

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Computing high quality phase-only holograms for holographic displays

Chakravarthula, Praneeth, Peng, Yifan, Kollin, Joel, Heide, Felix, Fuchs, Henry

Praneeth Chakravarthula, Yifan Peng, Joel Kollin, Felix Heide, Henry Fuchs, "Computing high quality phase-only holograms for holographic displays," Proc. SPIE 11310, Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR), 1131006 (21 February 2020); doi: 10.1117/12.2547647

SPIE.

Event: SPIE AR VR MR, 2020, San Francisco, California, United States

Computing high quality phase-only holograms for holographic displays

Praneeth Chakravarthula^a, Yifan Peng^b, Joel Kollin^c, Felix Heide^d, and Henry Fuchs^a

^aThe University of North Carolina at Chapel Hill, USA

^bStanford University, USA

^cMicrosoft Research, USA

^dPrinceton University, USA

ABSTRACT

Holography has demonstrated potential to achieve a wide field of view, focus supporting, optical see-through augmented reality display in an eyeglasses form factor. Although phase modulating spatial light modulators are becoming available, the phase-only hologram generation algorithms are still imprecise resulting in severe artifacts in the reconstructed imagery. Since the holographic phase retrieval problem is non-linear and non-convex and computationally expensive with the solutions being non-unique, the existing methods make several assumptions to make the phase-only hologram computation tractable. In this work, we deviate from any such approximations and solve the holographic phase retrieval problem as a quadratic problem using complex Wirtinger gradients and standard first-order optimization methods. Our approach results in high-quality phase hologram generation with at least an order of magnitude improvement over existing state-of-the-art approaches.

Keywords: Holographic displays, Phase-only holograms, Phase retrieval, Near eye display, Augmented reality, Wide field of view, Vergence-accommodation conflict

1. INTRODUCTION

Near-eye displays (NEDs) open a wide range of possibilities by integrating computer graphics into day-to-day human vision. NEDs for virtual reality (VR) have already become ubiquitous consumer products, while today's augmented reality (AR) devices partially inherit the advances of VR displays though are still limited to large form factors and are unable to allow for occlusion and continuous focus cues to avoid vergence-accommodation conflict. Approaches using conventional optics to overcome these challenges are often restricted to large setups.¹ Mass adoption of augmented reality (AR), however, relies on display devices that can seamlessly integrate into human vision, just like everyday prescription eyeglasses, and provide superior quality virtual imagery.

Digital holography, in theory, offers an elegant way to solve the complex optical design challenges for near-eye displays. Instead of physically controlling the emitted wavefront of light with cascades of (variable) refractive optics and conventional displays, holographic displays shift this process to computation. A phase-only digital hologram can encode a scene intensity image at a given distance into a phase pattern. Assuming for a moment that the encoding process is perfect, this approach allows for high-resolution and continuous focus cues. Building on precise phase modulation hardware and a large body of existing work in holography, the recent seminal works^{2,3} demonstrate impressive high-FOV holographic near-eye displays in a light-weight wearable form-factor. Although these approaches offer insight into the promise of holographic near-eye displays, they rely on heuristic holographic phase encoding schemes that severely restrict the achievable image quality and future research.

Similar to the work we first introduced as "Wirtinger Holography" in Ref. 4, we present in this paper an improved hologram generation algorithm where we depart from existing heuristic encoding methods and iterative approaches. Instead, we pose the hologram generation process as a quadratic problem which we solve using complex Wirtinger derivatives and first-order optimization methods. Specifically, our method does not require any relaxations but can be solved directly with standard methods. We solve for high quality Fresnel

Further author information: (Send correspondence to Praneeth Chakravarthula)
Praneeth Chakravarthula: E-mail: cpk@cs.unc.edu

holograms using quasi-Newton method or stochastic gradient descent methods and demonstrate at least a gain of 10 dB in PSNR values of the simulated holographic reconstructions. We further validate the proposed method on an experimental prototype and demonstrate that our approach eliminates severe artifacts present in existing methods. We recommend the reader to also refer to our previous paper Ref. 4 for a more comprehensive understanding of the algorithm.

Contributions Our primary contribution is an improved flexible algorithm for generating high-quality phase-only holograms for a compact form factor holographic near-eye display for augmented and virtual reality applications. Specific contributions are as follows:

- We present Wirtinger calculus and Wirtinger gradients in the context of optimization problems involving derivatives of complex variables.
- We use Wirtinger gradients to directly solve holographic phase retrieval as formal optimization problems using first-order optimization methods.
- We validate the proposed method in simulation and achieve an order of magnitude lower error in holographic reconstructions compared to previous methods.
- We demonstrate the proposed method on an experimental prototype holographic display and show that our approach reduces severe artifacts present in existing hologram generation methods.

2. RELATED WORK

Computational holograms are traditionally classified into Fourier holograms corresponding to far-field Fraunhofer diffraction, and Fresnel holograms which produce images in the near field, as determined by the Fresnel Number. The computational complexity between near-field and far-field holograms is substantial. While computing far-field holograms using 2D Fourier transforms can be accomplished using fast algorithms and optimized hardware, computing near-field Fresnel holograms admits no simple analytic solution. In fact, it is mathematically equivalent to inverting a generalized scalar diffraction integral.⁵ A majority of work on this topic investigates numerical techniques for generating Fresnel fringe patterns. In this section, we briefly summarize prior art related to hologram computation and display.

2.1 Traditional Phase Retrieval Algorithms

A Fourier hologram produces a flat image at a far distance Fourier plane, and such a hologram is often computed using traditional phase retrieval techniques. Phase retrieval is the method of recovering an unknown signal from the measured magnitude of its Fourier transform. Since the phase is lost in the measurement of the signal, the inverse problem of recovering it is generally ill-posed. However, the phase can be perfectly recovered, in theory, by solving a set of non-linear equations if the measurements are sufficiently oversampled.⁶ Early methods of phase retrieval included error reduction methods using iterative optimization,^{7,8} together with an assumption on a non-zero support of the real-valued signal, with applications in optics, crystallography, biology and physics. Extension of such iterative algorithm is the popular hybrid input-output (HIO) method,⁹ and others with various relaxations.^{10,11} Phase-retrieval methods using first-order non-linear optimization have been explored in the past to characterize complex optical systems,^{12–14} eventually also sparking recent work on using alternative direction methods for phase retrieval,^{15,16} non-convex optimization,¹⁷ and methods overcoming the non-convex nature of the phase retrieval problem by lifting, i.e. relaxation, to a semidefinite¹⁸ or linear program.^{19,20}

2.2 Computational Fresnel holograms

Fresnel hologram computation can be categorized into two classes. 1) Geometry-based techniques, which model three-dimensional scene geometry as a collection of emitters; either point-emitters (point-source method) or polygonal tiles (polygon-based method). The collective interference of these emitters with a reference wave is computed at a set of discretized locations throughout the combined field to generate a hologram of the scene.²¹ 2) Image-based techniques, which leverage the advantage of computer graphics rendering techniques along with wave propagation. Next, we review geometry and image-based methods.

Point-source methods Waters et al.²² were the first to propose using a collection of points rather than finite sized objects to model a scene for holographic image generation. By using a look-up table of precomputed elemental fringes, Lucente et al.²³ sped up the point-source hologram computation to under one second. Recent point-source based CGH computation methods leverage the parallelization of a computer graphics card.^{24–26} Most recent work of Maimone et al.² present point-source based CGH computation for holographic near-eye displays for both virtual and augmented reality. All of these methods have in common that the optical transmission process of different primitives is modeled to be independent, and hence, it is challenging to accurately represent view-dependent mutual-occlusion and shading. Moreover, a continuous parallax demands a very dense set of point-sources which requires a large compute budget.

