This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3023601, IEEE Transactions on Visualization and Computer Graphics

Correcting the Proximity Effect in Nanophotonic Phased Arrays

Xuetong Sun* Yang Zhang† Po-Chun Huang ‡ Niloy Acharjee§ Mario Dagenais ¶
Martin Peckerar ‖ Amitabh Varshney **

University of Maryland, College Park

(a) Without PEC (b) Basic Iterative PEC Gemayel *et al.* [13] (c) Iterative PEC (d) Proximal PEC

Fig. 1. Nanophotonic phased arrays suffer from thermal proximity effect where one pixel being heated affects the temperature of nearby pixels. This causes inaccurate phase modulation and noise in the formed image as shown in (a). The current state-of-the-art methods, as shown in (b), are not able to sufficiently correct the proximity effect for our use case. We propose two proximity effect correction (PEC) methods which are able to reduce the noise, as shown in (c) and (d). The proximal PEC method has the best correction effectiveness. Images are chosen from the Common Objects in Context dataset [30].

Abstract—Thermally modulated Nanophotonic Phased Arrays (NPAs) can be used as phase-only holographic displays. Compared to the holographic displays based on Liquid Crystal on Silicon Spatial Light Modulators (LCoS SLMs), NPAs have the advantage of integrated light source and high refresh rate. However, the formation of the desired wavefront requires accurate modulation of the phase which is distorted by the thermal proximity effect. This problem has been largely overlooked and existing approaches to similar problems are either slow or do not provide a good result in the setting of NPAs. We propose two new algorithms based on the iterative phase retrieval algorithm and the proximal algorithm to address this challenge. We have carried out computational simulations to compare and contrast various algorithms in terms of image quality and computational efficiency. This work is going to benefit the research on NPAs and enable the use of large-scale NPAs as holographic displays.

Index Terms—Nanophotonic phased array, proximity effect correction, proximal algorithms, phase-only hologram.

1 Introduction

Nanophotonic phased arrays (NPA) are phased arrays operating at optical wavelengths. An NPA is composed of an optical power distribution system, a phase modulation mechanism that can control the phase of each pixel individually, and antennas for propagating phase-modulated light into free space. NPAs can efficiently steer optical waves in the desired direction and do so efficiently without any moving parts. Thus, they have found widespread applications in optical communications (such as Khalighi and Uysal [20]), range-finding (such as Poulton *et al.* [54]), and other fields. Recently, Sun *et al.* [63] and Notaros *et al.* [44] demonstrated the use of the NPA as an augmented reality holographic display. In this application, the NPA generates an optical wavefront. Through careful phase modulation, text and images (or 3D scenes) can be displayed for the observer to see. This use of NPAs is of particular interest for graphics research into near-eye displays (NEDs).

Near-eye displays have become more popular with virtual and augmented reality (VR and AR) going mainstream. In the current VR and AR headsets, users' eyes can only focus at one or a small number of depths. This vergence-accommodation conflict (VAC) can lead to visual discomfort, nausea, and cybersickness. Focus-tunable displays have been shown to provide a preferable viewing experience [23]. Varifocal displays and light field displays have been proposed to address this issue but they are either limited in angular resolution or involve multiple moving components. Holographic displays recreate the optical wave field of 2D images or 3D scenes. Therefore, the light observed by the user looks as if it is coming from the real image or object, allowing viewing from different perspectives. Holographic displays also form images and 3D scenes at the desired depths for eyes to naturally focus

*e-mail: xtsun@umd.edu
†@yzhangdd@terpmail.umd.edu
‡hpcalex@umd.edu
§acharjee@terpmail.umd.edu
¶dage@umd.edu
‖peckerar@umd.edu
**varshney@cs.umd.edu

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

on, thereby addressing VAC.

A holographic display modulates the phase and/or amplitude of light at each pixel location to form the desired wavefront to be observed. Previously, Liquid Crystal on Silicon Spatial Light Modulators (LCoS SLMs) are the most common phase modulators for holographic displays. However, NPAs have gained popularity recently. One major advantage of the NPAs over SLMs is their fast refresh rate. One refresh cycle on the NPA includes heating each pixel to a certain temperature corresponding to the desired phase and cooling down to ambient temperature. The simulation with our NPA design shows that it can complete a refresh cycle within $10\,\mu s$, suggesting a 100kHz refresh rate which is uniquely suitable for dynamic content as well as time-division multiplexing operations such as color-switching and higher dynamic range. The only other microdisplay capable of kHz-refresh rates is the Digital Micromirror Device (DMD). It has been used for creating near-eye displays in Rathinavel *et al.* [55]. However, the DMD can only modulate binary amplitude patterns which limits the fidelity of its image in a single frame. The NPA is capable of continuous phase modulation. In addition, the NPAs support integrated light sources where a laser is coupled to the NPA via an optical fiber, distributed to each pixel and emitted by the on-board antennas. In this way, the NPA displays can be made more compact as optical operations such as beam expansion and polarization necessary for typical LCoS SLM or DMD display systems are not required. LCoS SLMs also suffer from pixel voltage crosstalk and non-linearity mapping between phase and voltage which hinder accurate phase modulation.

Thermally-modulated NPAs, such as those in Doylend *et al.* [9], Sun *et al.* [63] and our design, depend on the thermo-optical effect of the materials to modulate the phase of a pixel. The phase of a pixel can be changed from 0 to 2π by changing the temperature of the pixel. The precise and independent control of the phase is needed to ensure the successful display of desired 2D or 3D imagery. However, the thermal proximity effect (PE), the phenomenon where the temperature of one pixel affects the temperature of neighboring pixels, impacts the accuracy of phase modulation and degrades the observed image. A simulated example of the proximity effect negatively impacting the formed image can be seen in Figure 1. Therefore, in order to form the desirable images, this thermal proximity effect must be taken into consideration. This problem is further complicated by the fact that most NPAs only have heat generation mechanisms, with no cooling capability, making some phase patterns impossible to realize.

There has been limited study on the problem of thermal proximity effect on NPA holographic displays. But the proximity effect is present in many processes and devices. In beam electron lithography and optical communication using SLMs, the proximity effect is modeled as a convolution, such as in Peckerar *et al.* [49] and Persson *et al.* [53]. In the former case, the correction is conducted as a deconvolution with non-negative constraints on the solution. In the latter case, the Gerchberg-Saxton phase retrieval algorithm is slightly modified to produce a result that is somewhat resistant to the proximity effect. Sun *et al.* [64] also consider the thermal proximity effect on NPAs as a convolution and perform deconvolution on the phase under the assumption of perfect amplitude modulation.

In this study, we consider the practical but more challenging case of phase-only holograms on NPAs. Based on the thermo-electric simulations, we also model the proximity effect as a convolution. We propose two new methods for proximity effect correction. The first method works by integrating deconvolution into the phase retrieval algorithm. The second method uses the proximal algorithm to directly optimize the displayed image on the NPA considering the proximity effect.

We start by investigating how the proximity effect negatively affects the observed image quality on a phase-only NPA holographic display. We implement the two new proximity effect correction methods mentioned above and compare their correction effectiveness with existing solutions. Computational simulations are run with proximity effect of varying severity and both qualitative and quantitative results are presented. The factors affecting the correction effectiveness and processing time of the methods are also investigated.

In summary, the goal of this study is to explore the proximity effect on NPA holographic displays and propose new methods of correction[1]. To the best of our knowledge, this is the first detailed analysis of phase-only NPA-based holographic displays with the thermal proximity effect and the first to come up with effective solutions.

The rest of the paper is organized as follows. In Section 2 we briefly summarize the existing literature on several related topics including near-eye displays, NPAs, computer-generated holography, and proximity correction. In Section 3, we describe how we model the proximity effect on an NPA display. Then in Section 4, we describe our proposed proximity effect correction methods. In Section 5, we show through simulation the impact of proximity effect on the reconstructed far-field image quality and the results of our proximity effect correction methods. Discussions and limitations of our methods are presented in Sections 6 and 7.

2 Related Work

We introduce some of the existing work on the various topics related to this paper. This includes work on NPAs, NEDs, computer-generated holography and the proximity effect correction.

2.1 Nanophotonic Phased Arrays

The phased array has been an active area of research since its invention. The microwave version of the phased array was first realized in the mid 1940s [12] and since then it has seen widespread use in applications such as radio detection and ranging (RADAR), broadcasting, and astronomy. The invention of the optical version of the phased arrays, the nanophotonic phased arrays, was made possible due to several advances starting from the invention of the laser in the early 1960s and high-precision silicon photonics fabrication.

