

Focal Surface Displays

NATHAN MATSUDA, Oculus Research

ALEXANDER FIX, Oculus Research

DOUGLAS LANMAN, Oculus Research

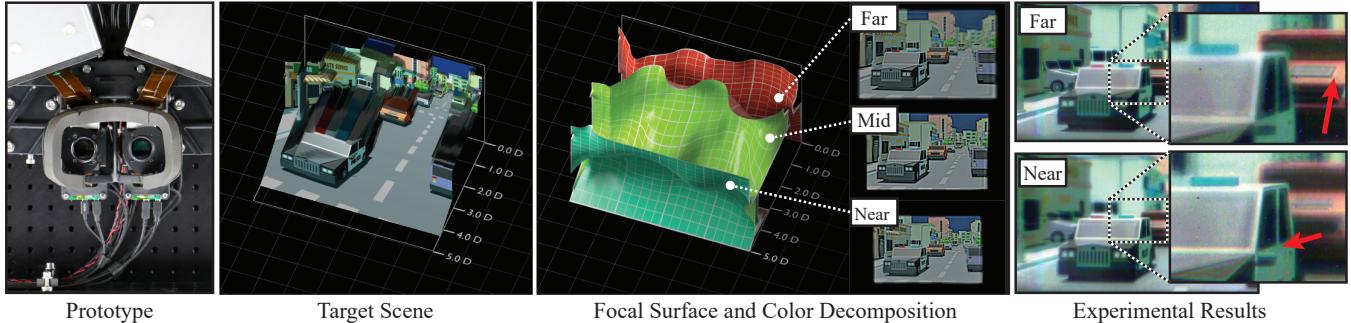


Fig. 1. A *focal surface display* augments a conventional head-mounted display (HMD) with a spatially programmable focusing element placed between the eyepiece and the underlying color display. This design depicts sharp imagery and near-correct retinal blur, across an extended depth of focus, by shaping one or more virtual images to conform to the scene geometry. Our rendering algorithm uses an input focal stack and depth map to produce a set of phase functions (displayed on a phase SLM) and RGB images (displayed on an OLED). Shown from left to right are: a prototype using a liquid crystal on silicon (LCOS) phase modulator, a target scene, the optimized focal surfaces and color images displayed across three frames, and the resulting focal stack, sampled at a near and far focus of 4.0 and 0.0 diopters, respectively. (Source imagery courtesy Unity Asset Store publisher “VenCreations.”)

Conventional binocular head-mounted displays (HMDs) vary the stimulus to vergence with the information in the picture, while the stimulus to accommodation remains fixed at the apparent distance of the display, as created by the viewing optics. Sustained vergence-accommodation conflict (VAC) has been associated with visual discomfort, motivating numerous proposals for delivering near-correct accommodation cues. We introduce *focal surface displays* to meet this challenge, augmenting conventional HMDs with a phase-only spatial light modulator (SLM) placed between the display screen and viewing optics. This SLM acts as a dynamic freeform lens, shaping synthesized focal surfaces to conform to the virtual scene geometry. We introduce a framework to decompose target focal stacks and depth maps into one or more pairs of piecewise smooth focal surfaces and underlying display images. We build on recent developments in “optimized blending” to implement a multifocal display that allows the accurate depiction of occluding, semi-transparent, and reflective objects. Practical benefits over prior accommodation-supporting HMDs are demonstrated using a binocular focal surface display employing a liquid crystal on silicon (LCOS) phase SLM and an organic light-emitting diode (OLED) display.

CCS Concepts: • Computing methodologies → Virtual reality;

Additional Key Words and Phrases: head-mounted displays, multifocal displays, caustics, freeform optics, vergence-accommodation conflict

ACM Reference format:

Nathan Matsuda, Alexander Fix, and Douglas Lanman. 2017. Focal Surface Displays. *ACM Trans. Graph.* 36, 4, Article 86 (July 2017), 14 pages.
DOI: <http://dx.doi.org/10.1145/3072959.3073590>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2017 ACM. 0730-0301/2017/7-ART86 \$15.00
DOI: <http://dx.doi.org/10.1145/3072959.3073590>

1 INTRODUCTION

A modern head-mounted display (HMD), as designed for virtual reality (VR) applications, is a simple construction placing viewing optics (e.g., a magnifying lens) between the user’s eye and a display screen. This configuration is replicated for binocular stereo configurations: one set of optics and one display, or portion of a display, is dedicated to each eye. In this manner, a binocular HMD depicts stereoscopic imagery such that the user perceives virtual objects with correct retinal disparity, which is the critical stimulus to vergence (the degree to which the eyes are converged or diverged to fixate a point) [Peli 1999].

