebox size and the field-of-view. Scattering-based étendue expansion, in which coherent light is focused into an image after being scattered by a static mask, is a promising avenue to break this tradeoff. However, to date, this approach has been limited to very sparse content consisting of, for example, only tens of spots.

In this work, we introduce new algorithms to scattering-based étendue expansion that support dense, photorealistic imagery at the native resolution of the spatial light modulator, offering up to a 20 dB improvement in peak signal to noise ratio over baseline methods. We propose spatial and frequency constraints to optimize performance for human perception, and performance is characterized both through simulation and a preliminary benchtop prototype. We further demonstrate the ability to generate content at multiple depths, and we provide a path for the miniaturization of our benchtop prototype into a sunglasses-like form factor.

CCS Concepts: • **Hardware → Displays and imagers**.

Additional Key Words and Phrases: computer generated holography, computational displays, augmented reality, near-eye displays

ACM Reference Format:
Grace Kuo, Laura Waller, Ren Ng, and Andrew Maimone. 2020. High Resolution Étendue Expansion for Holographic Displays. *ACM Trans. Graph.* 39, 4, Article 66 (July 2020), 14 pages. https://doi.org/10.1145/3386569.3392414

1 INTRODUCTION

Computer generated holography allows generation of an arbitrary light distribution from a flat, programmable spatial light modulator (SLM) by controlling the wavefront of a coherent beam of light. This technique is particularly promising for near-eye displays since it enables per-pixel focus control, computational correction of optical aberrations, and simple optical components suitable for miniaturization [Maimone et al. 2017].

However, current holographic displays suffer from a unique challenge: a tradeoff between field-of-view (FoV) and the size of the viewing eyebox, the area in which the eye must be located to see the image. Together, these two quantities describe the *étendue* of the display, a quantity which measures the product of the area and solid angle of emitted light from a surface in an optical system. In conventional, non-holographic displays (e.g., the Oculus Rift S) obtaining large étendue is generally not a challenge and can be provided, for example, by a display panel backlight that has large area and range of emission angles. However, in a holographic display the étendue

Authors' addresses: Grace Kuo, Facebook Reality Labs, University of California, Berkeley, gkuo@berkeley.edu; Laura Waller, University of California, Berkeley, waller@berkeley.edu; Ren Ng, University of California, Berkeley, ren@berkeley.edu; Andrew Maimone, Facebook Reality Labs, andrew.maimone@fb.com.

0730-0301/2020/7-ART66
https://doi.org/10.1145/3386569.3392414

is determined by the number of degrees of freedom (*i.e.* pixels) on the SLM.

For an immersive display, one generally desires a large FoV of $\geq 90°$. With current modulators, a holographic display with such a FoV would only afford an eyebox of approximately 1 mm. Small eyeboxes cause the image to disappear if the eye deviates slightly from the design position, including deviations from eye rotation. The brute force solution is to increase the pixel count of the SLM; however, a solution providing a 10 mm eyebox would require approximately one billion pixels. This solution is two orders of magnitude away from current technology and is inefficient since the pixel count far exceeds what can be resolved with the human eye.

Another proposed solution is to expand the étendue by augmenting the display with a static scattering mask, which can increase the angles of light diffracted from the display. The known or inferred mask pattern is taken into account when computing the hologram so that a coherent image can still be formed after scattering. However, past efforts using the approach [Buckley et al. 2006; Park et al. 2019; Yu et al. 2017b] have been limited to very simple scenery, consisting, for example, of only tens of spots. To make this approach practical, it must scale several orders of magnitude to achieve the resolution expected of modern displays.

In this work, we present a new algorithmic approach to scattering-based étendue expansion that preserves the native, high resolution of modern spatial light modulators. After being scattered by a mask, the wavefront from a holographic display has many more degrees-of-freedom than one can control with the spatial light modulator, resulting in very high resolution output, but also extreme noise. Our key innovation is to constrain the holographic image to the number of spatial frequencies that can be controlled by the modulator, so that noise is pushed to higher frequencies than can be resolved by a human viewer. The process essentially decouples étendue and pixel count in holographic displays and results in high quality output with a small to moderate loss in contrast. Unlike prior work, we show that our method scales well to complex, full-resolution, photographic images. We also demonstrate that spatial constraints can be used to programmatically redistribute the image quality and resolution in a holographic image, for example, to increase fidelity in the area around the user's fovea. We present a mathematical framework for optimizing étendue expanded holograms with scattering masks, provide optical simulations and characterize performance, and provide preliminary experimental results on a benchtop prototype. We also discuss current limitations and describe a potential path for implementing our design in a sunglasses-like form factor.

1.1 Contributions and Limitations

We provide algorithms for generating high-quality étendue-expanded holograms and evaluate results in simulation and on a hardware prototype. Specifically, we make the following contributions:

- An algorithm for generating holographic images through a scattering mask based on constrained non-convex optimization that significantly outperforms prior state-of-the-art methods
- The addition of frequency and spatial constraints to significantly improve image quality
- The first demonstration, both in simulation and experiment, of dense, photorealistic holograms with higher étendue than the native SLM

Our approach also suffers from some limitations and challenges. Image contrast is reduced as the étendue of the display is increased beyond the native support of the SLM, limiting the practical range of étendue expansion. Computation time is greater than past methods [Buckley et al. 2006; Park et al. 2019; Yu et al. 2017b] as we rely on iterative optimization, rather than a one-step method. Additionally, our current hardware prototype operates in a single color channel only, although full color operation has been demonstrated in past related works [Park et al. 2019]. As with other prior work featuring scattering masks with small features, our approach is sensitive to alignment and has not yet been demonstrated in a compact form factor suitable for the proposed virtual and augmented reality applications. We address these challenges in Section 5.1.

2 RELATED WORK

Holographic Displays. Holographic displays have shown promising results for virtual and augmented reality in a series of recent papers [Li et al. 2016; Maimone et al. 2017; Shi et al. 2017; Wakunami et al. 2016; Yaraş et al. 2009]. To highlight a few, Maimone et al. [2017] demonstrated a holographic display for augmented reality with wide FoV and sunglasses-like form factor, and demonstrated high quality, full-color holograms with real-time computation in a benchtop form factor. Shi et al. [2017] demonstrated the rendering of light field data as holograms to capture view-dependent effects. However, these systems were constrained by the low étendue of current SLMs, limiting either the FoV or eyebox of the displays.

More recently, several works have proposed methods for more effective use of the étendue of a holographic display by tracking the viewers' eyes and dynamically moving around a small eye box, also known as pupil steering. Jang et al. [2018] show pupil steering by changing the angle of light incident on an SLM with a mechanical mirror and an arrayed hologram. Kim et al. [2018] create several copies of the hologram and used a reflective display to control which copy is shown. Choi et al. [2020] also create copies of the hologram but effectively control which copy is used computationally. While showing promise, these pupil steered methods require precise and low-latency eye tracking, have complex and difficult to miniaturize optics, and have lower performance than non-pupil steered holographic displays. Our approach does not demand eye tracking and requires only the addition of a scattering mask in the optical system.

Focusing through Scattering Media. Our approach builds on prior work using wavefront shaping to focus light through an unknown scattering element. This concept was first described by Vellekoop and Mosk [2007] who formed a focal spot on the far side of a scattering material by optimizing the phase of a deformable mirror via a feedback loop. Popoff et al. [2010] describes a more efficient calibration technique in which the scattering material's transmission matrix is pre-characterized using a wavefront modulator, enabling computational creation of focal spots at any location without recalibrating; variations on this approach are prevalent in the literature [Conkey et al. 2012; Tao et al. 2015; Yoon et al. 2015]. Focusing through scattering has also been demonstrated with binary amplitude modulation [Akbulut et al. 2011; Yu et al. 2017a] and has been successfully employed for imaging through translucent materials [Chaigne et al. 2014; Conkey et al. 2015]. However, these works all assume that the scatterer is an unknown and undesirable obstacle.

In contrast, our proposed approach exploits properties of the scattering element, namely that it can diffract light to higher angles than natively supported by the spatial light modulator. This property was first used by Vellekoop et al. [2010] to generate sharper foci than achievable without the scattering media, and Yeh et al. [2019] used the concept for optical superresolution in imaging. Holographic displays that take advantage of increased angle from a scattering element have also been proposed, as we discuss below.

Étendue Expansion for Displays. Perhaps most conceptually similar to our work are prior holographic displays that use a diffractive mask in front of the SLM for the purpose of increasing étendue. Buckley et al. [2006] describe using a diffractive phase mask in front of a binary phase modulator to remove the twin image and simultaneously increase the viewing angle of the display. While a compelling idea, their experimental results are limited to a static prototype of a single very simple and sparse scene (a few letters of text) and temporal averaging is required to produce visually-pleasing results.

Yu et al. [2017b] created the first dynamic display with expanded étendue. Using an off-the-shelf diffuser as the scattering element, they demonstrated a large increase in étendue but could only create up to 15 foci simultaneously and required an intensive calibration that scaled with the number of pixels on the SLM, limiting use to low resolution modulators. Park et al. [2019] improved on the idea by replacing the unknown diffuser with a known diffractive amplitude mask or “photon sieve”, thus eliminating the calibration step and enabling use of higher resolution SLMs. The resulting display could also generate a large increase in étendue, but could only generate up to 75 focal spots simultaneously. In contrast to these approaches, our proposed method scales to dense, photorealistic holograms at the native resolution of the SLM. We choose a thin transparent mask for our scattering element, which has better light efficiency and a smaller DC term compared to the “photon sieve”, and we introduce spatial and frequency-based weighting to étendue expanded holograms.

Algorithms for Computer Generated Holography. A key component in any holographic display is the algorithm used to determine the pattern to display on the SLM. This is a particular challenge if the SLM affords phase-only control, which we assume in most of this work. To generate phase-only holograms, one option is to simply discard the amplitude, but iterative approaches, such as the popular algorithm by Gerchberg and Saxton [1972], increase image fidelity by allowing the phase at the image plane to vary. Georgiou et al. [2008] augmented this algorithm with “don’t care” regions which improve image quality at the expense of dedicating a high-noise region outside of the active part of the field of view. However, both these algorithms do not have explicitly defined cost functions, making it challenging to tune parameters to specific applications. In contrast, Zhang et al. [2017] explicitly define the problem in an optimization framework, allowing custom, application-specific loss functions, and they demonstrate improved results for optogenetic stimulation. Chakravarthula et al. [2019] use a similar framework and target their work towards displays, demonstrating high quality experimental results on color images. Similar to this prior work, our algorithm is based on explicitly solving an optimization problem, but we extend the approach to compensate for the scattering mask and introduce new loss functions based on perceptual metrics.