Optical beam steering using one-dimensional NPAs allows stable, rapid, and precise beam steering without mechanical motion while providing large scanning angles. Two-dimensional NPAs have several applications, including LiDAR scanning and free-space optical communications. What is of particular interest to us is its use as a holographic display. Sun *et al.* [63] demonstrate a novel NPA architecture capable of integrating an optical directional coupler, a 2π tunable phase modulator and a nanoantenna within a compact $9\,\mu m \times 9\,\mu m$ unit and an infrared image containing text information which can be observed in the far field. Building an NPA on a silicon nitride platform can break its operating wavelength limitation and expand its applications into the visible spectrum. Raval *et al.* [57] show a static image with the horizontal parallax and a viewing angle of 5° generated on an NPA operating at $635\,nm$ wavelength. Notaros *et al.* [44] demonstrate for the first time an NPA-based visible-light near-eye holographic projector where a holographic image encoding methodology has been used and a 3D cube can be observed by the eye. However, in most of the works mentioned above, the phase shift for each pixel has been hard-coded during manufacturing and cannot be dynamically changed. As a result, only static images were shown. A notable exception is the work by Sun *et al.* [63] where a small-scale (8×8) array with active phase tuning is used to demonstrate infrared beam steering. Our NPA design supports active phase tuning for visible light and addresses the challenge of correcting thermal proximity effect that comes with thermally modulated NPAs. A phased array with acousto-optic phase modulation has also been shown by Grinenko *et al.* [15].

2.2 Near-eye Displays

Traditionally, near-eye displays use binocular vision to provide the parallax depth cue [22, 42]. While these displays and the modern commercial headsets can provide a very good field-of-view, they can only afford a single focal distance which leads to the vergence-accommodation mismatch. A comprehensive study by Rolland *et al.* [58] provides an engineering specification on the number of focal distances that is required to suppress the vergence-accommodation mismatch. To provide multiple focal distances, Akeley *et al.* [2] use several beamsplitters placed at different distances from the viewer to reflect the content from a single LCD monitor. Dunn *et al.* [10] use a deformable membrane

[1]Example code available at https://www.cs.umd.edu/~varshney/NPA_PPEC.

controlled by airtight cavities to modulate the focal distance. A gaze tracker is used to sense the desired focal distance. Commercialized focus-tunable lenses have also been used address VAC and vision correction in near-eye displays in Chakravarthula *et al.* [5] and Xia *et al.* [67]. Akşit *et al.* [3] and Kim *et al.* [21] use a motorized linear stage to move the optical combiner and the MOLED display on the optical axis to adjust the focal distance.

Light field displays encode the radiance as a function of position and angle [28] and have been shown to approximate the accommodation to the extent limited by the angular resolution [37]. Light field displays have been implemented based on a variety of underlying hardware platforms such as microlens arrays [25], attenuating LCDs [17, 34, 36, 66], and holographic optical elements combiner layer with projectors [26, 27]. Pamplona *et al.* [46] and Huang *et al.* [18] show that light field displays can correct impaired vision such as myopia and presbyopia and higher-order aberrations. A method to compute the optimal imagery on such displays with multiple focal planes has been described [43]. However, the angular resolution of light field displays is limited by the number of different views which together with the spatial resolution is bound by the device's form factor and pixel density. Attempting to further increase pixel density can lead to undesirable diffraction artifacts. Padmanaban *et al.* [45] solve this problem to some degree with Overlap-Add Stereogram (OLAS) at the cost of more intensive computation. Mi *et al.* [7, 41] have shown that the accommodation problem can be circumvented with a type of retinal scanning display that directly projects the image onto the user's retina.

Holographic displays have been built with pre-recorded holographic optical elements [3, 21] or with dynamic phase spatial light modulators [35, 61] or a combination of both [19]. This technology has the potential to fully address some hard problems faced by the current generation of VR and AR near-eye displays such as vergence-accommodation conflict. However, the refresh rate and form factor of the holographic displays falls short of what the NPA technology can provide.

For more details on near-eye displays we refer the interested reader to the comprehensive survey by Koulieris *et al.* [24].

2.3 Computer-Generated Holography and Phase-Only Holograms

Various methods have been proposed for computer-generated holography (CGH). For Fourier holograms, the holographic pattern of an image in the far field can be calculated as its Fourier transform. Lohmann *et al.* [31] show that a binary representation of the Fourier hologram can achieve good reconstruction quality for letters in a 2D image.

Fresnel holograms are calculated by considering the various scene objects, such as points and polygons, as a collection of light sources. Waters [65] synthesized a Gabor-style hologram by calculating the interference pattern from the light emitted by each point on an object and the illumination beam. Directly computing the interference pattern from every point is computationally intensive. Lucente [32] uses a look-up table corresponding to the interference contribution from points at each location in the image space and achieves a $43\times$ speedup. Yoshikawa *et al.* [69] use approximation methods to calculate the distance from a point source to a holographic pixel. Stereograms are small regions of a hologram. They each record the diffraction pattern of a portion of the scene instead of the entire scene and are used in Lucente and Galyean [33], Xu *et al.* [68], and Shi *et al.* [61] to reduce the amount of computation. More recently, Eybposh *et al.* [11] train an unsupervised convolutional neural network that outputs the near-field phase for a given desired image in real-time. Peng *et al.* [52] employ a camera in the hologram calculation process to iteratively optimize the hologram based on captured observation from the camera. These holograms are used to train a neural network that can output high-quality holograms in real-time.

Ziegler *et al.* [72] develop a conversion technique between the light field representation and the holographic representation of a scene. Padmanaban *et al.* [45] use this conversion to develop the OLAS algorithm to escape the trade-off between spatial and angular resolution in traditional light-field displays. Ways to add occlusion or change lighting in a hologram have also been explored in Ziegler *et al.* [73].

While an optical wavefront contains both phase and amplitude, many currently available holographic displays are only capable of modulating either only the amplitude or only the phase. Akşit [1] use a learning-based optimization approach to calculate the amplitude hologram that enhances the spatial resolution by up to four times when illuminated with a fast-refreshing locally-addressable backlight. The NPAs we are investigating here is only capable of phase modulation. Therefore, a method to convert fully complex holograms to phase-only holograms is required. Gerchberg and Saxton [14] proposed the first of a family of iterative phase retrieval algorithms that generate a near-field phase pattern which coupled with a predefined near-field amplitude pattern can propagate to form the desired far-field amplitude. We discuss this algorithm and our modifications in Section 4.1. The original Gerchberg-Saxton (GS) algorithm uses Fourier transform to approximate the Fraunhofer diffraction over a long distance. Fresnel diffraction is also frequently used in a similar way when the propagation distances are shorter.

Several researchers have made modifications to the iterative phase-retrieval algorithm. Masuda *et al.* [40] added a binarization step after the propagation from the object plane to the hologram plane to extract a binary phase. Similarly, Persson *et al.* [53] and Gemayel *et al.* [13] add a convolution step to counter the pixel voltage crosstalk on SLMs. However, these methods require heavy padding of the image to be effective and in our opinion do not make the full of the limited resolution on holographic displays. Nevertheless, we include these methods as the most recent existing proximity effect correction method even though the cause of proximity effect in our case is different.

A recent work on phase retrieval is Chakravarthula *et al.* [6] using non-convex optimization based on Wirtinger derivative. Our proposed proximal proximity effect correction method also uses non-convex optimization and shows that it can effectively and efficiently correct thermal proximity effect in NPAs.

2.4 Proximity Effect Correction

Parikh [47] introduces a self-consistent technique that provides a mathematically unique solution to the PEC problem given the proximity function by solving a set of linear equations. This technique is accompanied by a second technique that compensates for the unaddressed regions by solving an overdetermined system of equations. Carroll [4] expresses the PEC problem as a linear programming problem with constraints derived from the addressed and unaddressed pixels.

Chow *et al.* [8] use an image deconvolution approach based on the convolution theorem to efficiently solve the PEC problem. Similarly, a matrix inversion approach to deconvolution is tested by Peckerar *et al.* [51]. These deconvolution-based methods are likely to produce negative entries in the result, which is not physically acceptable. Simple techniques have been proposed to address this problem such as setting them to zero (chopping) and adding a background value equal to the absolute of the minimum entry (shifting) in Marrian *et al.* [39]. A more advanced method to avoid negative entries is to add a regularizer to the cost function in the gradient-descent method. One such regularizer used in [39, 51] is proportional to the Shannon entropy of the pattern. Alternatively, Peckerar *et al.* [50] express the non-negative requirement as well as the incident dose as constraints in a linear programming problem that minimizes the total dose. Sun *et al.* [64] try a number of PEC methods including quadratic programming on an NPA model that suffers from proximity effect in phase modulation but assumes independent and ideal amplitude modulation. Our work addresses the more practical, but also more challenging case of thermal proximity effect correction on a phase-only NPA.

There are many more PEC methods in E-beam lithography that we are not able to cover in the entirety here. We refer interested readers to Li [29] for a thorough review.

3 NPA Holographic Display

This section describes the basic structure and operation of an NPA display. The description is based on an NPA of our design currently under development. The differences between this NPA design and previous NPAs such as those in Sun *et al.* [63] and Raval *et al.* [57] are

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3023601, IEEE Transactions on Visualization and Computer Graphics

out of the scope of this paper. The focus of this study is the thermal proximity effect which is common in all thermal-modulated NPAs and its correction methods. Therefore, we first explain the phase modulation in Section 3.1, followed by how we model and quantify the thermal proximity effect in Section 3.2.