Polygonal methods Polygonal methods for computing CGH also have existed for decades.²⁷ The basic idea of this approach to CGH is to represent a 3D object as a collection of tilted and shifted planes. The diffraction patterns from a tilted input plane can be computed by the fast Fourier transform (FFT) algorithm and an additional coordinate transformation. This analysis can also be done in the spatial frequency domain by using the translation and rotation transformations of the angular spectra permitting the use of FFT algorithms.²⁸ Moreover, a property function can be defined for each planar input surface to provide texture²⁹ and shading.^{29,30} Researchers also have explored geometric facet selection methods by ray tracing,³¹ silhouette methods^{32,33} and inverse orthographic projection techniques³⁴ to provide occlusion culling effects.

Image-based methods Image-based holography techniques can be broadly categorized into light field and layer-based methods. Light field holograms, also known as holographic stereograms, partition a hologram spatially into elementary hologram patches, each producing local ray distributions (images) that together reconstruct multiple views supporting intra-ocular occlusions.^{35–37} Holographic stereograms can be paired with the point-source method to enhance the image fidelity to provide accommodation and occlusion cues.^{3,38} Layer-based methods, in contrast, compute the hologram by slicing objects at multiple depths and superimposing the wavefronts from each slice on the hologram plane.^{39,40} Layer-based and light field methods can both be combined to produce view-dependent occlusion effects.^{41,42} While prior work on holographic phase retrieval addressed computing holograms for both 2D and 3D scenes, the focus of this paper lies on rendering high-quality images of scenes at a single depth plane. The method can be extended to volumetric scenes as discussed in Section ??.

3. WIRTINGER CALCULUS FOR OPTIMIZATION

In this section, we provide a quick introduction to Wirtinger calculus and Wirtinger derivatives in the context of optimization problems which involve derivatives of a complex variable. However, we refer interested readers to⁴³ for a detailed discussion on complex functions and analysis.

Consider a real-valued function of a real variable, for instance:

$$f : \mathbb{R} \ni x \mapsto y = f(x) \in \mathbb{R}, \quad (1)$$

the point x_{opt} , where $f(x)$ is maximum (or minimum) is obtained by taking the derivative of f concerning x and setting it zero. In other words, for x_{opt} , the following equation has to be valid

$$\left. \frac{df}{dx} \right|_{x_{opt}} = 0. \quad (2)$$

Note that here we assume $f(x)$ to be continuous in some domain $\mathbb{D}$, and the derivative to exist. However, whether the obtained solution gives a maximum or minimum value needs to be checked using additional conditions or higher-order derivatives.

We can define the derivative for a complex function of a complex variable, for instance

$$f : \mathbb{C} \ni z \mapsto w = f(z) \in \mathbb{C}, \quad (3)$$

as follows

$$f'(z_0) = \left. \frac{df}{dz} \right|_{z_0} = \lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0}. \quad (4)$$

If $f'(z)$ exists in the domain $\mathbb{D} \subset \mathbb{R}$, the function $f(z)$ is called an *analytic or a holomorphic* function in $\mathbb{R}$.

Now, a given complex function can be decomposed into two real functions, each depending on two real variables, say x and y , which are the real and imaginary parts of the complex variable z . Mathematically, this can be represented as

$$f(z) = f(x + jy) \equiv u(x, y) + jv(x, y); z = x + jy \quad (5)$$

It can be shown that for the above function $f(z)$ to be holomorphic, the corresponding component functions $u(x, y)$ and $v(x, y)$ need to satisfy the Cauchy-Riemann conditions⁴³

$$\frac{\partial u(x, y)}{\partial x} = \frac{\partial v(x, y)}{\partial y} \quad (6)$$

$$\frac{\partial v(x, y)}{\partial x} = -\frac{\partial u(x, y)}{\partial y} \quad (7)$$

If the above conditions are valid, then the complex derivative of a holomorphic function $f(z)$ can be expressed by the partial derivatives of the component real functions $u(x, y)$ and $v(x, y)$ as

$$\frac{df(z)}{dz} = \frac{\partial u(x, y)}{\partial x} + j \frac{\partial v(x, y)}{\partial x} \quad (8)$$

3.1 Real functions of complex variables

It can be seen from Equation 8 that if the function $f(z)$ is a real valued function, then the component function $v(x, y)$ which forms the imaginary part of $f(z)$ is zero, i.e. $v(x, y) = 0$. Following Cauchy-Riemann conditions as described in Equation 6, we see that

$$\frac{\partial u(x, y)}{\partial x} = \frac{\partial v(x, y)}{\partial y} = 0 \quad (9)$$

Thus, $u(x, y)$ is always a constant, and so is the function $f(z)$, and the derivative of such a function is always zero. In other words, a real function of complex variable is not holomorphic unless it is a real constant.

To work around this, instead of treating $f(z)$ as a real function of a complex variable, we regard $f(z) = u(x, y)$ as a function of two real variables. In the context of solving an optimization problem, it can be treated as a multidimensional real function. Therefore, to find the optimum value of $f(z)$, we want to find the optimum of $u(x, y)$, which requires

$$\frac{\partial u(x, y)}{\partial x} = 0 \quad \text{and} \quad \frac{\partial u(x, y)}{\partial y} = 0 \quad (10)$$

Note that $v(x, y) = 0$ since the function is real valued.

Both of the above real-valued equations for the optimal components x_{opt} and y_{opt} can be linearly combined into a more compact one complex-valued representation as follows

$$\alpha_1 \frac{\partial u(x, y)}{\partial x} + j \alpha_2 \frac{\partial u(x, y)}{\partial y} = 0 + j0 = 0 \quad (11)$$

where α_1 and α_2 are arbitrary real valued non-zero constants. Note that the above representation is valid since the real and imaginary parts are orthogonal, and the linear combination is only intended to devise a compact representation.

Let us now define a differential operator writing the real and imaginary parts of $z = x + jy$ as the tuple (x, y)

$$\frac{\partial}{\partial z} = \alpha_1 \frac{\partial}{\partial x} + j \alpha_2 \frac{\partial}{\partial y}. \quad (12)$$

Note that the operator defined in Equation 12 can also be applied to complex functions, because real cost functions can often be decomposed into complex components, for example $f(z) = |z|^2 = z \cdot \bar{z} = f_1(z)f_2(z)$, where $f_1(z), f_2(z) \in \mathbb{C}$. Also, observe that $\bar{z} = x - jy$ is the complex conjugate of $z = x + jy$.

The choice of values $\alpha_1 = \frac{1}{2}$ and $\alpha_2 = -\frac{1}{2}$ meets the requirements for defining the complex derivatives, and it was first introduced by *Wilhelm Wirtinger*. Wirtinger derivatives of a (complex) function $f(z)$ of a complex variable $z = x + jy$ are defined as the following linear partial differential operators of the first order

$$\frac{\partial f}{\partial z} = \frac{1}{2} \cdot \left(\frac{\partial f}{\partial x} - j \frac{\partial f}{\partial y} \right) \quad (13)$$

and

$$\frac{\partial f}{\partial \bar{z}} = \frac{1}{2} \cdot \left(\frac{\partial f}{\partial x} + j \frac{\partial f}{\partial y} \right) \quad (14)$$

For functions of many variables, $f : \mathbb{C}^n \ni z = [z_1, z_2, \dots, z_n]^T \mapsto w = f(z) \in \mathbb{R}$, all n partial derivatives with respect to the complex variables $z_1, \dots, z_n$ need to be calculated. In other words, the derivatives need to be calculated and combined into a vector again, and this *gradient* of the function can be defined as follows

$$\frac{\partial f}{\partial z} = \begin{bmatrix} \frac{\partial f}{\partial z_1} \\ \frac{\partial f}{\partial z_2} \\ \vdots \\ \frac{\partial f}{\partial z_n} \end{bmatrix} \quad \text{and} \quad \frac{\partial f}{\partial \bar{z}} = \begin{bmatrix} \frac{\partial f}{\partial \bar{z}_1} \\ \frac{\partial f}{\partial \bar{z}_2} \\ \vdots \\ \frac{\partial f}{\partial \bar{z}_n} \end{bmatrix} \quad (15)$$

Of course, the Wirtinger gradients defined in Equation 15 has to be equal to the zero vector $\mathbf{0} = [0, 0, \dots, 0]^T$. Note that for such multidimensional functions, one can separate and inspect individually the real and imaginary parts. But using the above definition, we can use simple arithmetic rules that can be expressed using vectors and matrices. This flexibility allows us to use formal optimization methods easily.