3 METHODS

3.1 Étendue of holographic displays

The étendue of a display is defined as the product of the display area with the solid angle of emitted light,

$$G = 4A\sin^2\theta, \tag{1}$$

where G is the étendue of a planar display with area A emitting light confined to a square pyramid of width 2θ around the display’s normal. Étendue is conserved through reflections, refraction, and free space propagation [Chaves 2017].

In a non-holographic display (e.g. a backlit LCD panel, OLED panel, etc.), each pixel on the display emits light over a large cone of angles, so the étendue is usually quite large and does not present major limitations to the optical design. However, in a holographic display, the maximum deflection angle of the light, θ, is determined by the pixel size, Δ,

$$\sin\theta = \frac{\lambda}{2\Delta}, \tag{2}$$

where λ is the wavelength of light. Substituting this equation into Eq. 1 yields the étendue of a holographic display,

$$G = \frac{\lambda^2 A}{\Delta^2} = \lambda N_x \times \lambda N_y, \tag{3}$$

where N_x and N_y are the number of pixels along each dimension of the SLM. Therefore, the étendue of a traditional holographic display is proportional to the total number of pixels.

3.2 Field-of-View and Eyebox Tradeoff

We will now consider why large étendue is desirable and how much is needed. For simplicity, we’ll center the discussion around “Fourier holography” in which a virtual image or volume is produced by the SLM at the Fourier (pupil) plane of a lens, assumed to be ideal with focal length f_1. (The same conclusions holds in the regime of “Fresnel holography” in which the virtual image is created directly in front of the SLM without additional optics.) In a near-eye display, there is typically an additional lens (focal length f_2) that projects the virtual image to optical infinity before the light enters the eye. Figure 2a shows a schematic of this scenario.

Consider an SLM with width w and maximum diffraction angle $\pm\theta$. Based on geometry, this results in a one-dimensional (1D) FoV and eyebox size given by:

$$\text{FoV} = 2\tan^{-1}\left(\frac{f_1}{f_2}\tan\theta\right) \approx 2\frac{f_1}{f_2}\theta \tag{4}$$

$$\text{eyebox} = \frac{f_2}{f_1}w. \tag{5}$$

Therefore, we can easily trade off between the FoV and eyebox size for a given SLM by choosing the ratio f_1/f_2. However, the FoV-eyebox product is fixed. To illustrate this, we apply the small angle

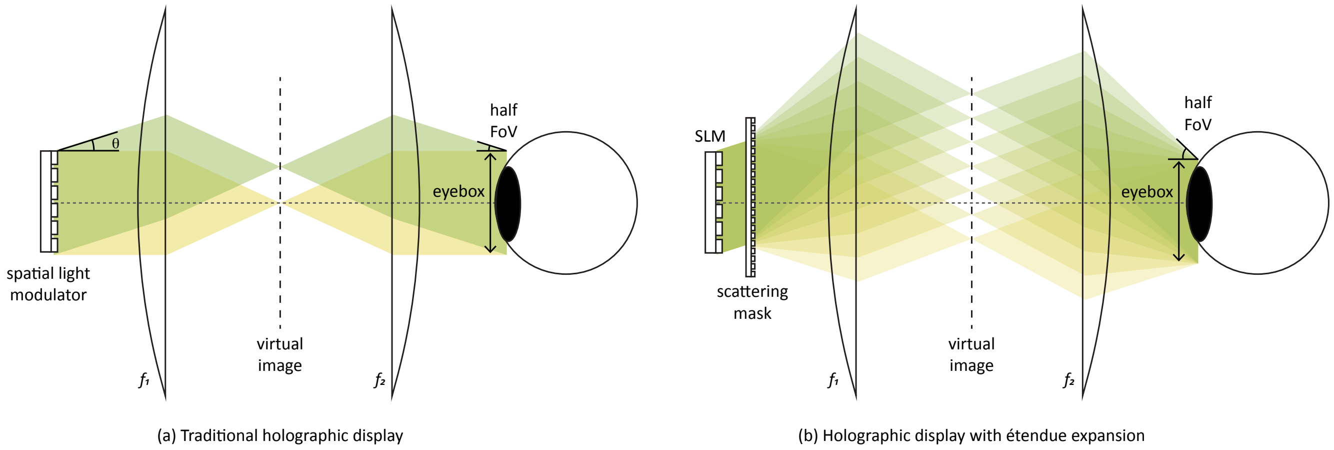

Fig. 2. In a traditional holographic display (a), the diffraction angle of pixels on the SLM determines the nominal FoV and the extent of the SLM determines the eyebox. These quantities can be exchanged by modifying the ratio of f_1/f_2, but the product is fixed and determined by the number of pixels on the SLM. (b) To overcome this trade-off, a scattering mask is placed in front the SLM. The wavefront coming off the SLM is scattered by the mask to a larger range of angles, thus increasing the FoV without decreasing the eyebox size.

approximation ($\sin\theta \approx \tan\theta \approx \theta$), and plug Eq. 2 into Eq. 4 to give the following relationship:

$$\text{FoV} \times \text{eyebox} \approx 2\theta w \approx \frac{\lambda w}{\Delta} = \lambda N_x. \tag{6}$$

Thus, the product of 1D FoV and eyebox is equal to the 1D étendue, and the same relationship extends to 2D as well.

To give an idea of the achievable design space, the highest pixel count commercial SLM today (Holoeye GAEA-2) has 4160 × 2464 pixels. To provide an immersive experience for the user with a horizontal FoV of 120° at wavelength $\lambda = 532$ nm, the eyebox size is only 1.05 mm. Therefore, even a small rotation of the eye will cause the pupil to leave the display's viewing eyebox, and the user will no longer be able to see the image. If we instead prioritize an eyebox of 10 mm, large enough to accommodate reasonable eye movement, the resulting FOV is only 12.7°. Simultaneously achieving both the 120° FoV and 10 mm eyebox would require $N_x \approx 32,500$ pixels, resulting in a total of over one billion pixels needed over the SLM area.

The brute force solution is to develop SLMs with more pixels, a pathway that is actively being researched. Most current phase SLMs use liquid crystal on silicon (LCoS) technology, and the highest resolution LCoS device known has 8192 × 4320 pixels [Sterling 2008]. However, this still leaves a gap of more than an order of magnitude between the achievable étendue of current technology and that required for the ideal display described above. Continuing to increase pixel count will encounter challenges in display bandwidth, power usage, pixel cross-talk [Moser et al. 2019], and device size.

In addition, a billion-pixel SLM is inefficient since it generates much higher resolution images than can be resolved with the human eye. An SLM with a horizontal pixel count of $N_x = 32,500$ and 120° FoV, as described above, would create visual stimuli with 270 pixels/degree angular resolution, far beyond the 60 pixels/degree limit of normal 6/6 or 20/20 vision. Furthermore, we know that humans have higher visual acuity in the center of the retina, a observation that has been exploited for improved computational efficiency through foveated rendering [Guenter et al. 2012; Kaplanyan et al. 2019; Patney et al. 2016]. As a result of both foveation and the limited resolution of the eye, the total number of degrees of freedom actually perceived by the user is far smaller than the number of SLM pixels needed for étendue purposes. What we truly desire is to decouple the étendue, and therefore the FoV and eyebox, from the number of SLM pixels and the display resolution. In the next section, we describe a strategy to achieve this by placing a static scattering mask in front of the SLM and computationally generating holograms that account for the limited resolution of human vision.

3.3 Scattering-Based Étendue Expansion

In a traditional holographic display, shown in Figure 2a, the image perceived by the viewer is the same as the intensity distribution at the virtual image plane, $I(\vec{x})$, which is described by

$$I(\vec{x}) = |y(\vec{x})|^2 = |\mathcal{F}\{s(\vec{u})\}|^2. \tag{7}$$

Here, $s(\cdot)$ is the complex field at the SLM, $y(\cdot)$ is the complex field at the image plane, $\mathcal{F}\{\cdot\}$ denotes 2D Fourier transform, and $\vec{x}$ and $\vec{u}$ are the coordinates at the image plane and SLM plane, respectively, which are related by $\vec{x} = \vec{u}/\lambda f_1$ [Goodman 2005]. Since the SLM pixel size, Δ, determines the maximum frequency displayable on the SLM, the virtual image has finite controllable extent corresponding to $\vec{x} \in [-\frac{1}{2\Delta}, \frac{1}{2\Delta}]^2$. Due to the discrete nature of the SLM pixels, the SLM also generates higher-order terms which manifest as replicas of diminishing intensity. However, these replicas cannot be controlled independently of the central region so they are not included in the FoV calculation; in fact, many holographic display systems physically filter them out.

To expand the étendue of the holographic display, we place a static scattering mask in front of the SLM, shown in Figure 2b. Unlike past systems that focused light through unknown media (e.g. biological tissue), here we can choose the scattering mask to have desirable properties, as discussed in Sec. 3.3.1. We specifically use a thin transparent mask with a known phase profile, $\alpha(\vec{u})$, resulting in a corresponding complex modulation function $m(\vec{u}) = \exp(j\alpha(\vec{u}))$. Since the mask is thin by design, with an optical path length deviation of at most one wavelength, we assume that the mask only affects the electric field at one plane, and using a relay system we set that plane to be directly conjugate to the SLM. Therefore, with the addition of the scattering mask, the intensity at the image plane is

$$I(\vec{x}) = |\mathcal{F}\{s(\vec{u})m(\vec{u})\}|^2, \tag{8}$$

which can also be written as a convolution between the far-field patterns of the SLM and the mask:

$$I(\vec{x}) = |\mathcal{F}\{s(\vec{u})\} * \mathcal{F}\{m(\vec{u})\}|^2. \tag{9}$$

Here, $*$ denotes a 2D convolution.