3.1 Phase Modulation

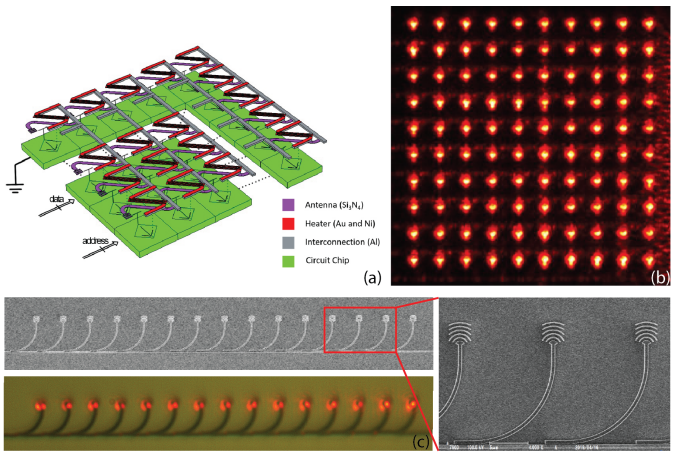

Fig. 2. The NPA display with electronic and optical layers is shown in (a). The optical power is evenly distributed into each pixel. Therefore, the near-field (b) has uniform amplitude. The antennas (c) propagate the near-field wavefront to form the image to be observed at the desired depth.

Figure 2(a) shows the schematics of the NPA we have developed. Each array unit (a pixel) consists of a tunable thermo-optical phase shifter that is coupled to an optical antenna. The refractive index of the phase shifter changes with temperature. Therefore, phase modulation can be achieved by heating each pixel. Special efforts have been put into reducing the physical size of the phase shifter while maintaining its sensitivity to temperature change. This way, the array can be small and a complete 2π phase shift can be achieved with a relatively small temperature range, thus achieving high energy efficiency. The optical power is evenly distributed into each pixel by a precise design of the directional couplers. Therefore, a uniform amplitude is achieved in the near field, as shown in Figure 2(b). The phase of the NPA holographic display pixel is modulated with the temperature. An electronic chip (green layer in Figure 2) is used to control the heating of each pixel.

There is a linear relationship between the phase ϕ_j of a pixel j and its temperature T_j (above room temperature), expressed in Equation 1.

$$\phi_j = \gamma T_j \tag{1}$$

Note that T_j must be non-negative for physical feasibility. From the simulations run on our design, γ is inferred to be $\frac{\pi}{175}\ \mathrm{rad}\,^{\circ}\mathrm{C}^{-1}$ for 630 nm red laser. With precise phase shifter design, γ is proportional to the wavelength, i.e. $\gamma = \frac{\pi}{147}\ \mathrm{rad}\,^{\circ}\mathrm{C}^{-1}$ for green (530 nm) and $\gamma = \frac{\pi}{124}\ \mathrm{rad}\,^{\circ}\mathrm{C}^{-1}$ for blue (445 nm) lasers. Given the desired 2D phase pattern of a hologram, we can find the desired 2D temperature profile $\tilde{T}$ with Equation 1.

3.2 Thermal Proximity Effect

The thermal proximity effect in the NPAs is the phenomenon where one pixel affects the neighboring pixel's temperature when it is being heated. Similar proximity effects are present in beam electron lithography and LCoS SLMs. In Peckerar *et al.* [49] and Persson *et al.* [53], the relation between the energy deposited and energy absorbed and the pixel voltage crosstalk are both modeled as convolutions. We also model the thermal proximity effect in the NPAs as a convolution. To quantify the proximity effect, we simulated a 5×5 array where we supplied the center pixel with power and measured the temperature on all pixels. An illustration of this thermal proximity effect can be seen in Figure 3. The data for this figure is obtained from fitting a Gaussian to the 25 values from the simulation on a 5×5 array which is described below.

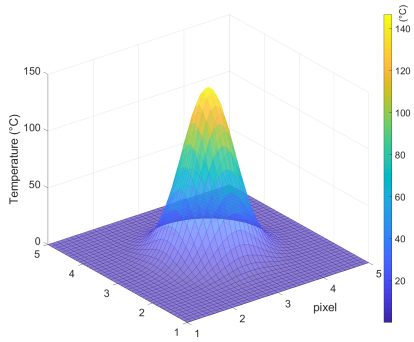

Fig. 3. Thermal crosstalk arising from power being supplied to the center pixel on a 5×5 array.

Using the measurement, we derive the following relationship:

$$T_{k\to j} = T_k e^{-\frac{d(j,k)^2}{\sigma^2}} \tag{2}$$

$T_{k\to j}$ refers to the temperature rise in pixel j caused by pixel k, T_j and T_k are the temperatures of pixels j and k before considering the proximity effect and $d(j,k)$ is the distance between pixels j and k measured in pixels (px). The temperature $\hat{T}_j$ of pixel j considering the proximity effect is the sum of T_j and the contributions from all surrounding pixels, expressed in Equation 3. The temperature of the entire NPA can be expressed as the convolution in Equation 4 where T and $\hat{T}$ are the 2D temperature profiles before and after considering the proximity effect and $*$ is the convolution operation.

$$\hat{T}_j = T_j + \sum_{k\neq j} T_{k\to j} \tag{3}$$

$$\hat{T} = T * K_\sigma \tag{4}$$

We derive that $\sigma = 0.66$ px based on the data obtained from thermal-electrical simulations. The convolution kernel K_σ in Equation 4 can be constructed using σ. We have used kernels of size 7×7 in our simulations.

The severity of the thermal proximity effect is quantified by the value of σ. Even though we derived that in our current design $\sigma = 0.66$ px, factors such as the thermal conductivity of the materials of the NPA and the pixel pitch could affect this value. For example, if the pixel pitch is reduced, heating one pixel would affect the temperature of its neighboring pixels more as they are closer to the pixel being heated. The impacts of the thermal proximity effect and the effectiveness of the correction methods are presented with varying levels of the thermal proximity effect in Section 5.

4 Proximity Effect Correction

In this section, we describe our novel proximity effect correction methods for phase-only NPA holographic displays. The first method is based on the Gerchberg-Saxton iterative phase retrieval algorithm. In this method, a deconvolution step is introduced to the source amplitude constraint in each iteration. We refer to this method as the Iterative Proximity Effect Correction (IPEC) method. The second method which we call the Proximal Proximity Effect Correction (PPEC) method, is based on the proximal algorithm. We use the proximal algorithm to directly find a near-field phase pattern that minimizes the difference between the far-field intensity it forms under the proximity effect and the desired image.

4.1 Iterative Proximity Effect Correction

The Gerchberg-Saxton iterative phase retrieval algorithm is a classic and well-known algorithm for calculating phase-only holograms. We do not wish to reiterate this algorithm in detail here but instead focus on how we integrate proximity effect correction into it. For completeness, we present a detailed description of this algorithm in the supplemental material.

The pseudocode for the IPEC method is presented in Algorithm 1. The inputs to this algorithm are $\mathcal{S}, \mathcal{T} \in \mathbb{R}^{M\times N}$, the source and target amplitude patterns each of size $M \times N$. The source amplitude $\mathcal{S}$ is the near-field amplitude on the NPA. According to Section 3, $\mathcal{S}$ is uniform in our setting. The target amplitude $\mathcal{T}$ is the amplitude we wish to form in the far field. The phase retrieval algorithm finds the near-field phase that coupled with $\mathcal{S}$ produces the far-field amplitude $\mathcal{T}$. A random initial phase $\phi_0 \in \mathbb{R}^{M\times N}$ may be used to initialize the algorithm. The function "phs" returns the phase of a complex wavefront. We use the Fourier hologram model where the propagation of the near-field wavefront to the observation plane is approximated with a Fourier transform. $\mathcal{F}$ represents the discrete Fourier transform. Let $P(\phi)$ be the function that applies the proximity effect to the phase pattern ϕ. P^{-1} is the reverse process.

Algorithm 1: Iterative Proximity Effect Correction

input : $\mathcal{S}, \mathcal{T}, \phi_0$
output : ϕ_r
$A = \mathcal{F}^{-1}(\mathcal{T} \cdot e^{i\phi_0})$;
while *Within maximum iteration* **do**
  $\phi_r = P^{-1}(\text{phs}(A))$;
  $B = \mathcal{S} \cdot e^{iP(\phi_r)}$ (source amplitude constraint);
  $C = \mathcal{F}(B)$;
  if *Termination condition met* **then**
    break;
  end
  $D = \mathcal{T} \cdot e^{i\text{phs}(C)}$ (target amplitude constraint);
  $A = \mathcal{F}^{-1}(D)$;
end

We integrate the proximity effect correction into the phase retrieval algorithm by modifying the step of the source amplitude constraint. Originally, the source amplitude constraint is applied by having B take the amplitude $\mathcal{S}$ and the phase of A. In our iterative proximity effect correction method, B takes the same amplitude $\mathcal{S}$ but the phase after taking the proximity effect and its correction into account. Specifically, the phase is $P(P^{-1}(\text{phs}(A)))$. P^{-1} takes into consideration the non-negative constraint. Therefore, in most cases, $P(P^{-1}(\phi)) \neq \phi$.