4. HIGH-QUALITY HOLOGRAPHIC PHASE RETRIEVAL

Computer generated holography (CGH) simulates the physical process of a real hologram recording and reconstruction using numerical methods. The computed digital hologram is displayed on a phase modifying spatial light modulator (SLM) that is illuminated by a laser source. The phase pattern displayed on the SLM introduces a delay to the phase of the incident wave on the SLM. This modulated wave travels a certain (propagation) distance before interfering to form the intended imagery. This wave propagation process can be treated as digital signal processing of light, and several numerical propagation methods exist for synthesizing the object waves in a digital hologram.

The numerical wave propagation methods broadly fall under either the fast Fourier transform (FFT) based methods or the convolution-based methods. While the FFT based methods are simple and fast to use, the sampling window size is proportional to the propagation distance causing the final image window size to vary for different image depths. This problem can, however, be overcome by using a multi-step Fresnel method⁴⁴ which adds the ability to control the sampling intervals. On the other hand, convolution-based methods require at least two FFTs in the computation, but the sampling window size and sampling intervals remain same as that of the source field, i.e. unlike FFT based methods, they are not dependent on the propagation distance.

Both FFT based methods or convolution-based methods often make (paraxial) approximations to make the numerical method computationally tractable. Such approximations result in phase errors which generally manifest as curvature errors. Computing the wave propagation using a more inconvenient formulation, such as

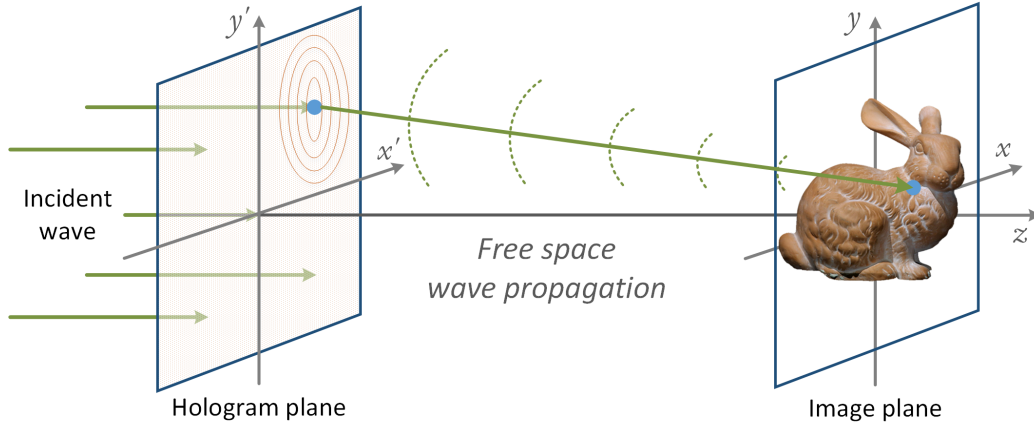


Figure 1: Schematic of scalar diffraction model, that the incident wave first gets modulated by the phase delays on the hologram, where each pixel then emits a sub-wave that propagates in free space to the image plane. The complex amplitude of one point on the image plane is the integration of sub-waves propagated from all pixels on the hologram plane.

the Kirchoff or Rayleigh-Sommerfeld formulations, result in an accurate solution but is computationally very expensive. However, such a formulation is suitable for both paraxial and non-paraxial wave fields.

In this section we introduce near-field holographic image formation for near-eye display applications using a band-limited angular spectrum method (ASM). Since the formulae for the ASM can be derived directly from the Kirchoff or Rayleigh-Sommerfeld diffraction theory without any approximations, the numerical simulation is more accurate compared to the other propagation models. We then discuss our algorithm for optimizing phase-only holograms using complex Wirtinger derivatives as discussed in the earlier Section 3.

4.1 Image formation model

The diffracted field at any arbitrary aperture can be calculated accurately using the Rayleigh-Sommerfeld (RS) diffraction integral. For a hologram illuminated using a laser source as shown in Figure 1, the diffracted wave field at a distance z due to the hologram aperture can be calculated using the RS integral as:

$$U(x, y; z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(x', y'; 0) \frac{\exp(jk\rho)}{\rho} \frac{1}{j\lambda} dx' dy' \quad (16)$$

where $U(x, y; 0)$ is the input source field just after the hologram aperture, $\rho = [(x - x')^2 + (y - y')^2 + z^2]^{1/2}$ and λ is the wavelength. It can be easily observed that the above diffraction integral can be rewritten as a convolution of the input source field with the propagation kernel (i.e. the impulse response of the system) $h(x, y, z)$:

$$U(x, y; z) = U(x, y; 0) * h(x, y; z) \quad (17)$$

where $*$ represents the 2D convolution operation in $x - y$ dimensions. The propagation kernel $h(x, y; z)$ here can be calculated as

$$h(x, y; z) = \frac{1}{j\lambda r} \exp\left(j \frac{2\pi r}{\lambda}\right) \quad (18)$$

where $r = \sqrt{x^2 + y^2 + z^2}$. Here, we can invoke the convolution theorem to express the above equation as a product of Fourier transforms as follows:

$$\mathcal{F}\{U(x, y; z)\} = \mathcal{F}\{U(x, y; 0)\} \mathcal{F}\{h(x, y; z)\} \quad (19)$$

where $\mathcal{F}$ is the Fourier transform operator.

The diffractive field from the SLM, if Fourier-analyzed across any plane, can be identified as plane waves traveling in different directions away from the hologram plane. The field amplitude across any point can be calculated as the summation of contributions of these plane waves, taking into account the phase shifts undergone during the propagation.⁴⁵ This is called the *angular spectrum propagation* of the wave field. The angular spectrum method (ASM) is equivalent to the Rayleigh-Sommerfeld solution and yield identical predictions of the diffracted wave field.⁴⁶ From the equation 19, the ASM propagated field $U(x, y; z)$ from a diffracting aperture with an input source field $U(x, y; 0)$ can be computed as:

$$U(x, y; z) = \mathcal{F}^{-1}(\mathcal{F}U(x, y; 0) \circ H) \quad (20)$$

The ASM transfer function H is given by:

$$H(f_x, f_y; z) = \begin{cases} \exp \left[j2\pi \frac{z}{\lambda} \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2} \right] & , \quad \sqrt{f_x^2 + f_y^2} < \frac{1}{\lambda} \\ 0 & , \quad \text{otherwise} \end{cases} \quad (21)$$

where f_x, f_y are the Fourier frequencies in the x and y direction, respectively. As can be observed from equation 21, the Fourier frequencies of the ASM transfer function are not independent but the frequencies in the z direction are dependent on those in the x and y direction.

Although the above formulation is of a continuous integrals and Fourier transforms, a numerical simulation involves a discrete summation and a discrete Fourier transform (DFT) performed on a regularly sampled grid. One must note that a DFT of an input sampled at regular intervals assumes periodicity of the field in both spatial and Fourier frequency domain, and therefore, the corresponding convolution that is computed using a fast Fourier transform (FFT) is a circular convolution. However, since the input source field and the propagation kernel are aperiodic functions, invoking convolution theorem as in equation 20 causes errors in the edges of the computed wave field $U(x, y; z)$.