Assuming the mask is defined on a discrete grid with pixel size Δ_m, the far field pattern of the mask, $\mathcal{F}\{m(\vec{u})\}$, has unique content over the extent $\vec{x} \in [-\frac{1}{2\Delta_m}, \frac{1}{2\Delta_m}]^2$. As with the SLM, higher-order terms due to the discrete nature of the mask create replicas of the central region, allowing us to think of the convolution in Eq. 9 as having circular boundary conditions. Therefore, the total extent of $I(\vec{x})$, which directly corresponds to the FoV, is equal to the extent of the far-field pattern of the mask itself. Using standard micro/nano fabrication techniques, it is straightforward to create a mask with smaller pixels than those on the SLM, thus increasing the display's FoV.

Importantly, adding the scattering mask into the system in this way does not change the total SLM size, and therefore the eyebox is unchanged. This means that the increase in FoV described above corresponds directly to an increase in étendue by a factor, q, determined by the ratio of the SLM and mask pixel sizes,

$$q = \left(\frac{\Delta}{\Delta_m}\right)^2. \tag{10}$$

Throughout this paper, we'll visualize the increase in étendue as a FoV expansion, but the total expanded étendue can easily be redistributed between FoV and eyebox by choosing appropriate focal length lenses, as described in Sec. 3.2.

3.3.1 *Mask Design.*

One important parameter of the mask design is the pixel size, Δ_m, which determines the étendue expansion factor (Eq. 10). In addition, if we assume no prior knowledge about the content to be displayed, Eq. 9 suggests that we want a uniform far-field mask pattern, $\mathcal{F}\{m(\vec{u})\}$, so that the intensity distribution, $I(\vec{x})$, is not biased towards any specific location in the FoV. To achieve this, we choose a random binary phase profile, with phase $\alpha(\vec{u})$ either 0 or π (equivalent to $m(\vec{u}) = \pm 1$) at each mask pixel. This design is a type of white noise and thus yields a flat far-field pattern. Although we chose a binary pattern for ease of fabrication, other mask designs with uniform far-field patterns will generate similar results. However, an amplitude-only mask, such as the "photon sieve" used by Park et al. [2019], will always have a strong zero-order (DC) term due to lack of "negative" values in the modulation function, which results in a uneven far-field distribution. The DC term diminishes if more of the mask is opaque, but this greatly reduces light efficiency compared to our transparent phase mask.

A practical factor in the mask design is our ability to computationally model the effect of the mask on the electric field. Theoretically, a multiple-scattering element with uniform far-field intensity could replace the thin mask in our system. However, calibration of multiple-scattering effects is intensive and, for fixed q, the number of parameters in the model scales quadratically with the number of pixels on the SLM [Popoff et al. 2010]. In contrast, our choice of a thin flat scattering mask enables efficient modeling by a single point-wise multiplication (Eq. 8), and the number of parameters describing the mask scales linearly with the number of SLM pixels for fixed q, allowing the model to be efficiently used in an iterative computational framework.

Our final practical consideration in the mask design is also related to computational tractability: to compute the Fourier transform of Eq. 8 with an FFT, it is necessary that both the SLM and mask be represented digitally on a uniform grid. Therefore, a mask with a pixel structure is convenient as it is easily represented in a discrete form. A non-pixelated mask with smoothly varying features could also be represented on a uniform grid but to accurately capture all features, the mask must be over-sampled, increasing compute time.

Although our mask choice is justified based on the reasoning above, improvements to the design may be revealed through end-to-end optimization [Sitzmann et al. 2018], which is a topic for future work.

3.4 Image Calculation Algorithm

Simply adding our thin scattering mask in front of an SLM is not sufficient to make a display. In fact, the mask would scramble the wavefront such that the viewer only sees a speckle field. Therefore, to generate an image after the mask, we must pre-compensate for the scattering in the SLM pattern. Furthermore, the number of degrees-of-freedom in the output image, $I(\vec{x})$ is higher than the number of controllable modes on the SLM by a factor of q, so it is impossible to generate arbitrary images. This can be thought of as an over-determined data fitting problem, in which we can only attempt to create the closest possible image (by some metric), but there will always be uncontrollable noise creating deviations between the target image and output image. We propose an optimization-based algorithm to reduce unwanted noise by allowing the algorithm to control phase at the image plane, and we further direct residual noise into perceptually less noticeable regions or spatial frequencies through custom loss functions based on human vision. We begin by describing the algorithm used in prior work, which we call one-step backpropagation, then introduce our improved version.

3.4.1 *One-step Backpropagation.*

Despite physical differences in their systems, prior state-of-the-art [Buckley et al. 2006; Park et al. 2019; Popoff et al. 2010; Yu et al. 2017b] all use the same underlying algorithm for computing the SLM pattern to generate target images through the scattering element.

Due to the reciprocity of light, a given target electric field, $\hat{y}(\vec{x})$, can be propagated backwards to the calculate the electric field at the SLM plane. For example, in our model (Eq. 8), the electric field at the SLM plane is

$$\hat{s}(\vec{u}) = \mathcal{F}^{-1}\{\hat{y}(\vec{x})\}m^*(\vec{u}), \tag{11}$$

where $m^*(\vec{u})$ is the complex conjugate of the mask function and $\mathcal{F}^{-1}\{\cdot\}$ is the inverse 2D Fourier transform. If this electric field were displayed on the SLM, it would exactly recreate the target electric field after the mask. However, this is impossible for two reasons. First, the SLM is phase-only, meaning that it can only display the phase of the complex field. Second, the resolution of the SLM pixels is limited and only one phase value can be displayed per pixel. To generate a valid SLM pattern, the one-step backpropagation method integrates the field over each SLM pixel, then throws away the amplitude to get a phase-only value:

$$p_i = \arg\left(\int_{\Delta_i} \mathcal{F}^{-1}\left\{\hat{y}(\vec{x})\right\} m^*(\vec{u})\, d\vec{u}\right), \tag{12}$$

where p_i is the phase value at pixel i on the SLM, the arg$(\cdot)$ operator takes the phase of the complex field, and the integral is over the area corresponding to the i-th pixel.

On the surface this appears to be an optimal approach without much room for improvement. However, this model optimizes for a target *electric field*, despite the fact that humans can only detect the intensity of light. To create a target intensity, $\hat{I}(\vec{x})$, with this approach, first an arbitrary phase $\phi(\vec{x})$ is assigned such that

$$\hat{y}(\vec{x}) = \sqrt{\hat{I}(\vec{x})}\,\exp\left(j\phi(\vec{x})\right), \tag{13}$$

and after the phase is assigned, it stays fixed when calculating the SLM pattern from Eq. 12. In contrast, in our approach we let the phase at the image plane be a free variable, since it is not detectable by the eye, which greatly improves image quality at the expense of increased computation. This idea has been previously applied to computer generated holography without scattering masks [Chakravarthula et al. 2019; Gerchberg and Saxton 1972; Zhang et al. 2017]. In addition, we introduce a flexible framework that allows us to incorporate different loss functions that can be tailored to the specific application of near-eye displays.

3.4.2 *Our Approach.*

To calculate the SLM pattern, we solve the following optimization problem to find the phase values $\vec{p}$ to be displayed on the SLM.

$$\vec{p} = \arg\left(\underset{\vec{s}}{\operatorname{argmin}}\,\mathcal{L}\left(I(\vec{x}), \hat{I}(\vec{x})\right)\right) \quad \text{subject to } |\vec{s}| = \vec{1}, \tag{14}$$

where $\vec{s}$ is the discrete complex field at the SLM pixels, the arg$(\cdot)$ operator takes the phase of the complex field, $|\cdot|$ denotes element-wise magnitude, $\hat{I}(\vec{x})$ is the target intensity at the image plane, and $I(\vec{x})$ is the output intensity calculated using Eq. 8. Finally, $\mathcal{L}(\cdot,\cdot)$ is a differentiable custom loss function that outputs a single real-valued similarity metric between the output and target intensities.

To take the discrete nature of the SLM into account, we solve directly for a single phase value at each pixel. Before computing the output intensity using Eq. 8, we generate a continuous representation from the discrete field, $\vec{s}$, as follows

$$s(\vec{u}) = \sum_i s_i\, R(\vec{u} - \vec{u}_i) \qquad R(\vec{u}) = \begin{cases} 1 & \vec{u} \in [-\Delta/2, \Delta/2]^2 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

where s_i is the complex field at the i-th pixel and $\vec{u}_i$ is the pixel's location on the SLM. Note that, although the equations here are presented as continuous, they must be discretized to solve Eq 14 digitally; for practical purposes Eq. 15 corresponds to upsampling the SLM pattern with a box filter. Since the mask has smaller pixels than the SLM, the necessary grid resolution to describe the system is always finer than the SLM pixel size, making this upsampling a critical step. See the Appendix for more details.

In addition to specifically solving for a phase-only discrete SLM pattern, we also improve on one-step backpropagation with our custom loss function, which acts only on the image intensities. Unlike the prior algorithm, the phase of the target image is never defined, and the phase of the output image can vary freely. If we simply want the output image to match the target as closely as possible, we can use the following sum of squares loss function:

$$\mathcal{L}\left(I(\vec{x}), \hat{I}(\vec{x})\right) = \frac{1}{2}\sum_{\vec{x}}\left(I(\vec{x}) - \hat{I}(\vec{x})\right)^2. \tag{16}$$

However, since the number of degrees-of-freedom on the SLM is significantly smaller than the number of free variables in the output image, we cannot perfectly match the target and output. Next, we introduce two different loss functions that prioritize features of the image that are more perceptually relevant. Other loss functions (for example, based on salient features in the image) could be readily incorporated in this optimization framework.

Frequency Constraints. The resolution of the human eye is limited to about 60 pixels/degree for normal vision. As described in Section 3.2, the achievable resolution of a large étendue holographic display can be several times higher than human perception. Therefore, we can improve the output by constraining the loss function to only penalize the lower spatial frequencies resolvable by the eye:

$$\mathcal{L}\left(I(\vec{x}), \hat{I}(\vec{x})\right) = \frac{1}{2}\sum_{\vec{u}}\left(c_f(\vec{u})\,\mathcal{F}\left\{I(\vec{x}) - \hat{I}(\vec{x})\right\}\right)^2, \tag{17}$$

where $c_f(\vec{u})$ is a low-pass filter. Setting the cutoff frequency of the low-pass filter to match human vision depends on the total FoV of the system (determined by the focal lengths of the relay optics, described in Sec. 3.2). To abstract away the choice of how to distribute eyebox and FoV, we set the cutoff frequency to correspond to the resolution achievable by the native SLM. Conveniently, this means that the resolution of the display remains constant as the étendue expansion factor increases, as long as the native SLM resolution is at least that of the human eye. To minimize ringing artifacts, we choose a fifth order Butterworth filter

$$c_f(\vec{u}) = \left(1 + \left(\frac{||\vec{u}||^2}{r_0^2}\right)^5\right)^{-1} \tag{18}$$

where $||\cdot||^2$ denotes the squared magnitude, and the cutoff frequency is $r_0 = \Delta_m\sqrt{N_x \times N_y/\pi}$ such that the number of controlled frequencies in the filter's passband matches the number of degrees-of-freedom on the SLM.