A good termination condition can be the difference, $\|\mathcal{T} - |C|\|_{\text{F}}$, between the amplitude of C and the target $\mathcal{T}$ becoming small enough. While this works well with the original phase retrieval algorithm, we find it hard to define a threshold when P^{-1} is introduced. Instead, we terminate when the minimum and maximum values of $|\mathcal{T} - |C||$ are not updated for several consecutive iterations.

In our setting, P is a convolution, i.e. $P(\phi) = K * \phi$ where ϕ is a 2D phase pattern. P^{-1} is therefore a deconvolution with non-negative constraints. We can express P^{-1} in the following form. $M_K \boldsymbol{x} = \boldsymbol{\phi}$ is the matrix form of the convolution. $M_K \in \mathbb{R}^{MN\times MN}$. $\boldsymbol{\phi}, \boldsymbol{x} \in \mathbb{R}^{MN\times 1}$.

$$P^{-1}(\boldsymbol{\phi}) = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|M_K \boldsymbol{x} - \boldsymbol{\phi}\|_2^2 \quad \text{s.t.} \quad \boldsymbol{x} \geq 0 \tag{5}$$

A basic deconvolution method is matrix inversion. If M_K is non-singular, which it is in our case, we can simply take $\boldsymbol{x} = M_K^{-1}\boldsymbol{\phi}$. In practice, it is faster and more robust to solve the linear system $M_K \boldsymbol{x} = \boldsymbol{\phi}$. However, this method is likely to produce a solution that contains negative entries corresponding to a temperature lower than the room temperature. This is physically impossible to achieve on the current design of the NPA.

A more robust deconvolution method is quadratic programming, which is formulated in Equation 6. Traditional quadratic programming algorithms such as the interior-point algorithm can be used. Alternatively, we can use gradient descent to minimize the cost function in Equation 6 and use a regularizer to ensure the solution is non-negative.

$$\min_{\boldsymbol{x}} \quad \frac{1}{2}\boldsymbol{x}^{\text{T}}(M_K^{\text{T}} M_K)\boldsymbol{x} + (-\phi^{\text{T}} M_K)\boldsymbol{x} \quad \text{s.t.} \quad \boldsymbol{x} \geq 0 \tag{6}$$

In the simulations, we use the quadratic programming deconvolution method described above.

4.2 Proximal Proximity Effect Correction

In the proximal proximity effect correction method, we set out to find the near-field phase pattern that minimizes the difference between the far-field intensity it forms under the proximity effect and the desired image I. Namely, given the image $I \in \mathbb{R}^{M\times N}$, we want to find the non-negative near-field phase pattern $X \in \mathbb{R}^{+M\times N}$ that minimizes $g(X)$ in Equation 7. $U(X)$ is the intensity of the wavefront the NPA with input phase X forms in the far field, which can be calculated in Equation 8. Again, $P(X)$ is the result of the proximity effect on the input phase X. It should be noted that the proximity effect should be applied to the temperature instead of the phase. However in our case, since there is a relationship between the phase and temperature on the NPA, we directly use the phase X for simplicity in the problem formulation.

$$g(X) = \|U(X) - I\|_F^2 \tag{7}$$

$$U(X) = |\mathcal{F}(e^{iP(X)})|^2 \tag{8}$$

Proximal algorithms are generally used for solving convex optimization problems [48]. However, as is shown in Heide *et al.* [16] and will be shown in Section 5, the proximal algorithms can also work well on our non-convex problems.

The proximal algorithms work by repeatedly evaluating the *proximal operator*. The proximal operator of a function f is defined as follows ([48]). λ is the step size parameter.

$$\text{prox}_{\lambda f}(v) = \underset{x}{\operatorname{argmin}}(f(x) + \frac{1}{2\lambda}\|x - v\|_2^2) \tag{9}$$

To enforce the non-negative constraint, we add an indicator function to the cost function $f(X)$ in Equation 10.

$$f(X) = g(X) + c(X) \tag{10}$$

$$c(X) = \mathcal{I}_{[0,+\inf)}(X) \tag{11}$$

Various proximal algorithms have been proposed. In this study, we use the accelerated proximal gradient algorithm. This algorithm splits the cost function into two parts, one of which is differentiable. In our case, $f(X)$ is the differentiable term. Let X^k and λ^k be the solution and the step size in iteration k. Y^k is the extrapolated solution in iteration k whose purpose is to move the solution a step further towards convergence. The algorithm is carried out iteratively according to Equation 12.

$$\begin{aligned} Y^{k+1} &= X^k + \omega^k(X^k - X^{k-1}) \\ X^{k+1} &= \text{prox}_{\lambda^k c}(Y^{k+1} - \lambda^k \nabla g(Y^{k+1})) \end{aligned} \tag{12}$$

A simple choice for ω^k is $\frac{k}{k+3}$ [48]. The proximal operator of the indicator function $\text{prox}_{\lambda c}(v)$ projects the input v into the non-negative space. The gradient of the function $g(X)$, $\nabla g(X)$, has been

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3023601, IEEE Transactions on Visualization and Computer Graphics

Original | Basic Iterative PEC Gemayel *et al.* [13] | Iterative PEC | Proximal PEC

Fig. 4. This figure compares the various Proximity Error Correction (PEC) methods. The proximity effect level used in the first two rows is $\sigma = 0.66$ px. At this level, the IPEC results are noisy but still reasonable. The proximity effect level for the bottom row is $\sigma = 0.8$ px. At this level, the results from the IPEC methods are incomprehensible but the Proximal PEC method provides readable results. The images in the first two rows are from the LIVE dataset (Sheikh *et al.* [59, 60]) and the newspaper image in the bottom row is from the CHRONIC dataset (Smits and Faber [62]).

derived in the supplemental material. The function "Re" takes the real components of a complex value. $\overline{X}$ is the complex conjugate of X. K is the convolution kernel.

$$\nabla g(X) = 4K * \mathrm{Re}\{[ie^{iP(X)} \cdot \mathcal{F}((U(X) - I) \cdot \overline{e^{iP(X)}})]\} \quad (13)$$

5 Simulations

We have carried out our computational simulations on a dataset containing 29 real-life images (Sheikh *et al.* [59, 60]) and an additional newspaper image containing text from the CHRONIC dataset (Smits and Faber [62]). Our software implementation uses Matlab running on an Intel 9900K CPU. We present the simulation results in color as our proximity effect correction methods are applicable to all wavelengths. Multi-color displays can be easily implemented by coupling three lasers into the NPA.

5.1 Implementation and Test

As mentioned before in Section 4.1, we perform the deconvolution by minimizing the residual $\|M_K \boldsymbol{x} - \boldsymbol{\phi}\|_2^2$ while maintaining $\boldsymbol{x} \geq 0$. Specifically, the cost function $f(\boldsymbol{x})$ to minimize in the proximal algorithm is the following.

$$f(\boldsymbol{x}) = g(\boldsymbol{x}) + c(\boldsymbol{x}) \quad (14)$$

$$g(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}(M_K^{\mathrm{T}} M_K)\boldsymbol{x} + (-\phi^{\mathrm{T}} M_K)\boldsymbol{x} \quad (15)$$

$$c(\boldsymbol{x}) = \mathcal{I}_{[0,\mathrm{inf}]}(\boldsymbol{x}) \quad (16)$$

We use the accelerated proximal gradient algorithm. The gradient of $g(\boldsymbol{x})$ is evaluated as the following.

$$\nabla g(\boldsymbol{x}) = M_K^{\mathrm{T}} M_K - \boldsymbol{\phi}^{\mathrm{T}} M_K \quad (17)$$

For the PPEC method, we use the accelerated proximal algorithm exactly like described in Section 4.2. We use the residual change to determine when to terminate the algorithm. Let X^k be the solution in iteration k. The residual is $g(X^k)$ from Equation 7. Let r^k the change in iteration k compared to the previous iteration. $r_k = \frac{\mathrm{abs}(g(X)^k - g(X)^{k-1})}{g(X)^{k-1}}$. We terminate the algorithm when r_k drops below 0.001.

In addition to showing the simulated reconstructions of the far-field images, we also quantify the quality of these images with *structural similarity index* (SSIM) from Wang *et al.* [71]. The power of the luminance, contrast, and structure terms we use in the SSIM are all 1. $\mathrm{SSIM}(I, R)$ measures the similarity between the input image I and the reference image R. When I and R are identical, $\mathrm{SSIM}(I, R) = 1$. As the similarity decreases, so does the SSIM. We have also considered two other image quality metrics, the classic Peak Signal-to-Noise Ratio (PSNR) and a neural network-based perceptual similarity by Zhang *et al.* [70]. We did not find any contradictions between these image metrics and direct observation in our tests. The larger the proximity effect level, the more noise is present and the lower the quality metric score gets. For more detail on these image quality metrics in our test, please refer to the supplemental material.