From basic signal processing we know that a circular convolution is nothing but a linear convolution with aliasing. One can easily visualize that if the diffracted wave field on the output (image) plane is small enough compared to the overall computation window, the aliasing errors induced due to the circular convolution may be to an acceptable level. However, if the diffracted field overflows the size of the computation window at the output (image) plane, the aliasing errors become pronounced and the results would deviate from the accurate Rayleigh-Sommerfeld integral.

To overcome such errors we zero-pad the input image to double the size along both the dimensions, effectively increasing the sampling window. After computing the diffraction integral, the computation window on the output plane is cropped to the original image size. Although one can zero-pad the image in many different ways, we symmetrically increase the size of the window on both dimensions.

The angular spectrum method, after zero-padding, results in reduced artifacts due to aliasing. Although accurate results can be obtained using ASM in near distances, the accuracy of the method decreases for distances that are greater than the sampling window size by a factor of about ten or more. This is caused primarily due to insufficient sampling and these errors can be avoided by extending the sampling window and computing the whole field.⁴⁷ Although a viable option, this necessitates larger computational window for longer propagation distances, thereby resulting in a huge computational load.

For an SLM with N_x pixels of pixel pitch Δ_x in the x direction, the size of the original computation window is $L_x = N_x \Delta_x$. However, the source sampling area is doubled to linearize the circular convolution thereby increasing the computation window size to $2L_x$. Therefore, the sampling interval of the transfer function can be computed as $\Delta f_x = \frac{1}{2L_x}$. Now, according to the Nyquist theorem, aliasing errors in the above sampled ASM transfer function can be avoided if the sampling frequency is at least twice the local signal frequency of the transfer function $H(f_x, f_y; z)$, as discussed in Ref. 45:

$$\Delta f_x^{-1} \geq 2 \left| \frac{1}{2\pi} \frac{\partial \phi}{\partial f_x} \right| \quad (22)$$

where $\phi = \angle H(f_x, f_y; z) = 2\pi \frac{z}{\lambda} \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2}$ is the phase of the ASM transfer function.

To avoid aliasing errors, one can limit the frequencies in the transfer function to satisfy the Nyquist sampling requirement, resulting in what is called a band-limited transfer function.⁴⁷ Note that although the sampling intervals can be flexible and chosen freely, for a digital holographic display, the sampling interval is generally fixed and is equal to that of the SLM pixel pitch. Hence, the permissible range of frequencies for the spatial frequency f_x (and similarly for f_y) for which the angular spectrum transfer function does not result in aliasing is also fixed for a given propagation distance and can be obtained as follows:

$$|f_x| \leq \frac{1}{\left[1 + (2\Delta f_x z)^2\right]^{\frac{1}{2}} \lambda} \equiv f_{x_{limit}} \quad (23)$$

Therefore, we band-limit the angular spectrum by clipping the transfer function within a bandwidth of $2f_{x_{limit}}$ to avoid numerical errors arising due to the sampling and aliasing. The updated band-limited transfer function can therefore be obtained as

$$H_{BL}(f_x, f_y; z) = H(f_x, f_y; z) \text{rect}\left(\frac{f_x}{2f_{x_{limit}}}\right) \text{rect}\left(\frac{f_y}{2f_{y_{limit}}}\right) \quad (24)$$

Note that if the limiting frequency of the band-limited transfer function, i.e. $f_{x_{limit}}$ and $f_{y_{limit}}$ is less than the maximum spatial frequency required for the reconstruction of the holographic image, a part of the diffraction field loses the frequency band necessary for accurate diffraction and this would result in errors in the reconstruction. However, limiting the bandwidth in the above mentioned fashion always satisfies the minimum spatial frequency requirement.

4.2 Phase hologram optimization

For a given phase-only hologram with phase pattern Φ , the diffraction field on the output image plane can be computed using the above band-limited angular spectrum transfer function as:

$$z = \mathcal{F}^{-1}(\mathcal{F}U \circ H_{BL}) \quad (25)$$

where $U = e^{j\Phi}$ is the complex phase hologram and H_{BL} is the band-limited ASM transfer function. Ideally, we want to compute phase patterns that result in reconstructions as close to the target images as possible. In other words, we want the error between the reconstructed image intensity $|z|^2$ produced by the hologram with phase pattern Φ and the target image I to be minimum. This allows us to pose the holographic phase retrieval problem as the following optimization problem:

$$\begin{aligned} \Phi_{opt} &= \min_{\Phi} \underbrace{f(|z|^2, I)}_{Err(\Phi)} \\ &= \min_{\Phi} f(|\mathcal{F}^{-1}(\mathcal{F}U \circ H_{BL})|^2, I) \end{aligned} \quad (26)$$

where f is a penalty function to compute the error between the target and reconstructed images. The above optimization problem can be solved using the Wirtinger gradient framework as described in 4. We briefly discuss here the Wirtinger gradients for our band-limited angular spectrum propagation method. We refer the curious reader to Chakravarthula et al. 4 for more details. Using these gradients we can solve the above optimization problem in equation 26 using first-order optimization methods.

4.2.1 Gradient for ASM propagation

We refer the reader to the discussion in our previous paper, Ref.,⁴ for a more detailed discussion on the Wirtinger gradients. However, we briefly discuss it here. In order to update the hologram phase patterns using a gradient descent optimization technique, we require the gradient of the error function in equation 26 with respect to the phase pattern Φ . This can be calculated applying the chain rule to equations 25 and 26 as follows:

$$\frac{dErr}{d\Phi} = \underbrace{\frac{df}{dz}}_1 \underbrace{\frac{dz}{dU}}_2 \underbrace{\frac{dU}{d\Phi}}_3 \quad (27)$$

Notice that Part-1 of the equation 27 requires us to compute the derivative of the scalar real-valued error f with respect to the complex diffractive wave field z . As discussed in Section 3, the derivative of this part is undefined. To overcome the undefined gradient, we approximate the partial of the scalar function of the complex vector as:

$$d(Err) = df(z) = \text{Re}\langle \nabla f, dz \rangle \quad (28)$$

where Re denotes the real part of a complex number and $\langle \cdot, \cdot \rangle$ denotes the inner product of two vectors. Note that the above definition is not the exact gradient but only an approximate definition to use with any first-order optimization techniques.⁴ The value of ∇f in the above definition is obtained using the complex Wirtinger derivatives:

$$\nabla f(z) = 2\nabla_{\bar{z}}f \quad (29)$$

which can be further simplified as:

$$\nabla_{\bar{z}}f = \frac{df}{d(|z|^2)} \circ 2\nabla_{\bar{z}}(|z|^2) \quad (30)$$

Observe that the second part of the above equation 30 can be reduced to

$$2\nabla_{\bar{z}}(|z|^2) = 2\nabla_{\bar{z}}(z\bar{z}) = 2z \quad (31)$$

Therefore, the part 1 of the gradient equation 27 is evaluated as follows:

$$\frac{df}{dz} = \nabla f = \frac{df}{d(|z|^2)} \circ 2z \quad (32)$$

As discussed in previous section, the wave field at the destination plane can be obtained using band-limited angular spectrum propagation as $z = \mathcal{F}^{-1}(\mathcal{F}U \circ H)$, and the image is computed as $|z|^2$. Computing Part 2 of the equation 27 using the Part 1 gradient computed in equation 32 yields:

$$\begin{aligned} d(Err(H)) &= \text{Re}\langle \nabla f, dz \rangle \\ &= \text{Re}\langle \nabla f, d(F^\dagger(H)(FU_1)) \rangle \\ &= \text{Re}\langle \nabla f, F^\dagger H F dU_1 \rangle \\ &= \text{Re}\langle F^\dagger H^\dagger F \nabla f, dH \rangle \\ &= \text{Re}\langle F^\dagger H^* F \nabla f, dH \rangle \end{aligned} \quad (33)$$

Finally evaluating Part 3 with the complex amplitude on the hologram plane as $U = \mathbf{exp}^{j\Phi}$, we derive the definition of gradient of the loss function with respect to the phase Φ as follows:

$$\begin{aligned} d(Err(\Phi)) &= \text{Re}\langle F^\dagger H^* F \nabla f, d(e^{j\Phi}) \rangle \\ &= \text{Re}\langle -je^{-j\Phi} F^\dagger H^* F \nabla f, d(\Phi) \rangle \end{aligned} \quad (34)$$

Since the phase Φ is real valued, the above inner-product definition can be read as:

$$d(Err(\Phi)) = \langle \text{Re}(-je^{-j\Phi} F^\dagger H^* F \nabla f), d(\Phi) \rangle, \nabla Err(\Phi) = \text{Re}(-je^{-j\Phi} F^\dagger H^* F \nabla f). \quad (35)$$

With the above gradient in hand, the phase patterns can be optimized for using standard first-order stochastic gradient descent solvers.