Spatial Constraints. It is well known that humans have foveated vision with the most visual acuity at the center of our gaze direction. This can be used to improve image quality by non-uniformly weighting the image loss based on the mostly likely regions viewed by the user, the most important content in the image, or the center of gaze as determined from eye tracking data. We combine this idea with the frequency constraints described above through the following loss function

$$\mathcal{L}\left(I(\vec{x}), \hat{I}(\vec{x})\right) = \frac{1}{2}\sum_{\vec{x}}\left(c_s(\vec{x})\mathcal{F}^{-1}\left\{c_f(\vec{u}')\,\mathcal{F}\left\{I(\vec{x}) - \hat{I}(\vec{x})\right\}\right\}\right)^2, \tag{19}$$

where $c_s(\vec{x})$ is a grayscale weight map describing the spatial importance of different regions of the image. This is conceptually similar to the work of Georgiou et al. [2008], in which noise is moved into "don't care" regions outside of the active FoV. Here we use spatial constraints to control the importance of imagery seen by the user and also provide non-binary weighting.

3.5 Extension to Multiple Focal Planes

One of the key advantages of holographic displays is the ability to display 3D content, which improves the realism of the display while helping alleviate visual fatigue from the vergence-accommodation conflict that plagues stereoscopic displays. In this section, we extend our image calculation algorithm to create content at multiple focal planes simultaneously. The intensity at distance z from the focal plane of the lens is

$$I_z(\vec{x}) = \left|\mathcal{P}_z\mathcal{F}\left\{s(\vec{u})m(\vec{u})\right\}\right|^2, \tag{20}$$

where $\mathcal{P}_z$ is the Frensel propagation operator which can be defined in Fourier space as $\mathcal{P}_z\{\cdot\} = \mathcal{F}^{-1}\{h_z(\vec{u})\mathcal{F}\{\cdot\}\}$ where

$$h_z(\vec{u}) = \exp\left(2\pi jz/\lambda\right)\exp\left(j\pi\lambda z||\vec{u}||^2\right), \tag{21}$$

and $||\cdot||^2$ is magnitude squared. By noting that $h_z(\vec{u}) = h_z(-\vec{u})$, we can efficiently calculate $I_z(\vec{x})$ as follows

$$I_z(\vec{x}) = \left|\mathcal{F}\left\{s(\vec{u})m(\vec{u})h_z(\vec{u})\right\}\right|^2. \tag{22}$$

As done by Zhang et al. [2017], we simultaneously optimize the intensity at all focal planes of interest by solving

$$\vec{p} = \arg\left(\underset{\vec{s}}{\operatorname{argmin}}\sum_{z}\mathcal{L}\left(I_z(\vec{x}), \hat{I}_z(\vec{x})\right)\right) \quad \text{subject to } |\vec{s}| = \vec{1}, \tag{23}$$

where $\hat{I}_z(\vec{x})$ is the target intensity at the focal plane at distance z, the summation is over the discrete number of z planes of interest, and the loss function can be set to any of those described for a single plane. Extensions to loss functions that account for interactions between planes are also possible but not explored in this work. As with the single plane version, $\vec{s}$ is converted to $s(\vec{u})$ via Eq. 15.

4 RESULTS

4.1 Simulation

We test our étendue expansion concept in a simulation implemented in MATLAB running on an Nvidia GeForce GTX 1060 GPU. We assume an SLM with 16 μm pixels and resolution 960 × 540. Our scattering mask is modeled as a thin phase element with binary phase (either 0 or π) randomly assigned to each pixel. The mask is the same physical size as the SLM and the pixel size of the mask, Δ_m, determines the étendue expansion factor, q, based on Eq. 10 with Δ_m = 8 μm, 4 μm, or 2.66 μm for 4×, 16×, and 36× étendue expansion, respectively. For a given target image, we solve Eq. 14 using projected gradient descent with Nesterov acceleration [Beck and Teboulle 2009]; the details of the algorithm are summarized in the Appendix. We use Eq. 8 to simulate the intensity at the image plane from the SLM pattern, then apply the low-pass filter described in Eq. 18 to simulate the perceptual effect of limited retinal resolution.

Figure 3 shows a simulation comparison of the image formation algorithms presented in Section 3.4. We implemented the one-step backpropagation algorithm used in prior work (Sec. 3.4.1) by solving Eq. 12 after generating a target electric field based on Eq. 13 where the phase, $\phi(\vec{x})$, is random from a uniform distribution, which we found yielded superior results compared to a constant phase. By not allowing optimization of the phase or additional constraints, one-step backpropagation results in low contrast images, even for only a 4× expansion factor, and contrast degrades rapidly as the expansion factor increases. We conclude that one-step backpropagation is better suited for the ultra-sparse scenes demonstrated in prior work [Park et al. 2019; Yu et al. 2017b] than for dense high resolution images.

We improve on prior state-of-the-art with our iterative optimization approach in which phase at the image plane is a free variable, and even without additional perceptual constraints, the improvement from such optimization is clearly apparent at all expansion factors. Adding frequency constraints, such that only low frequencies are optimized, further improves performance, yielding high quality, good contrast images for 4× expansion. However, these results begin to lose more contrast at higher expansion factors. Our spatial constraints can restore some of the contrast by prioritizing image fidelity in a subset of the FoV, thus improving contrast in this region at the expense of image quality in the periphery. Importantly, moving the spatial constraints to different regions of the FoV is easily accomplished algorithmically by changing the spatial weights, $c_s(\vec{x})$. The spatial constraints work equally well over any position in the FoV, and shifting position does not require physically moving components. An example of dynamically changing the spatial constraints is shown in the supplemental video. As discussed in Section 5, the spatial weights can be used statically or dynamically in conjunction with eye tracking.

A quantitative comparison of the methods is shown in Table 1. We calculate two metrics, peak signal to noise ratio (PSNR) and structural similarity (SSIM) [Hore and Ziou 2010]. When simulating images with spatial constraints, we apply a spatial map with a central region of value 1 that smoothly transitions into a peripheral region with value 0.1, shown in the inset of Figure 3. These regions approximately correspond to the viewing zones used by Patney

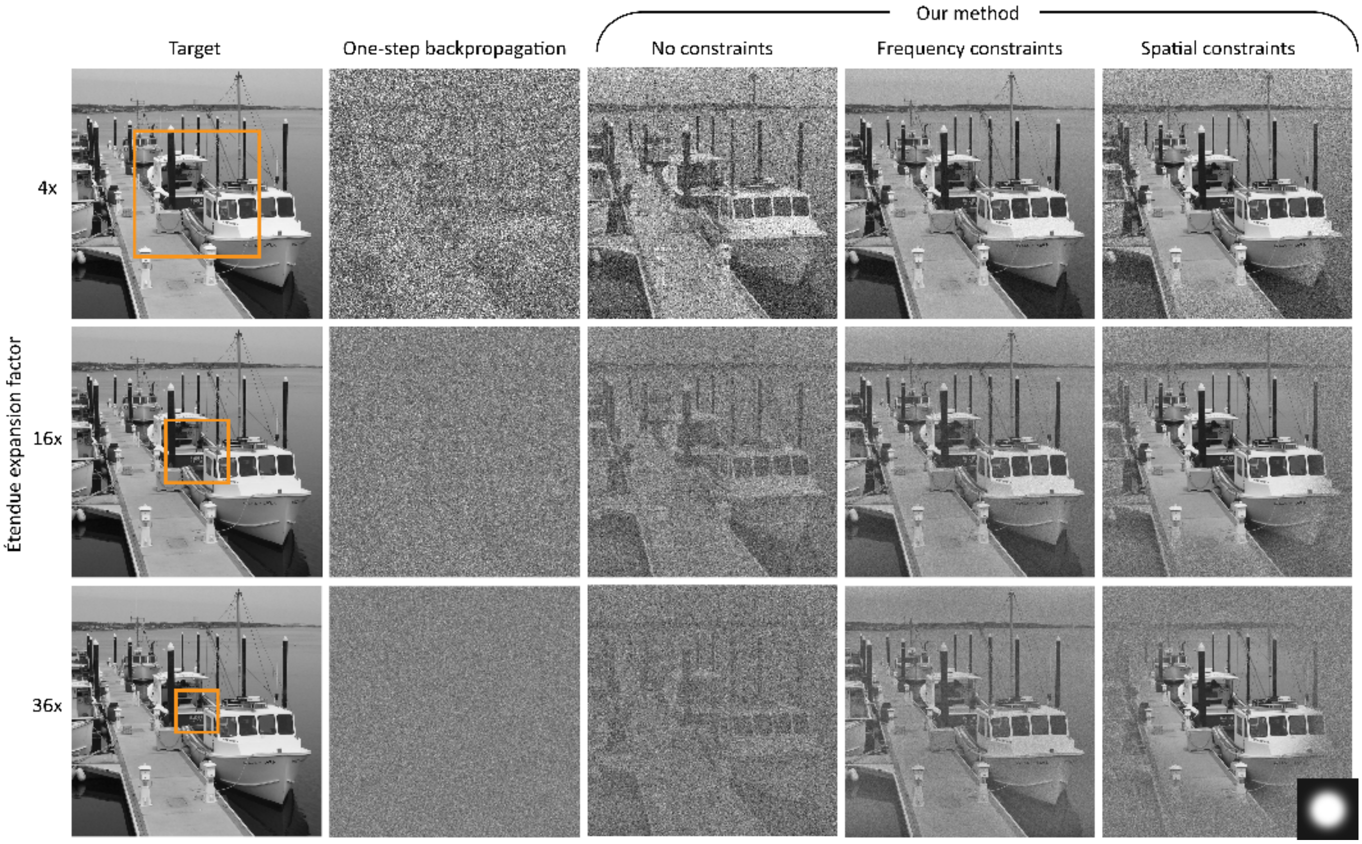

Fig. 3. Simulations comparing image formation algorithms for 4×, 16×, and 36× étendue expansion. The orange box indicates the FoV addressable by the native SLM without the scattering mask. With one-step backpropagation, the phase at the image plane is fixed and no perceptual constraints can be added, resulting in low contrast results that scale poorly to higher expansion factors. By allowing phase at the image plane be a free variable (i.e. "no constraints"), image quality is improved but still shows noisy results. With the addition of frequency constraints that prioritize the range of frequencies detectable by the visual system, noise and contrast are further improved. However, contrast degrades as the expansion factor increases. This can be mitigated by applying additional spatial constraints via a spatial weighting map (lower right inset). The spatial map used here is designed to approximately correspond to those used by foveated renderers [Patney et al. 2016], assuming our total FoV is set to 80°. Adding the spatial weights improves contrast and noise performance in the prioritized central region at the expense of the periphery, and the region of interest can easily be moved based on the viewer's gaze direction (see supplemental video for an example). All images are low-pass filtered to simulate the limited resolution of the visual system. Quantitative metrics are found in Table 1. Boat source image by Erick Bee (CC BY-SA 2.0).