5.2 Correction Effectiveness

We first present the qualitative visual results in Figure 4. The original images are shown on the left as references. We also include the method in Persson *et al.* [53] and Gemayel *et al.* [13] in the second column. We refer to this method as the Basic Iterative Proximity Effect Correction. The basic IPEC method skips the deconvolution step when applying the source amplitude constraint, namely using $B = \mathcal{S} \cdot e^{iP(\mathrm{phs}(A))}$ in Algorithm 1. The results of the IPEC method and the PPEC method are presented in the remaining two columns, respectively. The proximity effect level in the first two rows is $\sigma = 0.66$ (px) corresponding to our sample device. To show the impact of a greater proximity effect level especially on text-rich images, we use $\sigma = 0.8$ (px) in the third row with the image of a newspaper.

It can be seen from Figure 4, the IPEC method performs much better

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3023601, IEEE Transactions on Visualization and Computer Graphics

than the basic IPEC method and is able to produce reasonable results. However, it is still not able to effectively correct high levels of proximity effect. The PPEC method has the best correction effectiveness of all methods compared and is the only method that can recover text-rich images when the proximity effect level is high. At high proximity effect levels such as $\sigma = 0.80$ px, a vignette effect (where the center of the image is brighter than the border region) can be observed in the results of the basic IPEC as well as IPEC. This is because the proximity effect is modeled as a Gaussian convolution. When the proximity effect is high, the Gaussian kernel gets flatter and the near-field phase becomes more uniform. Since the Fourier transform is used to approximate the propagation, more energy gets focused around the center in the far field.

Figure 5 shows how the average image quality degrades with increasing amounts of proximity effect for the methods. We observe that the image quality degrades with increasing proximity effect σ for the IPEC method. The PPEC method retains the best image quality at all proximity effect levels. At $\sigma = 0.66$ px, the IPEC and PPEC methods improve the SSIM image quality metric by 121.26% and 249.98% compared to the basic IPEC method, respectively.

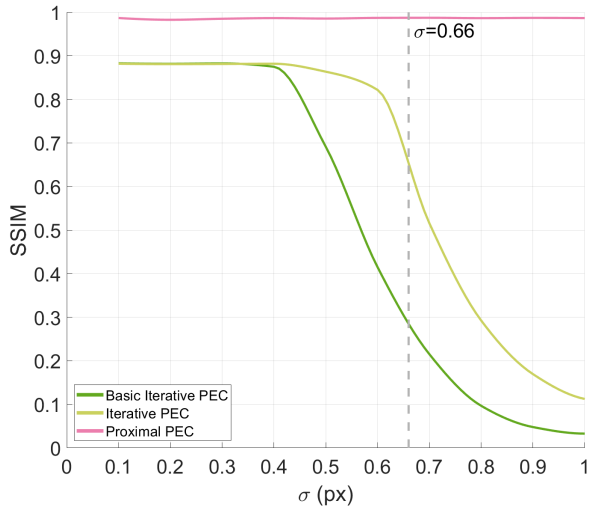

Fig. 5. Image quality degradation for various methods. This figure is plotted with interpolated data from simulations where σ ranges from 0.10 to 1.00 px with a 0.10 px interval.

5.3 Processing Time

We also look at the processing time of the proposed proximity effect correction methods. Two factors may impact the processing time, the severity of the proximity effect and the resolution of the image.

The processing time of the proposed methods with increasing thermal proximity effect levels is presented in Fig 6. It is clear at first glance that the IPEC method is very time-consuming. It is to be expected since IPEC is an iterative method and at the same time, the deconvolution in each iteration is also implemented in an iterative manner. The PPEC method is able to maintain a much faster processing speed than the IPEC method. The processing time for the PPEC method also increases with the proximity effect level. The basic iterative PEC method is the fastest of the three methods tested. However, as shown in Figure 4 and 5, this method does not provide sufficient proximity effect correction.

The processing time of the methods with increasing image resolution is shown in Figure 7. The processing time increases with image resolution for both the IPEC and PPEC methods.

5.4 Spatial Frequency of Images

Since the Gaussian filter used in the convolution modeling of the thermal proximity effect is a low-pass filter, we want to investigate whether

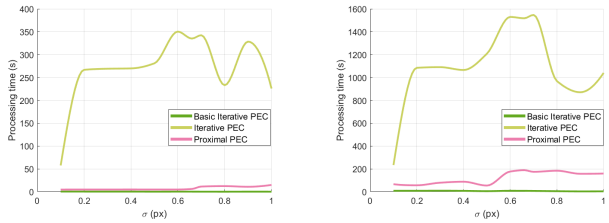

Fig. 6. The processing time of the proposed proximity effect correction methods with increasing levels of proximity effect at respective resolution.

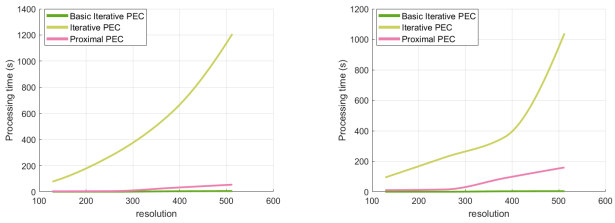

Fig. 7. The processing time of the proposed proximity effect correction methods on images of increasing resolution. The image aspect ratio is kept at 3 : 2. The resolution on the x-axis indicates the vertical resolution of the image.

the spatial frequency of the image affects the image quality of the proximity effect correction methods.

To test this, we prepare images filtered by low-pass filters with several frequency cutoff thresholds. We use the Butterworth low-pass filter to avoid the ringing artifacts in the filtered images. The low-pass filter we use is expressed in Equation 18. $f(u, v)$ is the normalized distance from (u, v) to the center of the spectrum. $f(u, v) = \sqrt{(\frac{2u}{M} - 0.5)^2 + (\frac{2v}{N} - 0.5)^2}$. f_c is the frequency cutoff threshold. We use frequency cutoff thresholds $f_c = 0.1, 0.2, 0.3, 0.4, 0.5$. The results of the PEC methods on the low-pass-filtered images are shown in Figure 8.

$$\text{BLPF}(I) = \mathcal{F}^{-1}[\mathcal{F}(I) \cdot H] \tag{18}$$

$$H(u, v) = \frac{1}{1 + [f(u, v)/f_c]^4} \tag{19}$$

As can be seen in Figure 8, there is no significant change of image quality with different image spatial frequencies. This may be because the thermal proximity effect mainly causes noise artifacts in the reconstructed images rather than blurring.

6 Discussions

In this work, we propose the iterative proximity effect correction which integrates deconvolution into the iterative phase retrieval algorithm and the proximal proximity effect correction which directly finds the phase that optimizes the formed image under thermal proximity effect. In this section, we discuss some details of the proposed methods.

6.1 Improving IPEC

As can be seen in Figure 5, the PPEC method achieves better image quality at all the thermal proximity effect levels tested. In fact, the PPEC method is able to maintain the image quality with increasing thermal proximity effect levels, albeit at the cost of slightly longer processing time (see Figure 6). IPEC methods perform reasonably well at lower proximity effect levels. But at high proximity effect levels, the IPEC method inevitably breaks down. An example with proximity effect level $\sigma = 0.8$ px can be seen in Figure 4 bottom row. However,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3023601, IEEE Transactions on Visualization and Computer Graphics

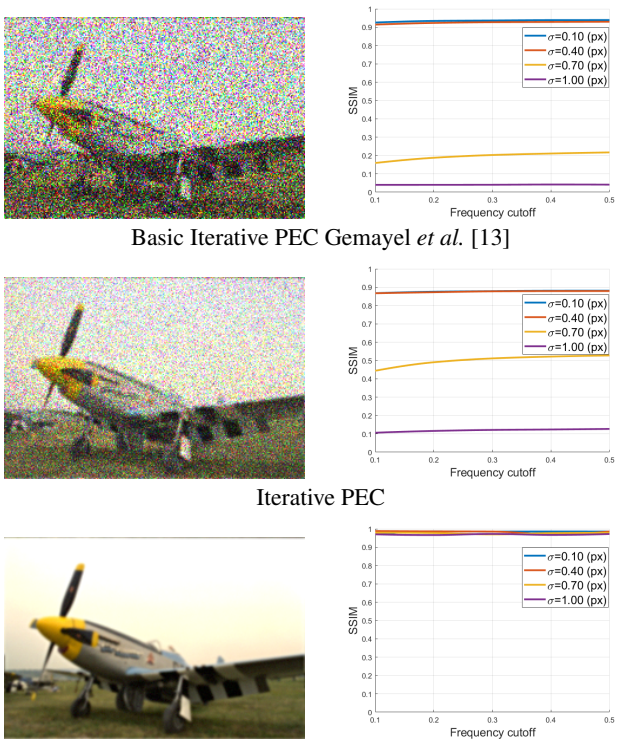

Fig. 8. The results of the basic iterative PEC method (top row) and the iterative PEC method (middle row) and the proximal PEC method (bottom row) on low-pass filtered images. The images (left column) are the results of the respective PEC methods applied to a low-pass filtered image with frequency cutoff threshold $f_c = 0.2$ at proximity effect level $\sigma = 0.7$ px. The image qualities of each PEC method at different proximity effect levels on images with different frequency cutoff thresholds are shown in the right column. We run the methods on each image five times and show the average image quality.