5. IMPLEMENTATION AND ANALYSIS

Our method eliminates severe artifacts that are caused by existing hologram generation methods. In order to demonstrate the performance of our method and experimentally validate our improved algorithm, we build a prototype holographic display system whose details are discussed in this section.

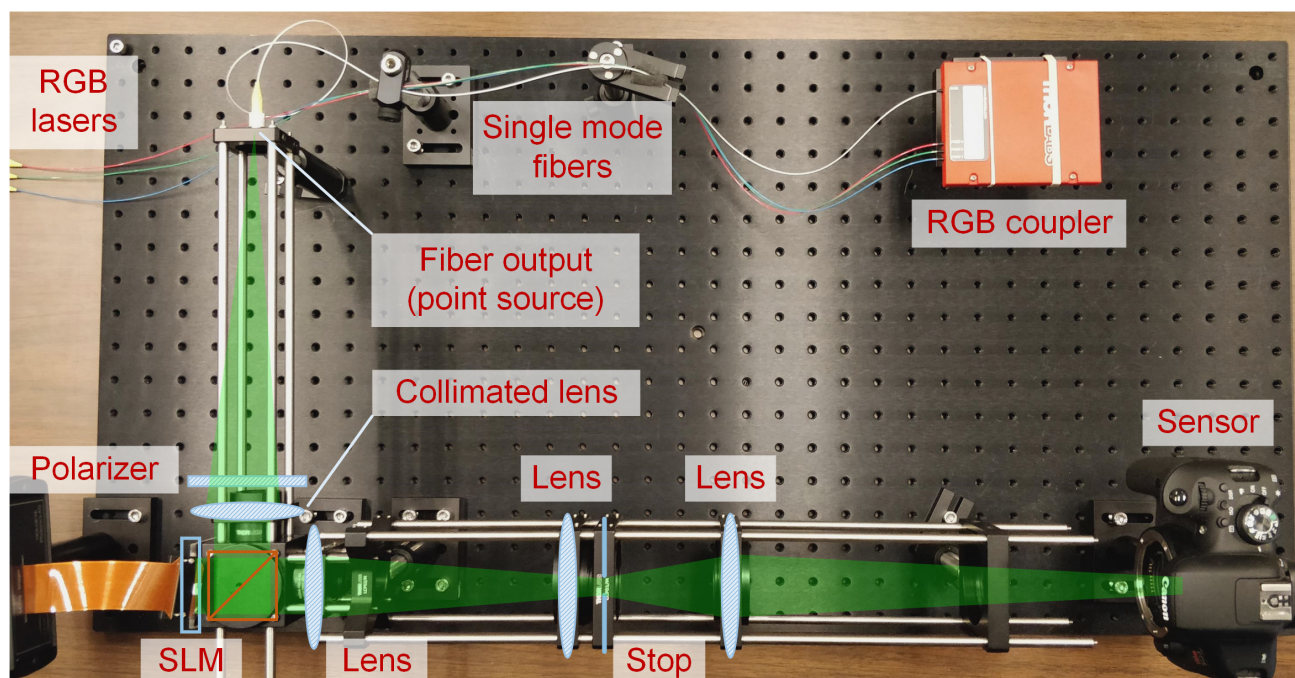


Figure 2: Schematic of our prototype holographic near-eye display. Specifically, we couple the lasers of three different single wavelengths into one single-mode fiber whose output end can be treated as point source. The light emitted from this end is polarized and collimated to incident on the SLM. We use a phase-only reflective type SLM to load and display holograms. The modulated wave is imaged and filtered by a couple of lenses, and eventually recorded as an intensity distribution on the sensor.

5.1 Hardware Implementation

The prototype display system, shown in figure 2 is same as the one demonstrated in Ref. 4. Our prototype display includes a HOLOEYE LETO-1 liquid crystal on silicon (LCoS) reflective phase-only spatial light modulator with a resolution of 1920×1080 . The pixel pitch of the SLM is $6.4 \mu\text{m}$, resulting in the active area of $12.28 \text{ mm} \times 6.91 \text{ mm}$. We use a single optical fiber that is coupled to three laser diodes emitting at wavelengths 446, 517, and 636 nm, in combination with collimating optics, to illuminate the SLM with a plane wave. The SLM is illuminated by the laser source in a color field sequential manner. To assess the display's image quality, we capture images from the prototype holographic display using a Canon Rebel T6i DSLR camera body (without camera lens attached) with an output resolution of $6,000 \times 4,000$, a pixel pitch of $3.72 \mu\text{m}$.

5.2 Software

We compute phase-only holograms on a PC with an Intel Xeon 2.4GHz CPU and 64GB of RAM, and an NVIDIA GeForce GTX 1080 GPU. We implement the proposed method using quasi-Newton optimization in MATLAB using the L-BFGS⁴⁸ quasi-Newton method. In particular, we use the minFunc optimization package.⁴⁹ For loss functions requiring stochastic gradient optimization, we implement the proposed method in TensorFlow⁵⁰ and use the Adam optimizer. Specifically, we use a learning rate of 0.001 and an exponential decay rate of 0.9 for the first moment estimate, and an exponential decay rate of 0.999 for the second moment estimates. The gradients obtained from backpropagation offered by TensorFlow are used for the loss function component and those parameterized by convolutional neural network to compute the complex Wirtinger gradient and feed it back to the optimizer for computing the optimal phase holograms. The full color holograms are generated by performing optimization sequentially for all three color channels.

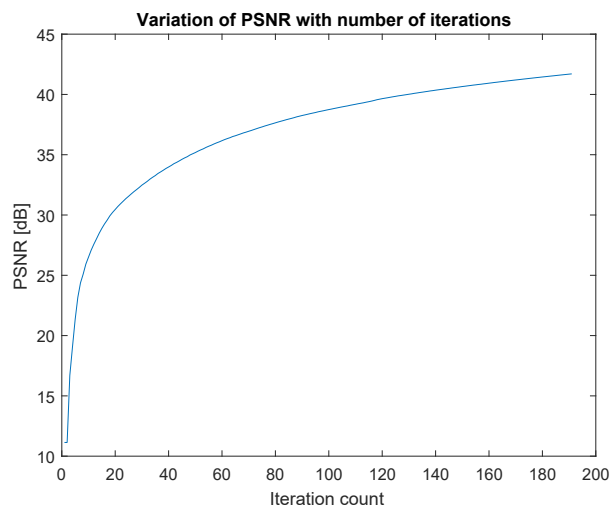


Figure 3: PSNR values plotted against number of iterations for Wirtinger holography optimization for a randomly initialized phase.

5.3 Initialization and termination criteria

We start the optimization process with an initial guess of the phase value ϕ , which could either be a random phase or a carefully crafted phase. Carefully initializing the optimizer results in faster convergence compared to random (or shallow random) initialization. It can be easily understood that the optimizer continuously searches for the optimal phase hologram for a given target intensity (image). However, one must note that there can be several phase holograms that can produce a given target intensity, making our objective function an extremely non-convex problem with several local minima. Therefore, initializing with a random phase distribution, intuitively, expands the search space for determining an optimal phase hologram. Owing to a larger domain of possible phase holograms, the particular choice of initial phase value may affect the convergence speed and numerical stability.