et al. [2016] for foveated rendering if the FoV of our display is set to 80° (although there is no direct mapping between the values in our spatial map and the sampling factors used by Patney et al. [2016]). To fairly compare the images with spatial constraints, metrics are calculated over both the whole FoV and over a subregion (the central quarter of the image) corresponding to the area prioritized by the spatial constraints. Quantitatively, our method with frequency constraints performs best when considering the whole FoV, and our method with spatial constraints performs best when considering only a subregion. As visible in Figure 3, the main difference between the target and output images is due to contrast reduction.

The frequency constraints are critical to achieving low-noise images. In Figure 4 we compare the unconstrained loss function (Eq. 16) and the loss function with frequency constraints (Eq. 17) *before* applying the low-pass filter that approximates the visual system. Without constraints, the raw unfiltered output is more faithful to the target image, as evidenced by its higher PSNR. With frequency constraints, the unconstrained noise is moved into higher frequencies that are filtered out. The filtered result is visually less noisy and higher contrast. In this example, the frequency constraints improve PSNR by over 10 dB in the filtered result.

4.2 Experimental Prototype

We validate our simulations with a benchtop prototype built with a 1080p phase only LCoS SLM (Holoeye PLUTO-2). To reduce sensitivity to alignement errors, we bin the SLM pixels 2×2 resulting in an effective resolution of 960×540 with $16\,\mu$m pixels. For our scattering mask, we use a binary phase mask with a random pattern of $4\,\mu$m pixels with phase values of either 0 or π, resulting in an étendue expansion of $q = 16\times$. The mask was fabricated with

Table 1. Numerical comparison of image formation algorithms, averaged over 40 different natural images, where higher values indicate more similarity to the target. To quantify the effects of the spatial constraints, metrics were calculated over both the whole FoV (top) and over a cropped region one quarter the size of the FoV corresponding to the area prioritized by the spatial constraints (bottom). All results and target images are filtered according to Eq. 18 before calculating metrics, and images are normalized to have the same mean. The best performing algorithm in each category is in bold; when considering the whole FoV, our method with frequency constraints performs best, but can be improved in a subregion of the FoV by applying spatial constraints.

		4× expansion		16× expansion		36× expansion	
		PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM
Full FoV	One-step backpropagation	12.059	0.064	12.720	0.197	12.891	0.389
	Ours: No constraints	18.510	0.286	15.128	0.363	14.217	0.509
	Ours: Frequency constraints	**26.958**	**0.624**	**18.486**	**0.552**	**16.167**	**0.633**
	Ours: Spatial constraints	19.267	0.472	15.341	0.455	14.371	0.541
Partial FoV	One-step backpropagation	12.163	0.069	13.044	0.194	13.272	0.383
	Ours: No constraints	18.996	0.316	15.883	0.373	14.912	0.515
	Ours: Frequency constraints	27.412	0.655	19.773	0.572	17.344	0.650
	Ours: Spatial constraints	**36.811**	**0.903**	**27.266**	**0.829**	**21.500**	**0.806**

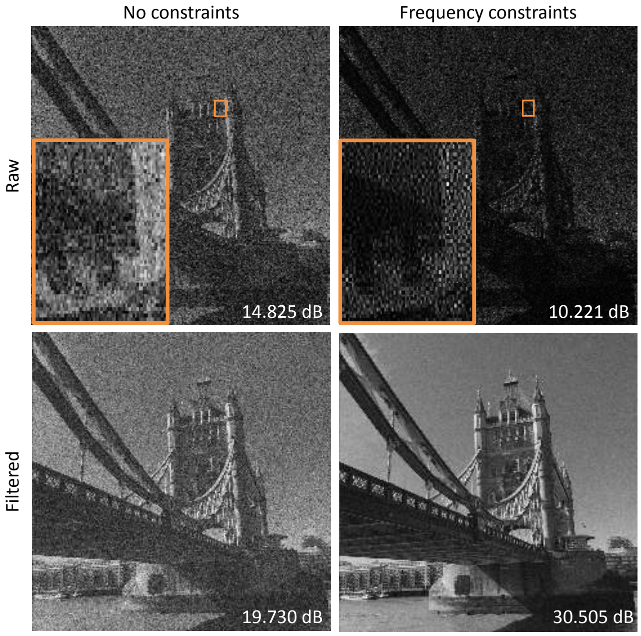

Fig. 4. By applying frequency constraints during optimization of the SLM pattern, noise is moved into higher frequencies that are imperceptible to the viewer, except through contrast loss. The unconstrained version is more similar to the target image before filtering; after low-pass filtering the output to simulate the viewer's experience (bottom), the frequency constrained result is much closer to the target. This example has 4× étendue expansion and PSNR is reported in the bottom right of each image.

lithography, and we assume the pattern is known. One of the challenges of LCoS SLMs is the relatively low diffraction efficiency that results in a strong reflection of un-modulated light, called the DC term. When the DC term passes through the mask, it scatters and creates background haze that reduces contrast. Therefore, we add a $4f$ system consisting of two Pentax FA 645 lenses (75 mm focal length, F/2.8) and place an opaque DC block (chrome on glass mask) at the Fourier plane to remove the DC term. This also allows us to relay the SLM directly onto the mask, to match the model in Eq. 8. The SLM is illuminated by a collimated beam from a laser with $\lambda = 660$ nm. We pre-calibrate the SLM phase [Bergeron et al. 1995] and measure and compensate for flatness deviations on the SLM [Xun and Cohn 2004].

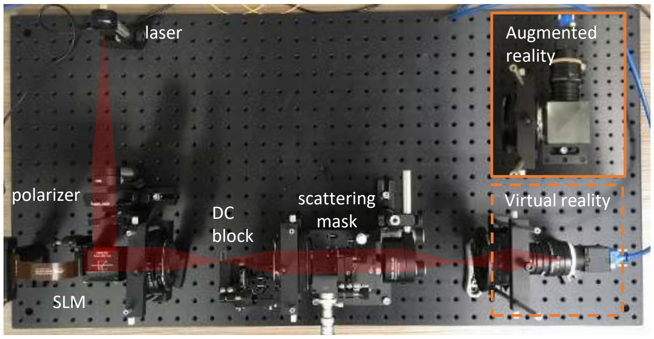

Fig. 5. Our benchtop prototype can be arranged either in a virtual reality configuration, shown above, or in an augmented reality configuration by including a beamsplitter in the imaging path, shown in inset.

After the mask, a relay system de-magnifies the image by a factor of 2×; finally, images of our display are captured with a monochromatic camera (FLIR Blackfly S BFS-U3-200S6M) with a f = 16 mm, F/1.4 C-mount lens. As with the simulations, captured experimental images are low-pass filtered to simulate the effect of low-pass filtering in the visual system. All non-linear processing in the camera, such as gamma and black level, are turned off and there are no adjustments to the black level in post processing. Figure 5 shows a schematic of the experimental prototype.

4.2.1 Mask Alignment and Calibration. The alignment of the scattering mask is critical for good performance since the SLM pattern is only valid when the mask position matches the simulation. Our custom designed mask includes coarse alignment markers (three 600 μm squares of constant phase) that are visible to the human