Fig. 9. (a) shows the result of TDM application to the IPEC method with 10 different initial phase. (b) shows the of result of the padding technique applied to IPEC method. The center signal region is shown. The image is padded from 171×256 to 512×512. The initial and result padded images are shown in (c) and (d), respectively. The thermal proximity effect level is $\sigma = 0.66$ px for both techniques.

some techniques can be applied to the IPEC method to further improve image quality.

The first technique is time-division multiplexing (TDM). We observe that the artifacts in the IPEC results are noise artifacts. We can run the IPEC method multiple times with different initial phase ϕ_0 and display the different resulting phase patterns in consecutive frames. This would increase the processing time of the IPEC method. The TDM technique uniquely takes advantage of the fast refreshing nature of the NPA. We can simulate the TDM by averaging the resulting far-field images, which are shown in Figure 9(a). Obviously, if IPEC is run for n times, the TDM technique would take n times the processing time of the IPEC method.

The other technique is to pad the desired image with a background region. In the iterative phase retrieval algorithm, a signal region and a background region on the wavefront can be specified. The two regions have no intersection. We want the signal region to display the desired amplitude pattern therefore the target amplitude constraint is only applied in the signal region. The background region offers more freedom to phase manipulation. We have included more detail on the iterative phase retrieval algorithm in the supplemental material. An example of this technique is shown in Figure 9(b). The padding technique uses more pixels for a given image than PPEC and wastes a portion of the display. Padding also increases the processing time.

A comparison of the two techniques with the IPEC and PPEC methods is shown in Figure 10. As can be seen, both techniques perform much better than the original IPEC method without the techniques. The two techniques achieve comparable image qualities to the PPEC method at lower proximity effect ($\sigma < 0.7$ px). This indicates that the IPEC method has the potential to achieve comparable correction effectiveness as the PPEC method. However, it is necessary to find fast deconvolution methods for the IPEC method to be efficient.

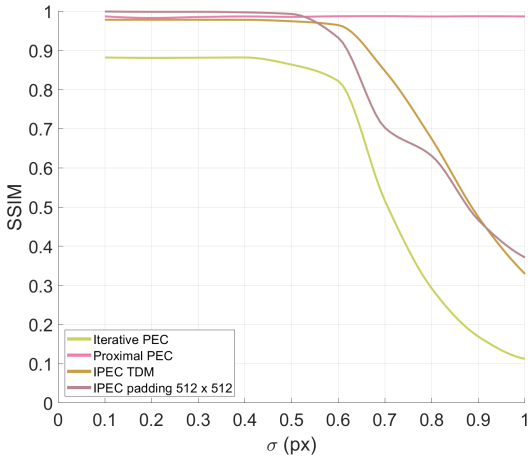

Fig. 10. IPEC with TDM and padding techniques compared to the original IPEC method and the PPEC method.

6.2 Pixel Pitch and Thermal Proximity Effect

Reducing the pixel pitch is one of the most effective ways of improving the quality of a holographic display. Given the form factor of the display, reducing the pixel pitch not only increases the number of pixels but also improves the field-of-view of the display. The importance of pixel pitch has been discussed in a number of studies on LCoS based holographic displays such as Shi *et al.* [61] and Jang *et al.* [19]. The same principle also applies to NPA holographic displays. It is easy to see that when the pixel pitch decreases, the pixels are more densely packed and each pixel is affected more by the heat from neighboring pixels. This is highly likely to lead to a more severe proximity effect. The PPEC method

is the only effective correction method at the moment to address this problem.

6.3 Choice of Proximal Algorithm

There are multiple variations of the proximal algorithm. Two examples are the proximal gradient algorithm and the ADMM (alternating direction method of multipliers). Both algorithms can be used to minimize the function $f(X) = g(X) + c(X)$ in Equation 10. In the proximal gradient algorithm, let X^k be the solution in iteration k. The update in iteration $k+1$ is the following.

$$X^{k+1} = \text{prox}_{\lambda c}(X_k) \tag{20}$$

In the ADMM algorithm, the updates in each iteration are the following. X^k and Z^k converge to each other and are in the domains of f and g respectively.

$$X^{k+1} = \text{prox}_{\lambda f}(Z^k - U^k) \tag{21}$$

$$Z^{k+1} = \text{prox}_{\lambda g}(X^{k+1} + U^k) \tag{22}$$

$$U^{k+1} = U^k + X^{k+1} - Z^{k+1} \tag{23}$$

The main difference between the two algorithms is the evaluation of the proximal operator of the function $g(X)$. $g(X)$ is a non-linear function, and minimizing it is the core of the proximal proximity effect correction problem. Since it is difficult to evaluate $\text{prox}_{\lambda g}$, we use the proximal gradient algorithm which only involves calculating the gradient of $g(X)$. In practice, we use the accelerated version of the proximal gradient algorithm. Its update steps are described in Section 4.2. The accelerated proximal gradient algorithm has an extrapolation step that can be interpreted as moving the solution a step faster towards the optimal.

In the implementation of the quadratic programming deconvolution in the IPEC method, we also use the proximal algorithm. Even though there exists closed-form evaluation of the function $g(\boldsymbol{x})$ in Equation 15 (see Parikh and Boyd [48]), it still involves solving a linear system with tens of thousands of variables. We have found that the accelerated proximal gradient implementation is faster than the ADMM implementation.

6.4 Iterations of PPEC

Here we show how the image quality improves over the iterations in the PPEC method. An example with proximity effect level $\sigma = 0.66$ px is shown in Figure 11.

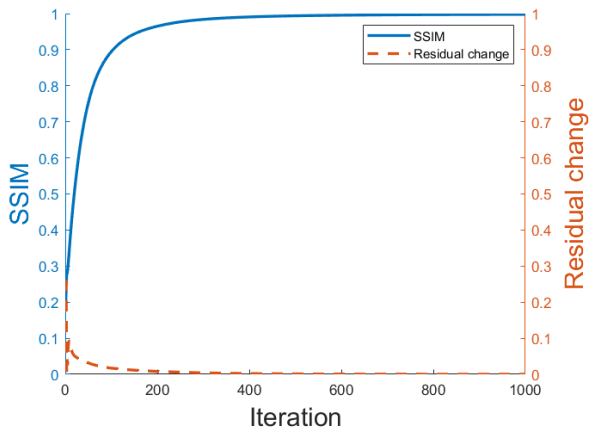

Fig. 11. Image quality improves over the iterations. The orange dashed line indicates the change in the residual compare to the previous iteration. $\sigma = 0.66$ px in this case.

The image quality improves over the iterations. We use the residual change r^k to determine when to terminate the algorithm. We terminate the algorithm when r^k drops below 0.001 at which point the image quality changes very little. The simulated observations of the PPEC method at several iterations before termination can be seen in the supplemental material.

7 Limitations and Future Work

Our simulations have shown that the proposed PPEC method achieves very good correction effectiveness across different proximity effect levels. However, there are still limitations to this approach. The NPA's properties can help addresses some of the limitations. Some of the properties may pose problems as well. Here we list some of the limitations and propose potential solutions.

7.1 Amplitude Modulation and Diminished Reality

NPAs have the advantage of fast refresh rate and integrated light source. However, NPAs also have limitations. The integrated light source makes it difficult for NPAs to modify the light from the real scene in augmented reality. Reflective or transmissive SLMs can modify the wavefront from the real scene that reaches them, achieving diminished reality or occlusion of real objects in AR, such as in Rathinavel *et al.* [56]. Combining LCoS SLMs can expand the capabilities of NPA holographic displays.

7.2 Choice of Initial Phase

The initial phase value of X in the PPEC method has an effect on the result. We find that a constant value often leads to poor image quality or an excessively long time to convergence. We use random values in our implementation and about 1% of the PPEC results suffer from slight speckle noise. We currently deal with this by detecting the noise and rerun PPEC with a different initial value. Heide *et al.* [16] propose to use the result from the iterative phase retrieval algorithm as the initial value for the phase retrieval using the proximal algorithm. We do not think it is a viable choice for us because the iterative phase retrieval would add to the processing time of the method. Still, the intermediate result of the iterative method after only a few iterations may offer better robustness for our method.

7.3 Acceleration

We have implemented the proposed PEC methods in Matlab running on the CPU for fast prototyping. Currently, it takes our implementation of the PPEC method 6.61s on average to process an image in our simulations. We expect a speed-up of at least 100 times when we move to parallelized implementation on the GPU, similar to the speed-up reported in Padmanaban *et al.* [45]. It is possible to bring the processing time to real-time rates through hardware improvement and massive parallelization in the future.

7.4 Larger Proximity Effect for PPEC

We have shown that the PPEC method effectively corrects the proximity effect within the tested range (from $\sigma = 0.10$ px to 1.00 px). While this is a realistic range in which our NPA design lies, the PPEC method may not suffice at higher proximity effect levels. For example, at $\sigma = 1.50$ px, the average SSIM of the PPEC method is 0.48, and at $\sigma = 2.00$ px, the average SSIM drops to 0.30.