Let us consider one pixel on the hologram (SLM) plane. The complex amplitude of the hologram pixel constitutes a unit amplitude and a phase value lying in the range $[0, 2\pi]$. In other words, the complex amplitude of the hologram pixel can lie anywhere on a unit circle. Suppose if the phase is known to lie within a certain range $[\theta_{min}, \theta_{max}]$, the correct complex number determining the complex amplitude of the hologram pixel must lie on an arc of the unit circle which is defined by the angles θ_{min} and θ_{max} .

In our case, the possibility of the range of phase values for each hologram pixel remains unclear. However, given the knowledge of some initial phase value producing the target intensity pattern, it is reasonable to assume the optimal phase value might lie somewhere close to the initial phase value. Initializing the optimizer with a carefully crafted phase, for example, a few iterations of Gerchberg-Saxton algorithm or a double-phase encoding, thus offers the promise of finding optimal phase faster. Figure 3 shows the PSNR values with iteration count for our framework for a randomly initialized phase, and it can be observed that the PSNR improves very quickly initially, and then slowly increases with an increasing number of iterations. For the results reported, we terminate the optimizer once the PSNR values reach convergence.

5.4 Analysis

Our holography method eliminates much of the ringing artifacts and generates high quality holographic reconstructions in comparison to the recently demonstrated modified GS⁴⁴ and double phase encoding² phase hologram generation algorithms. Figure 4 compares in simulation the state-of-the-art double phase encoding phase holograms and Wirtinger holograms, and present SSIM error maps as an indication of the probability of perceptually detecting ringing errors in the reconstructed image. Many other results of our method are reported in our previous paper (Ref. 4) and recommend the readers to also refer to it along with this paper.

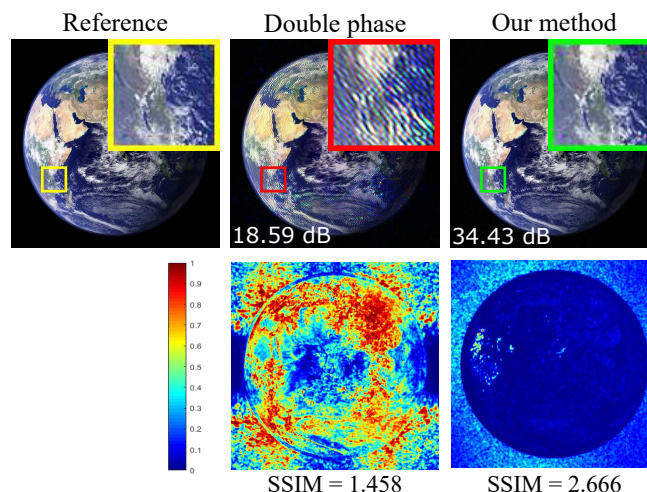


Figure 4: Probability of detection of errors for the simulated image also shown in the main manuscript. Note that the Wirtinger holograms produce much less perceptual error as compared to that of double phase encoding method.

We also show in simulation the results for computing high-quality holograms for 3D scenes. Computing 3D holograms is typically a difficult problem which often results in unwanted noise and reduction in contrast. Computing a 3D hologram in a multi-plane fashion using our abovementioned method is possible. However, if the depth planes are densely spaced, the optimizer forces light to change its phase rapidly while propagating in ways which are sometimes physically not feasible. Therefore the optimal phase hologram that is generated might contain noise which can degrade the image quality in comparison to 2D holograms. Although all 3D holographic rendering methods suffer from these problems, the holograms computed using full Fresnel diffraction integral without paraxial approximation and heuristic encoding schemes might produce fewer artifacts than optimizing for dense focal stack of 3D scene data. However, since the eye can only focus at one focal plane at a time, one can render holographic imagery at the eye accommodation plane by changing the scene focus globally. This needs to be accompanied by an eye tracker. Note, that although the depth detection capability progressively degrades with retinal eccentricity,⁷ object points which are closely spaced in the region where eye is foveated suffer if not presented with an accurate depth of field blur.² However, such regions can be addressed by providing the depth of field blur in the image space.⁵¹ Figure 5 shows simulated hologram results generated by our framework for a 3D scene by the method of dynamic global scene focus. Notice that the depth of field blur is appropriately represented without compromising on the image quality. Jointly optimizing for all depth planes to generate an artifact-free full 3D hologram remains an exciting direction for future work.

We also validate the method experimentally on the real holographic display as described above. Figure 6 present several holographic images reconstructed on our prototype holographic display. It can be seen that our Wirtinger holograms feature more detail, higher resolution, and higher light efficiency compared to other existing methods.

6. CONCLUSION

We have presented a phase retrieval method for holographic neareye displays, which departs from existing heuristic algorithms, and instead relies on formal optimization using complex Wirtinger derivatives. This framework has allowed us to formulate holographic phase retrieval as a quadratic optimization problem that can be solved using first-order optimization methods. We have validated that the resulting holograms have an order of magnitude lower error, i.e. more than 10 dB PSNR improvement, when compared to recent state-of-the-art approaches. We have assessed the proposed method in simulation and using an experimental prototype system. Wirtinger holography eliminates the severe artefacts present in existing methods, and, in particular, enables us to minimize ringing and chromatic artefacts. In addition to reducing error by an order of magnitude when compared to