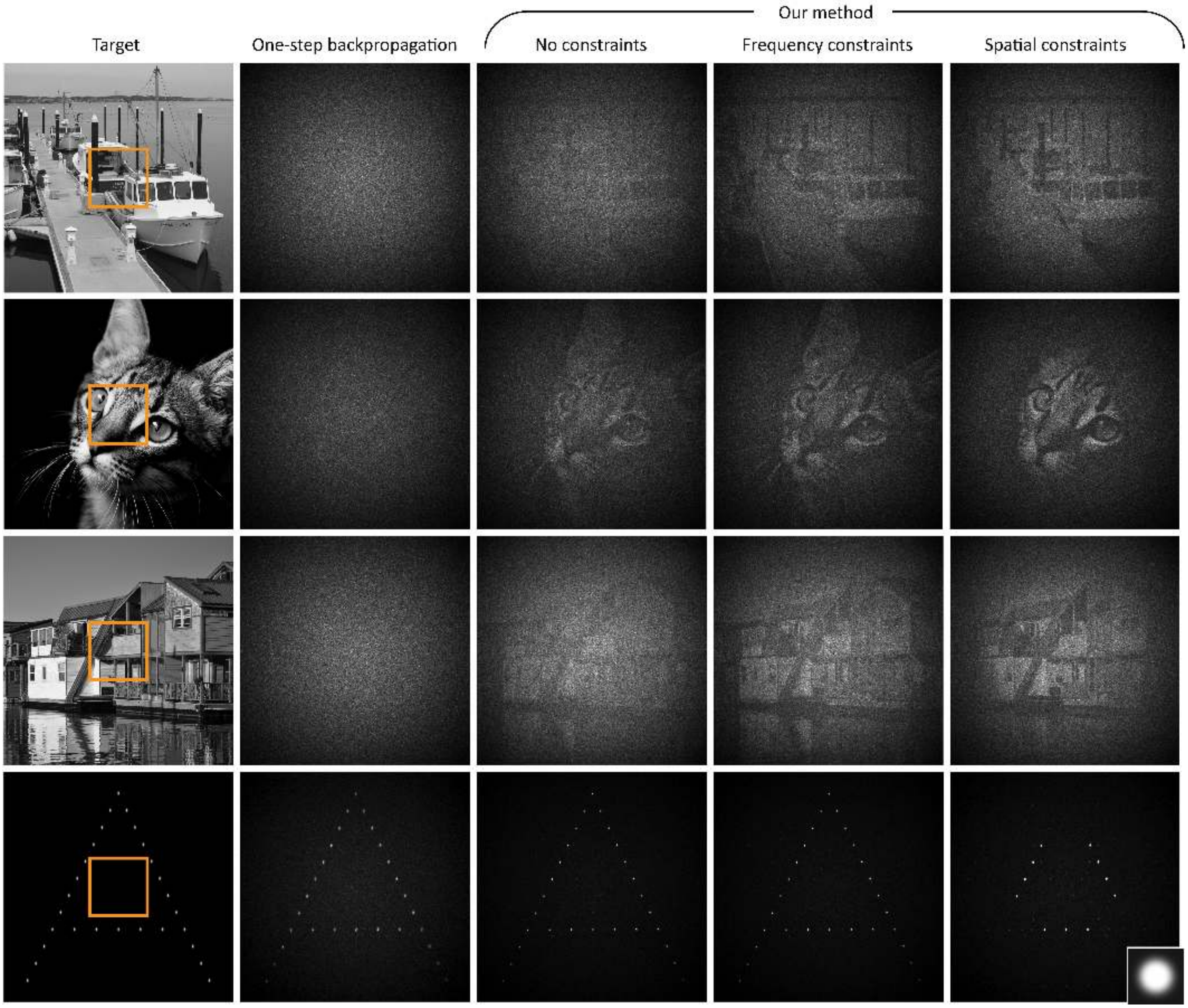

Fig. 6. Experimental results from our benchtop prototype with 16× étendue enhancement beyond the native SLM. Although the one-step backpropagation algorithm is effective on sparse scenes (bottom row), it has very low contrast on dense imagery. Contrast is improved by our iterative method and further enhanced by applying frequency and spatial constraints in the loss function. The orange box shows the native FoV of the SLM. Boat source image by Erick Bee (CC BY-SA 2.0); cat source image by Lali Masriera (CC BY 2.0); floating houses source image by Madeleine Deaton (CC BY 2.0).

eye and are used to approximately position the mask in the system. Since the mask pattern in known *a priori*, we can use our algorithm to calculate an SLM pattern that generates a single focal spot on the camera. The mask position, which is controlled by a 6-axis motion stage (Thorlabs Max313D, APY002, KM100C), is fine-tuned over the six degrees of rigid transformation to maximize the spot intensity on the camera.

In addition, geometric distortion from the $4f$ system can cause non-rigid misalignment between the mask and SLM. We coarsely compensate for this effect with the following procedure. First, we split the mask area in 3×5 subsections. For each subsection, we computationally modify our simulated mask pattern, $m(\vec{u})$, by translating the subsection of interest. We then calculate a new SLM pattern to produce a focal spot based on the modified mask. Using the camera for feedback, we optimize each subsection's translation to maximize the spot intensity, and we combine the optimal translations in a piecewise fashion to form a new mask pattern which is used for all future images. Note that we only apply translations of integer pixel values since we find interpolation of the mask pattern results in artifacts in the displayed image.

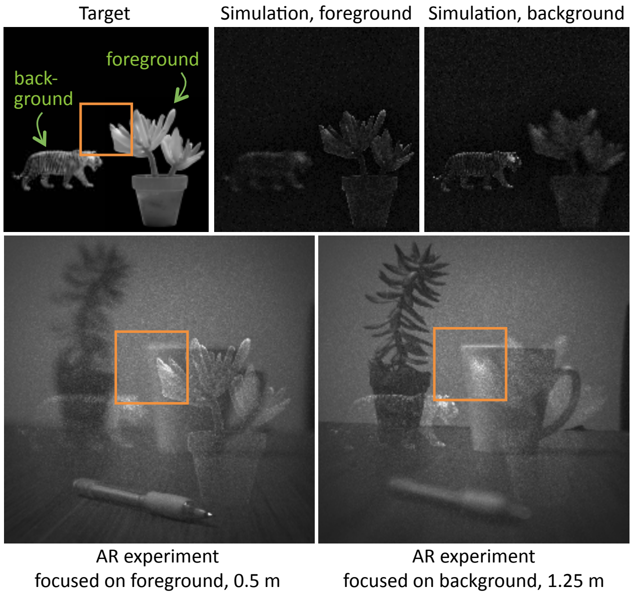
Target
Simulation, foreground
Simulation, background
foreground
back-ground
AR experiment focused on foreground, 0.5 m
AR experiment focused on background, 1.25 m

Fig. 7. Augmented reality prototype demonstration with multi-plane content shown at two different focal distances. The orange box represents the native FoV of the SLM without étendue expansion.

4.2.2 Experimental Results. Figure 6 shows captured images from our experimental prototype comparing the image calculation algorithms described in Section 3.4. All images are at 16× expansion. Although the one-step backpropagation algorithm works well for sparse scenes, it does not extend to dense, photographic imagery. As with the simulation results, we see strong improvement in contrast when using our algorithm compared to prior work, and contrast is further enhanced when using the frequency constraints. Applying spatial constraints creates higher contrast but only over a limited region determined by our provided spatial weighting map. Areas that are not prioritized by the spatial map may appear to be "missing" content due to reduced contrast in these regions, but recall that the spatial map is user-specified and can easily be translated to any location, enabling high quality content anywhere in the FoV. Although our experimental results do not yet match the quality of the simulations (discussed more in Sec. 5.1), ours is the first prototype to demonstrate dense, higher resolution imagery outside of the SLM's native FoV.

We further demonstrate the multi-plane capabilities of our system in an augmented reality (AR) prototype. We re-arrange the imaging path of our benchtop prototype to include a beamsplitter, creating a see-through path (see inset of Figure 5). Figure 7 shows a multi-plane image captured through our AR setup, displaying a small plant in the foreground and a tiger in the background. This highlights the advantages of a holographic display compared to a stereoscopic display: the holographic images contain correct monocular focal cues, which can be seen when the hologram is defocused. However, as with other holographic displays, we cannot easily display occlusions and all virtual objects appear transparent.

5 DISCUSSION

Our work represents a step towards practical holographic near-eye displays by breaking the trade-off between FoV and eyebox size. As an illustrative example, an ideal display may have a 120° FoV and a 1 cm eye box, such that the eye can rotate freely and maintain view of the image. Such a configuration could theoretically be achieved with our method by using emerging 8K SLMs (which have been demonstrated [Sterling 2008]) and a scattering mask with 16× expansion. Note that without the expansion mask, an 8K SLM scaled to have a 120° FoV would provide 66 pixels/degree of resolution, just above what normal human vision can perceive. Therefore, after applying our frequency constraints, which limits the final output resolution to the native SLM resolution, noise is pushed into imperceptible frequencies.

However, based on our simulations in Figure 3, there is a visible reduction in contrast at 16× étendue expansion. This might be acceptable in some scenarios, in particular for augmented reality in which content tends to be sparse, and light from the world changes the perception of contrast. For cases where the content is dense and the contrast loss is too severe, we propose three potential solutions, simulated in Figure 8. Although current commercial SLMs have phase-only modulation, complex modulation is a potential solution to increase performance, and is often achieved with cascaded modulators [Shi et al. 2017]. If complex SLMs are not available, another solution is based on fixed foveation: although a large FoV is important for an immersive display, the eye tends to only rotate within ±18° on average before the head moves [Fang et al. 2015]. Conservatively we will assume that half of the 120° FoV is almost always in peripheral vision and only the central 60° must be highly optimized. In Fig. 8, we simulate a fixed foveation falloff using our spatial constraints and show restored contrast in the center region, without the use of eye tracking. However, if eye tracking is available, the contrast can be further improved by moving around a smaller tracked foveal region based on the viewer's gaze direction.

5.1 Challenges and Future Work

Although our work demonstrates progress toward more practical holographic displays, there is still additional work to be done to achieve a full-color display with high resolution, complete focal depth cues, and a sunglasses-like form factor. We discuss some key challenges below.

Model Mismatch. Although our experimental system shows the potential of scattering-based étendue expansion, the contrast and quality of the early prototype is noticeably lower than that of the simulation. We conjecture that this is due to the high sensitivity of the system to alignment errors, particularly if the scattering mask is not at the correct location. Our alignment procedure removes misalignment that can be represented as a rigid transformation, but can only coarsely account for non-rigid distortion. Since the image of the SLM is relayed through a $4f$ system, we encounter geometric distortion from the lenses, even though the lenses are well corrected for aberrations. In Figure 9, we simulate the effect of a small amount of geometric distortion in the SLM pattern, less than a 30 μm translation at the edge of the SLM. Even after applying our coarse correction procedure, we see that the simulations with

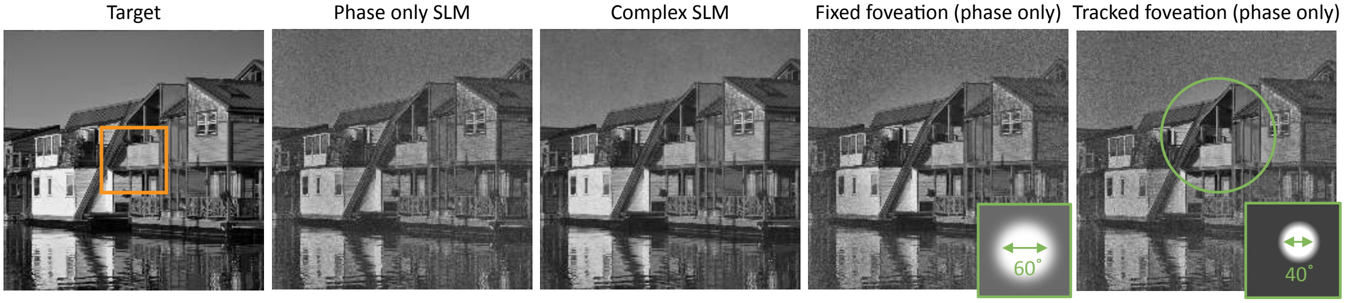

Fig. 8. We propose three methods to further improve contrast for 16× étendue enhancement beyond the performance of our baseline frequency constrained method (shown in second panel). First, with emerging complex-valued SLM technology, image contrast can be improved over the whole FoV. Second, since the eye tends to rotate only within a limited range before the head moves, in a 120° display a large portion of the FoV is almost always in the periphery. Therefore, we can improve contrast in the center by applying our spatial constraints in a fixed fovation pattern that does not require eye tracking. Finally, if eye tracking is available, our spatial constraints can be used to improve quality in a dynamically changing subregion, highlighted by the green circle. The orange box indicates the native FoV, and the spatial maps used in the foveated simulations are shown in the insets. Floating houses source image by Madeleine Deaton (CC BY 2.0).