7.5 3D Holograms

In this study, we have explored proximity effect correction on NPAs to form 2D images. Holographic displays truly shine at displaying realistic 3D scenes at the correct depth. The most naive approach would be to generate the phase-only hologram of a 3D scene either with double-phase encoding (such as in Maimone *et al.* [35]) or multi-plane phase retrieval (such as Makowski *et al.* [38]) and apply deconvolution to the phase-only hologram. Or we can integrate deconvolution into the multi-plane phase retrieval just like the proposed IPEC method. Noise and discontinuity in the depth can be expected from these approaches. Another approach that takes advantage of the fast refresh rate of the NPA is time-division multiplexing. Similar to what is suggested in

Chakravarthula *et al.* [6], we can divide the 3D scene into a stack of images at different depth and apply PPEC to each image. The results are displayed in consecutive frames. The PEC methods need to adapt to a different wave propagation model that specifies the propagation distance such as the Fresnel propagation or Fraunhofer propagation. We also plan to look at direct optimization methods that optimize the 3D scene as a whole.

8 Conclusion

Nanophotonic phased arrays have several advantages in its use as a holographic display. The integrated light source enables small-factor integration and the fast refresh rate is ideal for dynamic content. The use of NPA in a near-eye display has been demonstrated in Notaros *et al.* [44] and other works. However, the use of a large scale dynamically modulated NPA as a holographic display has not been achieved. We propose two proximity effect correction methods to address the thermal proximity effect issue faced by NPA holographic displays. The proximal proximity effect correction is the obvious choice as it is faster and more effective in correcting proximity effect at all proximity effect levels. The iterative proximity effect correction is effective at low proximity effect levels and the correction effectiveness can be improved with TDM or padding techniques. The IPEC method is also readily applicable to multi-plane scenes. However, more efficient deconvolution methods with non-negative constraints are required for the IPEC method to be practical.

Acknowledgement

We would like to thank the anonymous reviewers for the helpful comments that have significantly improved this paper. We also greatly appreciate the insightful suggestions by Ward Lopes and David Luebke from NVIDIA. This work has been supported in part by the NSF Grants 15-64212, 18-23321, and the State of Maryland's MPower initiative. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the research sponsors.

References

[1] K. Akşit. Patch scanning displays: spatiotemporal enhancement for displays. *Opt. Express*, 28(2):2107–2121, Jan 2020. doi: 10.1364/OE.380858

[2] K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks. A stereo display prototype with multiple focal distances. *ACM Trans. Graph.*, 23(3):804–813, Aug. 2004. doi: 10.1145/1015706.1015804

[3] K. Akşit, W. Lopes, J. Kim, P. Shirley, and D. Luebke. Near-eye varifocal augmented reality display using see-through screens. *ACM Transactions on Graphics (TOG)*, 36(6):189, 2017.

[4] A. M. Carroll. Proximity-effect correction with linear programming. *Journal of Applied Physics*, 52(1):434–437, 1981.

[5] P. Chakravarthula, D. Dunn, K. Akşit, and H. Fuchs. 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.

[6] P. Chakravarthula, Y. Peng, J. Kollin, H. Fuchs, and F. Heide. Wirtinger holography for near-eye displays. *ACM Trans. Graph.*, 38(6), Nov. 2019. doi: 10.1145/3355089.3356539

[7] J. Chen, L. Mi, C. P. Chen, H. Liu, J. Jiang, and W. Zhang. Design of foveated contact lens display for augmented reality. *Opt. Express*, 27(26):38204–38219, Dec 2019. doi: 10.1364/OE.381200

[8] D. G. L. Chow, J. F. McDonald, D. C. King, W. Smith, K. Molnar, and A. J. Steckl. An image processing approach to fast, efficient proximity correction for electron beam lithography. *Journal of Vacuum Science & Technology B: Microelectronics Processing and Phenomena*, 1(4):1383–1390, 1983. doi: 10.1116/1.582705

[9] J. K. Doylend, M. J. R. Heck, J. T. Bovington, J. D. Peters, L. A. Coldren, and J. E. Bowers. Two-dimensional free-space beam steering with an optical phased array on silicon-on-insulator. *Opt. Express*, 19(22):21595–21604, Oct 2011. doi: 10.1364/OE.19.021595

[10] D. Dunn, C. Tippets, K. Torell, P. Kellnhofer, K. Akşit, P. Didyk, K. Myszkowski, D. Luebke, and H. Fuchs. Wide field of view varifocal near-eye display using see-through deformable membrane mirrors. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1322–1331, 2017.

[11] M. H. Eybposh, N. W. Caira, M. Atisa, P. Chakravarthula, and N. C. Pégard. DeepCGH: 3D computer-generated holography using deep learning. *Opt. Express*, 28(18):26636–26650, Aug.

[12] A. J. Fenn, D. H. Temme, W. P. Delaney, and W. E. Courtney. The development of phased-array radar technology. *Lincoln Laboratory Journal*, 12(2):321–340, 2000.

[13] P. Gemayel, B. Colicchio, A. Dieterlen, and P. Ambs. Cross-talk compensation of a spatial light modulator for iterative phase retrieval applications. *Applied optics*, 55(4):802–810, 2016.

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

[15] A. Grinenko, M. P. MacDonald, C. R. P. Courtney, P. D. Wilcox, C. E. M. Demore, S. Cochran, and B. W. Drinkwater. Tunable beam shaping with a phased array acousto-optic modulator. *Opt. Express*, 23(1):26–32, Jan 2015. doi: 10.1364/OE.23.000026

[16] F. Heide, S. Diamond, M. Nießner, J. Ragan-Kelley, W. Heidrich, and G. Wetzstein. Proximal: Efficient image optimization using proximal algorithms. *ACM Trans. Graph.*, 35(4), July 2016. doi: 10.1145/2897824.2925875

[17] F.-C. Huang, K. Chen, and G. Wetzstein. The light field stereoscope: Immersive computer graphics via factored near-eye light field displays with focus cues. *ACM Trans. Graph.*, 34(4):60:1–60:12, July 2015. doi: 10.1145/2766922

[18] F.-C. Huang, G. Wetzstein, B. A. Barsky, and R. Raskar. Eyeglasses-free display: Towards correcting visual aberrations with computational light field displays. *ACM Trans. Graph.*, 33(4):59:1–59:12, July 2014. doi: 10.1145/2601097.2601122

[19] C. Jang, K. Bang, G. Li, and B. Lee. Holographic near-eye display with expanded eye-box. *ACM Trans. Graph.*, 37(6), Dec. 2018. doi: 10.1145/3272127.3275069

[20] M. A. Khalighi and M. Uysal. Survey on free space optical communication: A communication theory perspective. *IEEE Communications Surveys Tutorials*, 16(4):2231–2258, 2014.

[21] J. Kim, Y. Jeong, M. Stengel, K. Akşit, R. Albert, B. Boudaoud, T. Greer, J. Kim, W. Lopes, Z. Majercik, P. Shirley, J. Spjut, M. McGuire, and D. Luebke. Foveated AR: Dynamically-foveated augmented reality display. *ACM Trans. Graph.*, 38(4):99:1–99:15, July 2019. doi: 10.1145/3306346.3322987

[22] K. Kiyokawa. A wide field-of-view head mounted projective display using hyperbolic half-silvered mirrors. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, ISMAR '07, pp. 1–4. IEEE Computer Society, Washington, DC, USA, 2007. doi: 10.1109/ISMAR.2007.4538848

[23] R. Konrad, E. A. Cooper, and G. Wetzstein. Novel optical configurations for virtual reality: Evaluating user preference and performance with focus-tunable and monovision near-eye displays. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pp. 1211–1220. ACM, New York, NY, USA, 2016. doi: 10.1145/2858036.2858140

[24] G. A. Koulieris, K. Akşit, M. Stengel, R. K. Mantiuk, K. Mania, and C. Richardt. Near-eye display and tracking technologies for virtual and augmented reality. *Computer Graphics Forum*, 38(2):493–519, 2019. doi: 10.1111/cgf.13654

[25] D. Lanman and D. Luebke. Near-eye light field displays. *ACM Transactions on Graphics (TOG)*, 32(6):220, 2013.

[26] S. Lee, C. Jang, S. Moon, J. Cho, and B. Lee. Additive light field displays: Realization of augmented reality with holographic optical elements. *ACM Trans. Graph.*, 35(4):60:1–60:13, July 2016. doi: 10.1145/2897824.2925971

[27] S. Lee, C. Jang, S. Moon, B. Lee, J. Cho, and B. Lee. See-through light field displays for augmented reality. In *SIGGRAPH ASIA 2016 Virtual Reality Meets Physical Reality: Modelling and Simulating Virtual Humans and Environments*, SA '16, pp. 3:1–3:2. ACM, New York, NY, USA, 2016. doi: 10.1145/2992138.2992142

[28] M. Levoy and P. Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pp. 31–42. ACM, 1996.