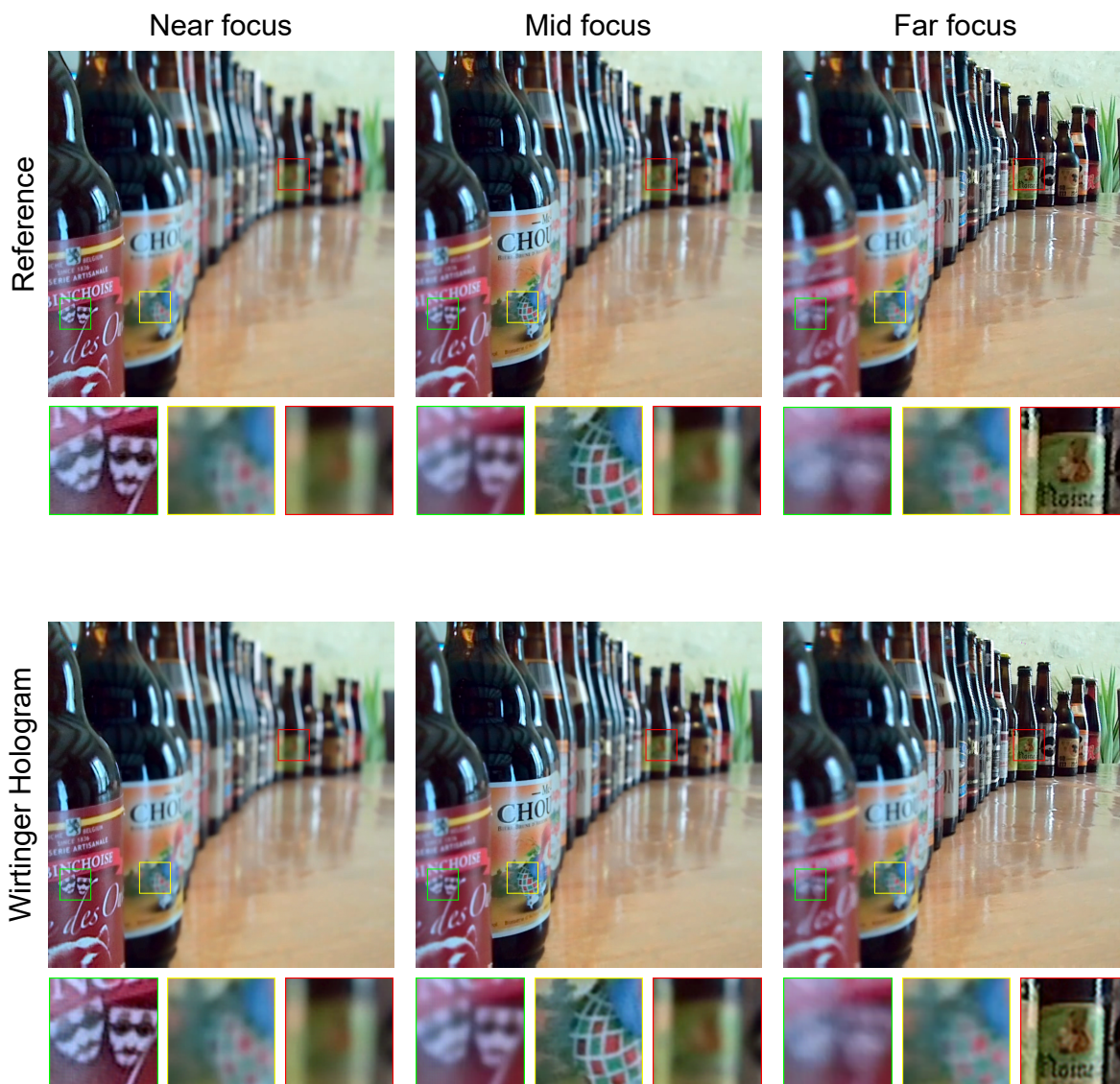


Figure 5: Simulated holograms for a 3D scene generated using fast depth switching and dynamic global scene focus for an assumed eye tracked position. Observe that the depth of field effects are appropriately represented without compromising on the holographic image quality. The depth of the above scene is scaled to make the near distance at 200mm and far distance at 500mm.

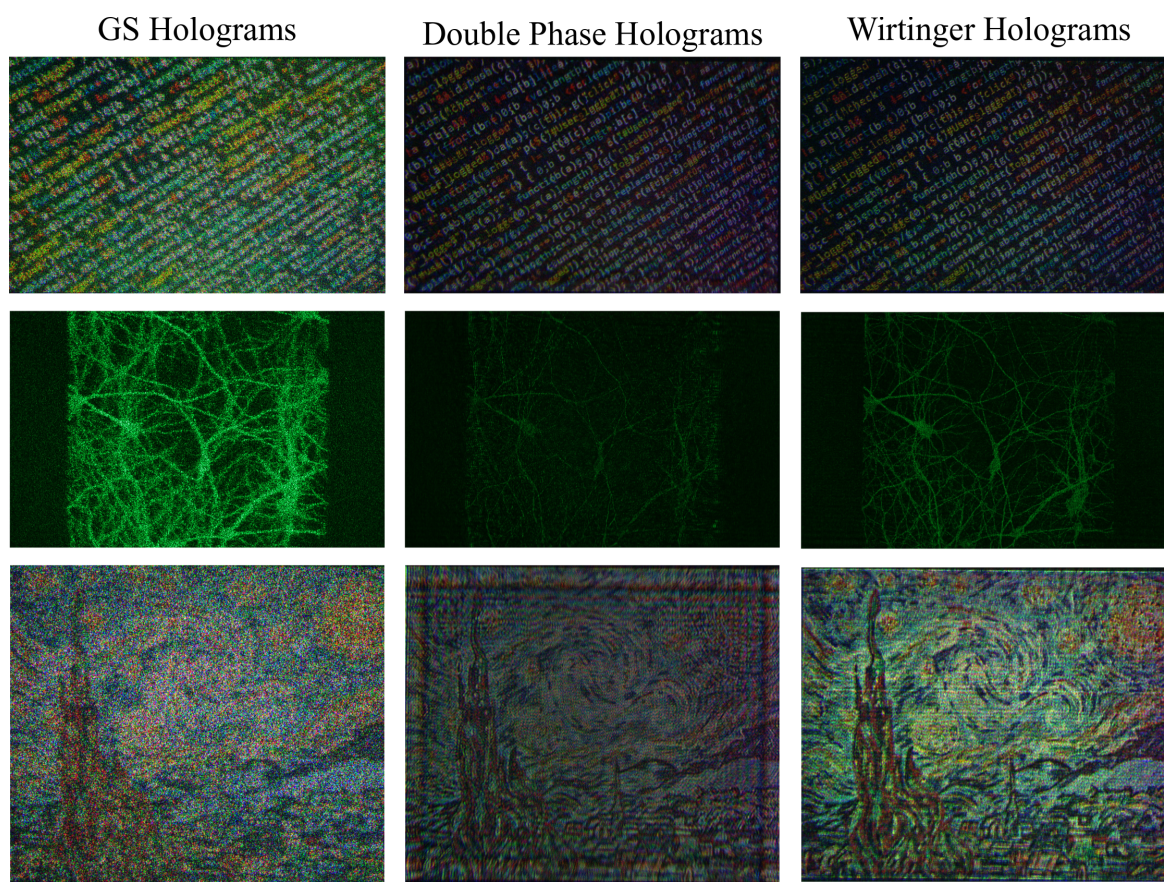


Figure 6: Results reconstructed by modified GS holograms, Double phase encoding of Fresnel holograms, and Wirtinger Holograms.

existing methods, we have validated the flexibility of the proposed system. We are excited to believe that our method enables future research towards the vision of artifact-free high-quality holography.

ACKNOWLEDGMENTS

The authors wish to thank Bernard Kress for lending the HOLOEYE LETO-I SLM, Roarke Horstmeyer for many fruitful discussions and also lending the laser diode controller, Andreas Georgiou and Nicolas Pegard for useful suggestions and Pavan Chandra Konda and Jim Mahaney for help with the hardware prototype and experimental captures. This work was partially supported by National Science Foundation (NSF) Grant # 1405847, by NSF Grant # 1840131, an unrestricted gift from Google, and by the BeingTogether Centre, a collaboration between Nanyang Technological University (NTU) Singapore and University of North Carolina (UNC) at Chapel Hill, supported by UNC and the Singapore National Research Foundation, Prime Minister's Office, Singapore under its International Research Centres in Singapore Funding Initiative.

REFERENCES

- [1] Chakravarthula, P., Dunn, D., Akşit, K., and Fuchs, H., "Focusar: Auto-focus augmented reality eyeglasses for both real world and virtual imagery," *IEEE transactions on visualization and computer graphics* **24**(11), 2906–2916 (2018).
- [2] Maimone, A., Georgiou, A., and Kollin, J. S., "Holographic near-eye displays for virtual and augmented reality," *ACM Transactions on Graphics (TOG)* **36**(4), 85 (2017).