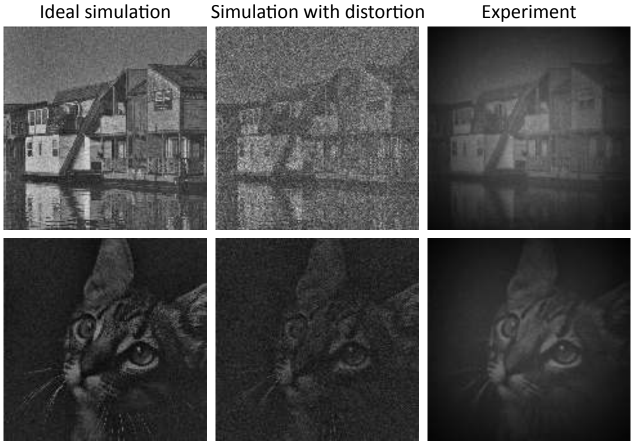

Fig. 9. Performance of our experimental prototype does not yet match the simulations. We ascribe this mismatch to the high sensitivity of the mask alignment, and although we precisely align the mask through the 6 degrees of freedom of rigid transformation, we do not accurately account for effects such as geometric distortion. Here, we simulate the effect of a small amount of distortion and observe that this model mismatch creates qualitatively similar contrast to the experimental results. Additional vignetting from the imaging system is also apparent in the experimental images. Floating houses source image by Madeleine Deaton (CC BY 2.0); cat source image by Lali Masriera (CC BY 2.0).

model mismatch yield qualitatively similar results to the experiment. This problem could be mitigated in future work by omitting the $4f$ system and using a volume grating to remove the DC term [Bang et al. 2019]. We also observe vignetting in experimental results, which can be alleviated with improved design of the imaging relay system. We find that using all the SLM pixels without binning results in further degradation in image quality, which we suspect is due to higher tolerances on alignment and increased cross-talk between the SLM pixels [Moser et al. 2019].

Miniaturization. The prototype presented in this work is intended as a proof-of-concept; the final design is ideally a wearable display with a sunglasses-like form factor. Starting with the design presented by Maimone et al. [2017], which had promising form factor and FoV but very limited eyebox, we propose integrating our scattering mask into the holographic optical element that acts as an image combiner. Figure 10 shows a simplified schematic of this idea. A display with a traditional holographic image combiner, shown on the left, is recorded by interfering two beams to create a volume hologram that relays the projected light to the eye box. To fabricate the holographic image combiner with the encoded scattering mask (Fig. 10 right), we propose placing a lithography-printed phase mask, like the one used in our benchtop prototype, in front of the holographic optical element during recording. We expect this will result in an image combiner that both relays the projector light to the eye box and implements the scattering mask in one compact optical element.

As in our benchtop system, we need an accurate model of the effect of the scattering mask such that our algorithm can compensate for the scattering in the SLM pattern. In the miniaturized off-axis configuration, we can no longer assume that the SLM and mask are parallel or at the same plane. However, these effects can be physically modeled with minimal additional compute by including free-space propagation using Fresnel or angular spectrum methods [Goodman 2005] and modelling electric fields at non-parallel planes using the fast approach described by Matsushima et al. [2003].

Color. Our prototype displays images in the red channel only; there are several additional considerations for full color display. Maimone et al. [2017] provide full color by field sequential operation, i.e., displaying holograms for the red, green, and blue channels in rapid succession on the SLM in conjunction with synchronized laser sources. Wavelength-multiplexed volume holograms are used for the static optical components, which allow independent operation for each color with very little crosstalk. A per-channel system calibration process is also used to reduce any residual differences between color channels. We could also apply the strategy of Maimone et al. [2017] by using field sequential operation on our SLM, replacing

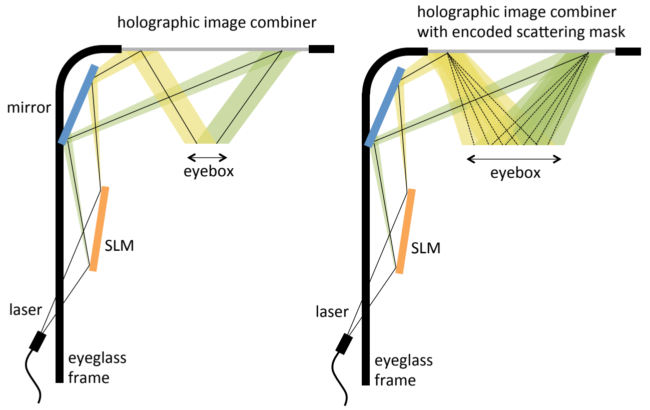
holographic image combiner
mirror
eyebox
SLM
laser
eyeglass frame
holographic image combiner with encoded scattering mask
eyebox
SLM
laser
eyeglass frame

Fig. 10. Proposed future scheme for integrating our étendue expansion mask into an sunglasses-like form factor display. (Left) Schematic of compact holographic display prototype based on the work of Maimone et al. [2017]. (Right) By encoding the scattering mask in the holographic image combiner, the eyebox can be increased without sacrificing FoV, adding additional optical components, or compromising form factor.

our binary phase scattering mask with a wavelength-multiplexed volume hologram optimized for each color channel, and performing the calibration procedure of Section 4.2 for each channel. We expect that his method will be successful for our proposed display as each color channel can be optimized and calibrated independently; however, we have yet to experimentally validate this method.

Compute Time. Currently, compute is performed offline and takes about 3 minutes to generate the hologram for the binned 960 × 540 SLM with 16× étendue expansion in MATLAB code on a GeForce GTX 1060 GPU. Accelerating compute was not prioritized in this work, but will be necessary to make a practical real-time display. Improved hardware acceleration will be critical, and future work incorporating temporal consistency into the hologram calculation might further reduce compute time for each frame.

Perceptual effects. In this work, we model the human visual system simply as a low-pass filter that removes high spatial frequencies. However, it is possible human subjects may actually perceive high frequencies in the image in more complicated ways. In order to ensure that the viewer experiences the desired effect, future work includes perceptual studies of the display architecture and development of corresponding biologically-inspired loss functions.

5.2 Conclusion

In this work, we described a physically simple approach to increase the étendue of holographic near-eye displays. Our method requires only a static scattering mask, making our design well suited for miniaturization, and unlike prior methods, scales to complex, full resolution imagery. Using enhanced algorithms as well as spatial and frequency constraints, we showed that we can increase étendue significantly (e.g. 4× to 16×) through simulation and preliminary experimental results. Although experimental results must be improved for practical use, we are encouraged at the prospect of new algorithms enabling practical future holographic displays using emerging spatial light modulator technology.

ACKNOWLEDGMENTS

We thank the team at Facebook Reality Labs for their support on this work. Changwon Jang, Douglas Lanman, Gang Li, Heeyoon Lee, Scott McEldowney, Tanya Malhotra, and Kiseung Bang provided discussion and advice. Giuseppe Calafiore fabricated the binary phase mask with support from Jessica Chau, Elliott Franke, and Emily Makoutz. Some of the source images shown in the figures were provided under Creative Commons licenses (CC BY 2.0[1] and CC BY-SA 2.0[2]), as indicated in the figure captions.

REFERENCES

Duygu Akbulut, Thomas J Huisman, Elbert G van Putten, Willem L Vos, and Allard P Mosk. 2011. Focusing light through random photonic media by binary amplitude modulation. *Optics express* 19, 5 (2011), 4017–4029.

Kiseung Bang, Changwon Jang, and Byoungho Lee. 2019. Compact noise-filtering volume gratings for holographic displays. *Optics letters* 44, 9 (2019), 2133–2136.

Amir Beck and Marc Teboulle. 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences* 2, 1 (2009), 183–202.

Alain Bergeron, Jonny Gauvin, François Gagnon, Denis Gingras, Henri H Arsenault, and Michel Doucet. 1995. Phase calibration and applications of a liquid-crystal spatial light modulator. *Applied optics* 34, 23 (1995), 5133–5139.

Edward Buckley, Adrian Cable, Nic Lawrence, and Tim Wilkinson. 2006. Viewing angle enhancement for two-and three-dimensional holographic displays with random superresolution phase masks. *Applied optics* 45, 28 (2006), 7334–7341.

Emmanuel J Candes, Xiaodong Li, and Mahdi Soltanolkotabi. 2015. Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory* 61, 4 (2015), 1985–2007.

Thomas Chaigne, Jérôme Gateau, Ori Katz, Emmanuel Bossy, and Sylvain Gigan. 2014. Light focusing and two-dimensional imaging through scattering media using the photoacoustic transmission matrix with an ultrasound array. *Optics letters* 39, 9 (2014), 2664–2667.

Praneeth Chakravarthula, Yifan Peng, Joel Kollin, Henry Fuchs, and Felix Heide. 2019. Wirtinger holography for near-eye displays. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 213.

Julio Chaves. 2017. *Introduction to nonimaging optics.* CRC press.

Myeong-Ho Choi, Yeon-Gyeong Ju, and Jae-Hyeung Park. 2020. Holographic near-eye display with continuously expanded eyebox using two-dimensional replication and angular spectrum wrapping. *Opt. Express* 28, 1 (Jan 2020), 533–547. https://doi.org/10.1364/OE.381277

Donald B Conkey, Antonio M Caravaca-Aguirre, Jake D Dove, Hengyi Ju, Todd W Murray, and Rafael Piestun. 2015. Super-resolution photoacoustic imaging through a scattering wall. *Nature communications* 6 (2015), 7902.

Donald B Conkey, Antonio M Caravaca-Aguirre, and Rafael Piestun. 2012. High-speed scattering medium characterization with application to focusing light through turbid media. *Optics express* 20, 2 (2012), 1733–1740.