[29] P. Li. A review of proximity effect correction in electron-beam lithography. *arXiv preprint arXiv:1509.05169*, 2015.

[30] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft COCO: Common objects in context. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, eds., *Computer Vision –*

ECCV 2014, pp. 740–755. Springer International Publishing, Cham, 2014.

[31] A. Lohmann and D. Paris. Synthesis of binary holograms. *IEEE Journal of Quantum Electronics*, 2(4):153–153, 1966.

[32] M. E. Lucente. Interactive computation of holograms using a look-up table. *Journal of Electronic Imaging*, 2(1):28–35, 1993.

[33] M. E. Lucente and T. A. Galyean. Rendering interactive holographic images. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pp. 387–394. ACM, 1995.

[34] A. Maimone and H. Fuchs. Computational augmented reality eyeglasses. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 29–38. IEEE, 2013.

[35] A. Maimone, A. Georgiou, and J. S. Kollin. Holographic near-eye displays for virtual and augmented reality. *ACM Transactions on Graphics (TOG)*, 36(4):85, 2017.

[36] A. Maimone, D. Lanman, K. Rathinavel, K. Keller, D. Luebke, and H. Fuchs. Pinlight displays: wide field of view augmented reality eyeglasses using defocused point light sources. In *ACM SIGGRAPH 2014 Emerging Technologies*, p. 20. ACM, 2014.

[37] A. Maimone, G. Wetzstein, M. Hirsch, D. Lanman, R. Raskar, and H. Fuchs. Focus 3D: Compressive accommodation display. *ACM Trans. Graph.*, 32(5):153–1, 2013.

[38] M. Makowski, M. Sypek, I. Ducin, A. Fajst, A. Siemion, J. Suszek, and A. Kolodziejczyk. Experimental evaluation of a full-color compact lensless holographic display. *Optics express*, 17(23):20840–20846, 2009.

[39] C. R. Marrian, S. Chang, and M. C. Peckerar. Proximity correction for electron beam lithography. *Optical Engineering*, 35(9):2685–2693, 1996.

[40] K. Masuda, Y. Saita, R. Toritani, P. Xia, K. Nitta, and O. Matoba. Improvement of image quality of 3D display by using optimized binary phase modulation and intensity accumulation. *Journal of Display Technology*, 12(5):472–477, 2015.

[41] L. Mi, C. P. Chen, Y. Lu, W. Zhang, J. Chen, and N. Maitlo. Design of lensless retinal scanning display with diffractive optical element. *Opt. Express*, 27(15):20493–20507, Jul 2019. doi: 10.1364/OE.27.020493

[42] H. Nagahara, Y. Yagi, and M. Yachida. Super wide viewer using catadioptrical optics. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '03, pp. 169–175. ACM, New York, NY, USA, 2003. doi: 10.1145/1008653.1008683

[43] R. Narain, R. A. Albert, A. Bulbul, G. J. Ward, M. S. Banks, and J. F. O'Brien. Optimal presentation of imagery with focus cues on multi-plane displays. *ACM Trans. Graph.*, 34(4):59:1–59:12, July 2015. doi: 10.1145/2766909

[44] J. Notaros, M. Raval, M. Notaros, and M. R. Watts. Integrated-phased-array-based visible-light near-eye holographic projector. In *2019 Conference on Lasers and Electro-Optics (CLEO)*, pp. 1–2, 2019.

[45] N. Padmanaban, Y. Peng, and G. Wetzstein. Holographic near-eye displays based on overlap-add stereograms. *ACM Trans. Graph.*, 38(6), Nov. 2019. doi: 10.1145/3355089.3356517

[46] V. F. Pamplona, M. M. Oliveira, D. G. Aliaga, and R. Raskar. Tailored displays to compensate for visual aberrations. *ACM Trans. Graph.*, 31(4):81:1–81:12, July 2012. doi: 10.1145/2185520.2185577

[47] M. Parikh. Corrections to proximity effects in electron beam lithography. I. theory. *Journal of Applied Physics*, 50(6):4371–4377, 1979.

[48] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014. doi: 10.1561/2400000003

[49] M. Peckerar, R. Bass, and K. W. Rhee. Sub-0.1 μ electron-beam lithography for nanostructure development. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, 18(6):3143–3149, 2000.

[50] M. Peckerar, D. Sander, A. Srivastava, A. Foli, and U. Vishkin. Electron beam and optical proximity effect reduction for nanolithography: New results. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, 25(6):2288–2294, 2007.

[51] M. C. Peckerar, S. Chang, and C. R. K. Marrian. Proximity correction algorithms and a co-processor based on regularized optimization. I. description of the algorithm. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, 13(6):2518–2525, 1995. doi: 10.1116/1.588385

[52] Y. Peng, S. Choi, N. Padmanaban, J. Kim, and G. Wetzstein. Neural holography with camera-in-the-loop training. *ACM Transactions on Graphics (TOG)*, 39(6), 2020.

[53] M. Persson, D. Engström, and M. Goksör. Reducing the effect of pixel crosstalk in phase only spatial light modulators. *Optics express*, 20(20):22334–22343, 2012.

[54] C. V. Poulton, A. Yaacobi, D. B. Cole, M. J. Byrd, M. Raval, D. Vermeulen, and M. R. Watts. Coherent solid-state lidar with silicon photonic optical phased arrays. *Opt. Lett.*, 42(20):4091–4094, Oct 2017. doi: 10.1364/OL.42.004091

[55] K. Rathinavel, H. Wang, A. Blate, and H. Fuchs. An extended depth-at-field volumetric near-eye augmented reality display. *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2857–2866, 2018.

[56] K. Rathinavel, G. Wetzstein, and H. Fuchs. Varifocal occlusion-capable optical see-through augmented reality display based on focus-tunable optics. *IEEE Transactions on Visualization and Computer Graphics*, 25(11):3125–3134, 2019.

[57] M. Raval, A. Yaacobi, and M. R. Watts. Integrated visible light phased array system for autostereoscopic image projection. *Opt. Lett.*, 43(15):3678–3681, Aug 2018. doi: 10.1364/OL.43.003678

[58] J. P. Rolland, M. W. Krueger, and A. Goon. Multifocal planes head-mounted displays. *Appl. Opt.*, 39(19):3209–3215, Jul 2000. doi: 10.1364/AO.39.003209

[59] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11):3440–3451, Nov 2006. doi: 10.1109/TIP.2006.881959

[60] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. Live image quality assessment database release 2. `http://live.ece.utexas.edu/research/quality`, 2004. Accessed: 2019-04-03.

[61] L. Shi, F.-C. Huang, W. Lopes, W. Matusik, and D. Luebke. 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.

[62] T. Smits and W. Faber. Chronic (classified historical newspaper images). KB Lab: The Hague. `http://lab.kb.nl/dataset/chronic-classified-historical-newspaper-images`, 2018.

[63] J. Sun, E. Timurdogan, A. Yaacobi, E. S. Hosseini, and M. R. Watts. Large-scale nanophotonic phased array. *Nature*, 493(7431):195, 2013.

[64] X. Sun, Y. Zhang, P.-C. Huang, M. Dagenais, M. Peckerar, and A. Varshney. Proximity effect on nanophotonic phased arrays. In *Frontiers in Optics + Laser Science APS/DLS*, p. JW3A.101. Optical Society of America, 2019. doi: 10.1364/FIO.2019.JW3A.101

[65] J. P. Waters. Holographic image synthesis utilizing theoretical methods. *Applied physics letters*, 9(11):405–407, 1966.

[66] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar. Tensor displays: Compressive light field synthesis using multilayer displays with directional backlighting. *ACM Trans. Graph.*, 31(4), July 2012. doi: 10.1145/2185520.2185576

[67] X. Xia, Y. Guan, A. State, P. Chakravarthula, K. Rathinavel, T. Cham, and H. Fuchs. Towards a switchable ar/vr near-eye display with accommodation-vergence and eyeglass prescription support. *IEEE Transactions on Visualization and Computer Graphics*, 25(11):3114–3124, 2019.

[68] S. Xu, F. Farbiz, S. Solanki, X. Liang, and X. Xu. Adaptive computation of computer-generated holograms. In *Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, VRCAI '08, pp. 35:1–35:2. ACM, New York, NY, USA, 2008. doi: 10.1145/1477862.1477908

[69] H. Yoshikawa, S. Iwase, and T. Oneda. Fast computation of fresnel holograms employing difference. In *Practical Holography XIV and Holographic Materials VI*, vol. 3956, pp. 48–56. International Society for Optics and Photonics, 2000.

[70] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018.

[71] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004. doi: 10.1109/TIP.2003.819861

[72] R. Ziegler, S. Bucheli, L. Ahrenberg, M. Magnor, and M. Gross. A bidirectional light field-hologram transform. In *Computer Graphics Forum*, vol. 26, pp. 435–446. Wiley Online Library, 2007.

[73] R. Ziegler, S. Croci, and M. Gross. Lighting and occlusion in a wave-based framework. In *Computer Graphics Forum*, vol. 27, pp. 211–220. Wiley Online Library, 2008.