- [3] Shi, L., Huang, F.-C., Lopes, W., Matusik, W., and Luebke, D., "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), 236 (2017).
- [4] Chakravarthula, P., Peng, Y., Kollin, J., Fuchs, H., and Heide, F., "Wirtinger holography for near-eye displays," *ACM Transactions on Graphics (TOG)* **38**(6), 1–13 (2019).
- [5] Underkoffler, J. S., *Toward accurate computation of optically reconstructed holograms*, PhD thesis, Massachusetts Institute of Technology (1991).
- [6] Bates, R., "Fourier phase problems are uniquely solvable in more than one dimension. i: Underlying theory," *Optik (Stuttgart)* **61**, 247–262 (1982).
- [7] Lesem, L., Hirsch, P., and Jordan, J., "The kinoform: a new wavefront reconstruction device," *IBM Journal of Research and Development* **13**(2), 150–155 (1969).
- [8] Gerchberg, R. W., "A practical algorithm for the determination of the phase from image and diffraction plane pictures," *Optik* **35**, 237–246 (1972).
- [9] Fienup, J. R., "Phase retrieval algorithms: a comparison," *Applied optics* **21**(15), 2758–2769 (1982).
- [10] Luke, D. R., "Relaxed averaged alternating reflections for diffraction imaging," *Inverse problems* **21**(1), 37 (2004).
- [11] Bauschke, H. H., Combettes, P. L., and Luke, D. R., "Hybrid projection–reflection method for phase retrieval," *JOSA A* **20**(6), 1025–1034 (2003).
- [12] Gonsalves, R., "Phase retrieval from modulus data," *JOSA* **66**(9), 961–964 (1976).
- [13] Lane, R., "Phase retrieval using conjugate gradient minimization," *Journal of Modern Optics* **38**(9), 1797–1813 (1991).
- [14] Fienup, J. R., "Phase-retrieval algorithms for a complicated optical system," *Applied optics* **32**(10), 1737–1746 (1993).
- [15] Wen, Z., Yang, C., Liu, X., and Marchesini, S., "Alternating direction methods for classical and ptychographic phase retrieval," *Inverse Problems* **28**(11), 115010 (2012).
- [16] Marchesini, S., Tu, Y.-C., and Wu, H.-t., "Alternating projection, ptychographic imaging and phase synchronization," *Applied and Computational Harmonic Analysis* **41**(3), 815–851 (2016).
- [17] Zhang, J., Pégard, N., Zhong, J., Adesnik, H., and Waller, L., "3d computer-generated holography by non-convex optimization," *Optica* **4**(10), 1306–1313 (2017).
- [18] Candes, E. J., Strohmer, T., and Vershynin, V., "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics* **66**(8), 1241–1274 (2013).
- [19] Goldstein, T. and Studer, C., "Phasemax: Convex phase retrieval via basis pursuit," *IEEE Transactions on Information Theory* (2018).
- [20] Bahmani, S. and Romberg, J., "Phase Retrieval Meets Statistical Learning Theory: A Flexible Convex Relaxation," in [*Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*], Singh, A. and Zhu, J., eds., *Proceedings of Machine Learning Research* **54**, 252–260, PMLR, Fort Lauderdale, FL, USA (20–22 Apr 2017).
- [21] Benton, S. A. and Bove Jr, V. M., [*Holographic imaging*], John Wiley & Sons (2008).
- [22] Waters, J. P., "Holographic image synthesis utilizing theoretical methods," *Applied physics letters* **9**(11), 405–407 (1966).
- [23] Lucente, M. E., "Interactive computation of holograms using a look-up table," *Journal of Electronic Imaging* **2**(1), 28–34 (1993).
- [24] Petz, C. and Magnor, M., "Fast hologram synthesis for 3d geometry models using graphics hardware," in [*Proc. SPIE*], **5005**, 266–275 (2003).
- [25] Masuda, N., Ito, T., Tanaka, T., Shiraki, A., and Sugie, T., "Computer generated holography using a graphics processing unit," *Optics Express* **14**(2), 603–608 (2006).
- [26] Chen, R. H.-Y. and Wilkinson, T. D., "Computer generated hologram from point cloud using graphics processor," *Applied optics* **48**(36), 6841–6850 (2009).
- [27] Leseberg, D. and Frère, C., "Computer-generated holograms of 3-d objects composed of tilted planar segments," *Applied optics* **27**(14), 3020–3024 (1988).

- [28] Tommasi, T. and Bianco, B., "Computer-generated holograms of tilted planes by a spatial frequency approach," *JOSA A* **10**(2), 299–305 (1993).
- [29] Matsushima, K., "Computer-generated holograms for three-dimensional surface objects with shade and texture," *Applied optics* **44**(22), 4607–4614 (2005).
- [30] Ahrenberg, L., Benzie, P., Magnor, M., and Watson, J., "Computer generated holograms from three dimensional meshes using an analytic light transport model," *Applied optics* **47**(10), 1567–1574 (2008).
- [31] Kim, H., Hahn, J., and Lee, B., "Mathematical modeling of triangle-mesh-modeled three-dimensional surface objects for digital holography," *Applied optics* **47**(19), D117–D127 (2008).
- [32] Matsushima, K. and Nakahara, S., "Extremely high-definition full-parallax computer-generated hologram created by the polygon-based method," *Applied optics* **48**(34), H54–H63 (2009).
- [33] Matsushima, K., Nakamura, M., and Nakahara, S., "Silhouette method for hidden surface removal in computer holography and its acceleration using the switch-back technique," *Optics express* **22**(20), 24450–24465 (2014).
- [34] Jia, J., Liu, J., Jin, G., and Wang, Y., "Fast and effective occlusion culling for 3d holographic displays by inverse orthographic projection with low angular sampling," *Applied optics* **53**(27), 6287–6293 (2014).
- [35] Lucente, M. and Galyean, T. A., "Rendering interactive holographic images," in [*Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*], 387–394, ACM (1995).
- [36] Yamaguchi, M., Hoshino, H., Honda, T., and Ohyama, N., "Phase-added stereogram: calculation of hologram using computer graphics technique," in [*Proc. SPIE*], **1914**, 25–31 (1993).
- [37] Smithwick, Q. Y., Barabas, J., Smalley, D. E., and Bove, V. M., "Interactive holographic stereograms with accommodation cues," in [*Practical Holography XXIV: Materials and Applications*], **7619**, 761903, International Society for Optics and Photonics (2010).
- [38] Zhang, H., Zhao, Y., Cao, L., and Jin, G., "Fully computed holographic stereogram based algorithm for computer-generated holograms with accurate depth cues," *Optics express* **23**(4), 3901–3913 (2015).
- [39] Bayraktar, M. and Özcan, M., "Method to calculate the far field of three-dimensional objects for computer-generated holography," *Applied optics* **49**(24), 4647–4654 (2010).
- [40] Zhao, Y., Cao, L., Zhang, H., Kong, D., and Jin, G., "Accurate calculation of computer-generated holograms using angular-spectrum layer-oriented method," *Optics express* **23**(20), 25440–25449 (2015).
- [41] Chen, J.-S. and Chu, D., "Improved layer-based method for rapid hologram generation and real-time interactive holographic display applications," *Optics express* **23**(14), 18143–18155 (2015).
- [42] Zhang, H., Zhao, Y., Cao, L., and Jin, G., "Layered holographic stereogram based on inverse fresnel diffraction," *Applied optics* **55**(3), A154–A159 (2016).
- [43] Remmert, R., [*Theory of complex functions*], vol. 122, Springer Science & Business Media (2012).
- [44] Peng, Y., Dun, X., Sun, Q., and Heidrich, W., "Mix-and-match holography," *ACM Transactions on Graphics* (2017).
- [45] Goodman, J. W., [*Introduction to Fourier optics*], Roberts and Company Publishers (2005).
- [46] Shen, F. and Wang, A., "Fast-fourier-transform based numerical integration method for the rayleigh-sommerfeld diffraction formula," *Applied optics* **45**(6), 1102–1110 (2006).
- [47] Matsushima, K. and Shimobaba, T., "Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields," *Optics express* **17**(22), 19662–19673 (2009).
- [48] Liu, D. C. and Nocedal, J., "On the limited memory bfgs method for large scale optimization," *Mathematical programming* **45**(1-3), 503–528 (1989).
- [49] Schmidt, M., "minfunc: unconstrained differentiable multivariate optimization in matlab," *Software available at <http://www.cs.ubc.ca/~schmidt/Software/minFunc.htm>* (2005).
- [50] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al., "Tensorflow: A system for large-scale machine learning," in [*12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*], 265–283 (2016).
- [51] Cholewiak, S. A., Love, G. D., Srinivasan, P. P., Ng, R., and Banks, M. S., "Chromablur: Rendering chromatic eye aberration improves accommodation and realism," *ACM Transactions on Graphics (TOG)* **36**(6), 210 (2017).