Yu Fang, Ryoichi Nakashima, Kazumichi Matsumiya, Ichiro Kuriki, and Satoshi Shioiri. 2015. Eye-head coordination for visual cognitive processing. *PloS one* 10, 3 (2015).

A Georgiou, J Christmas, N Collings, J Moore, and WA Crossland. 2008. Aspects of hologram calculation for video frames. *Journal of Optics A: Pure and Applied Optics* 10, 3 (2008), 035302.

Ralph W Gerchberg and W. 0. Saxton. 1972. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* 35 (1972), 237–246.

Joseph W Goodman. 2005. *Introduction to Fourier optics.* Roberts and Company Publishers.

Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. 2012. Foveated 3D graphics. *ACM Transactions on Graphics (TOG)* 31, 6 (2012), 164.

Alain Hore and Djemel Ziou. 2010. Image quality metrics: PSNR vs. SSIM. In *2010 20th International Conference on Pattern Recognition.* IEEE, 2366–2369.

Changwon Jang, Kiseung Bang, Gang Li, and Byoungho Lee. 2018. Holographic Near-Eye Display with Expanded Eye-Box. *ACM Trans. Graph.* 37, 6, Article Article 195 (Dec. 2018), 14 pages. https://doi.org/10.1145/3272127.3275069

Anton S Kaplanyan, Anton Sochenov, Thomas Leimkühler, Mikhail Okunev, Todd Goodall, and Gizem Rufo. 2019. DeepFovea: neural reconstruction for foveated rendering and video compression using learned statistics of natural videos. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–13.

[1] https://creativecommons.org/licenses/by/2.0/

[2] https://creativecommons.org/licenses/by-sa/2.0/

Mugeon Kim, Sungjin Lim, Geunseop Choi, Youngmin Kim, Hwi Kim, and Joonku Hahn. 2018. Expanded Exit-Pupil Holographic Head-Mounted Display With High-Speed Digital Micromirror Device. *ETRI Journal* 40, 3 (2018), 366–375. https://doi.org/10.4218/etrij.2017-0166 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.4218/etrij.2017-0166

Gang Li, Dukho Lee, Youngmo Jeong, Jaebum Cho, and Byoungho Lee. 2016. Holographic display for see-through augmented reality using mirror-lens holographic optical element. *Optics letters* 41, 11 (2016), 2486–2489.

Andrew Maimone, Andreas Georgiou, and Joel S Kollin. 2017. Holographic near-eye displays for virtual and augmented reality. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 85.

Kyoji Matsushima, Hagen Schimmel, and Frank Wyrowski. 2003. Fast calculation method for optical diffraction on tilted planes by use of the angular spectrum of plane waves. *JOSA A* 20, 9 (2003), 1755–1762.

Simon Moser, Monika Ritsch-Marte, and Gregor Thalhammer. 2019. Model-based compensation of pixel crosstalk in liquid crystal spatial light modulators. *Optics express* 27, 18 (2019), 25046–25063.

Jongchan Park, KyeoReh Lee, and YongKeun Park. 2019. Ultrathin wide-angle large-area digital 3D holographic display using a non-periodic photon sieve. *Nature communications* 10, 1 (2019), 1304.

Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 179.

SM Popoff, G Lerosey, R Carminati, M Fink, AC Boccara, and S Gigan. 2010. Measuring the transmission matrix in optics: an approach to the study and control of light propagation in disordered media. *Physical review letters* 104, 10 (2010), 100601.

Liang Shi, Fu-Chung Huang, Ward Lopes, Wojciech Matusik, and David Luebke. 2017. Near-eye light field holographic rendering with spherical waves for wide field of view interactive 3D computer graphics. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 236.

Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein. 2018. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–13.

Rod Sterling. 2008. JVC D-ILA high resolution, high contrast projectors and applications. In *Proceedings of the 2008 workshop on Immersive projection technologies/Emerging display technologiges*. ACM, 10.

Xiaodong Tao, Dare Bodington, Marc Reinig, and Joel Kubby. 2015. High-speed scanning interferometric focusing by fast measurement of binary transmission matrix for channel demixing. *Optics express* 23, 11 (2015), 14168–14187.

Ivo M Vellekoop, Aart Lagendijk, and AP Mosk. 2010. Exploiting disorder for perfect focusing. *Nature photonics* 4, 5 (2010), 320.

Ivo M Vellekoop and AP Mosk. 2007. Focusing coherent light through opaque strongly scattering media. *Optics letters* 32, 16 (2007), 2309–2311.

Koki Wakunami, Po-Yuan Hsieh, Ryutaro Oi, Takanori Senoh, Hisayuki Sasaki, Yasuyuki Ichihashi, Makoto Okui, Yi-Pai Huang, and Kenji Yamamoto. 2016. Projection-type see-through holographic three-dimensional display. *Nature communications* 7, 1 (2016), 1–7.

Xiaodong Xun and Robert W Cohn. 2004. Phase calibration of spatially nonuniform spatial light modulators. *Applied optics* 43, 35 (2004), 6400–6406.

Fahri Yaraş, Hoonjong Kang, and Levent Onural. 2009. Real-time phase-only color holographic video display system using LED illumination. *Applied optics* 48, 34 (2009), H48–H53.

Li-Hao Yeh, Shwetadwip Chowdhury, Nicole A Repina, and Laura Waller. 2019. Speckle-structured illumination for 3D phase and fluorescence computational microscopy. *Biomedical Optics Express* 10, 7 (2019), 3635–3653.

Jonghee Yoon, KyeoReh Lee, Jongchan Park, and YongKeun Park. 2015. Measuring optical transmission matrices by wavefront shaping. *Optics Express* 23, 8 (2015), 10158–10167.

Hyeonseung Yu, KyeoReh Lee, Jongchan Park, and YongKeun Park. 2017b. Ultrahigh-definition dynamic 3D holographic display by active control of volume speckle fields. *Nature Photonics* 11, 3 (2017), 186.

HyeonSeung Yu, KyeoReh Lee, and YongKeun Park. 2017a. Ultrahigh enhancement of light focusing through disordered media controlled by mega-pixel modes. *Optics express* 25, 7 (2017), 8036–8047.

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

A APPENDIX

Here we provide additional details on the iterative algorithm used in this work. Although there are several options for solving Eq. 14, we use projected gradient descent with Nesterov acceleration, which is based on the work of Beck and Teboulle [2009] and summarized in Algorithm 1. In the algorithm, μ is the user-defined step-size, $\nabla\mathcal{L}$ is the gradient of the loss with respect to $\vec{s}$, and $\text{prox}\{\cdot\}$ is the proximal operator that constrains the SLM pattern to be phase-only:

$$\text{prox}\{\vec{s}\} = \vec{s}/|\vec{s}|. \tag{24}$$

The proximal operator is equivalent to setting the amplitude of every element of $\vec{s}$ to one.

Since implementation of the algorithm requires discretization of all variables, we summarize the complete loss function below in discrete vector representation.

$$\begin{aligned}
\mathcal{L} &= \tfrac{1}{2}||\vec{g}||^2, \\
\vec{g} &= \vec{c}_s \odot \mathcal{F}^{-1}\{\vec{c}_f \odot \mathcal{F}\{\vec{I} - \hat{I}\}\}, \\
\vec{I} &= |\vec{y}|^2, \\
\vec{y} &= \mathcal{F}\{\vec{m} \odot U\vec{s}\}.
\end{aligned} \tag{25}$$

Here, $\mathcal{L}$ is the scalar loss, $\vec{g}$ is a new intermediate variable, $\odot$ represents element-wise multiplication, and U is an upsampling operation with a box filter (the discrete version of Eq. 15). $\vec{c}_f, \vec{c}_s, \vec{I}, \hat{I}, \vec{y}$ and $\vec{m}$ are discrete vector versions of $c_f(\vec{u}')$, $c_s(\vec{x})$, $I(\vec{x})$, $\hat{I}(\vec{x})$, $y(\vec{x})$ and $m(\vec{u})$ respectively. The above equations use the custom loss function from Eq. 19, but the other loss functions can be achieved by setting $\vec{c}_s$ or $\vec{c}_f$ (or both) to all ones.

The gradient $\nabla\mathcal{L}$ is with respect to the complex variable $\vec{s}$, so we calculate the gradient using Wirtinger derivatives [Candes et al. 2015].

$$\nabla\mathcal{L} = \left(\frac{d\mathcal{L}}{d\vec{s}}\right)^* = \left(\frac{d\vec{I}}{d\vec{s}}\right)^* \left(\frac{d\mathcal{L}}{d\vec{I}}\right). \tag{26}$$

Note that $\mathcal{L}$ and $\vec{I}$ are both real valued, so there is no need for the complex conjugate on the second term. Since $\frac{d\vec{I}}{d\vec{s}}$ yields a matrix, we can think of it as an operator that acts on the vector $\frac{d\mathcal{L}}{d\vec{I}}$. This yields the gradient needed for Algorithm 1:

$$\begin{aligned}
\nabla\mathcal{L} &= U^T\left\{\vec{m}^* \odot \mathcal{F}^{-1}\left\{\vec{y} \odot \left(\frac{d\mathcal{L}}{d\vec{I}}\right)\right\}\right\}, \\
\left(\frac{d\mathcal{L}}{d\vec{I}}\right) &= \mathcal{F}^{-1}\{\vec{c}_f \odot \mathcal{F}\{\vec{c}_s \odot \vec{g}\}\},
\end{aligned} \tag{27}$$

where U^T is the downsampling (binning) operation such that $\nabla\mathcal{L}$ has the same number of elements as $\vec{s}$

ALGORITHM 1: Proj. Gradient Descent with Nesterov Acceleration

```
Initialize: $\vec{s}_0$ with uniformly distributed random phase
$t_0 = 1, \vec{s}'_0 = \vec{s}_0$
Repeat:
    $\vec{s}'_{k+1} \leftarrow \text{prox}\{\vec{s}_k - \mu\nabla\mathcal{L}\}$
    $t_{k+1} \leftarrow \frac{1+\sqrt{1+4t_k^2}}{2}$
    $\vec{s}_{k+1} \leftarrow \vec{s}'_{k+1} + \frac{t_k - 1}{t_{k+1}}\left(\vec{s}'_{k+1} - \vec{s}'_k\right)$
```