VR viewing optics typically create a virtual, erect, magnified image of the display screen, located at a fixed focal distance from the user [Cakmakci and Rolland 2006]. Thus, current VR HMDs do not correctly depict retinal blur, which is the critical stimulus to accommodation (the eyes’ focusing response). The resulting vergence-accommodation conflict (VAC) has been attributed as a source of visual discomfort: viewers report eye strain, blurred vision, and headaches with prolonged viewing [Shibata et al. 2011]. VAC has also been linked to perceptual consequences, affecting eye movements and the ability to resolve depth [Hoffman et al. 2008].

A multitude of “accommodation-supporting” HMD architectures have been proposed to depict correct or near-correct retinal blur, thereby mitigating VAC (see Table 1). As surveyed by Kramida [2016], these architectures are distinguished by the fidelity to which they synthesize retinal blur. At one end of the spectrum are designs that effectively extend the user’s depth of focus (DOF), allowing the virtual image to remain sharp, independent of the user’s accommodative state. This includes varifocal displays that dynamically adjust the focus of the HMD, contingent on the detected eye gaze. While addressing blurred vision induced by VAC, such displays

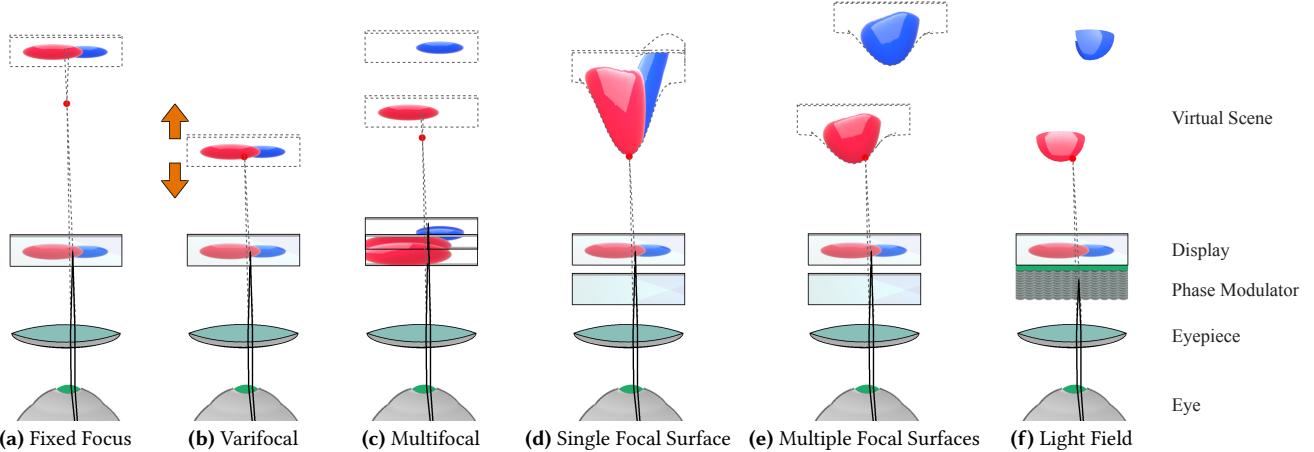


Fig. 2. Focal surface displays generalize the concept of manipulating the optical focus of each pixel on an HMD. Configurations (d,e) augment a fixed-focus HMD (a) with a programmable phase modulator placed between the eyepiece and display. (b) Varifocal HMDs use a globally addressed tunable lens. (c) Multifocal displays may use a high-speed tunable lens and display to create multiple focal planes. (f) In contrast, certain light field HMD concepts fall at the other end of this spectrum, using a finely structured phase modulator (a microlens array) placed near the display. (d) In this paper, we consider designs existing between these extremes in which a phase modulator locally adjusts the focus to follow the virtual geometry, generalizing varifocal and multifocal concepts. (e) Similar to multifocal displays, multiple focal surfaces can be synthesized with high-speed phase modulators and displays.

cannot correctly depict retinal blur; instead, blur can only be synthetically rendered. At the other end of the spectrum are designs that correctly reproduce the optical wavefront of a physical scene, including holographic displays and, in certain circumstances, light field displays. As reported by Kramida, such displays are not yet practical, due to the limited resolution, field of view, and image quality achievable with today’s hardware. As a result, a third category of accommodation-supporting HMDs is under active investigation: those that create “near-correct” (approximated) retinal blur.

Approximate blur for multifocal displays has been studied extensively. Multifocal displays consist of a superposition of multiple virtual images spanning a range of focal depths. The first multifocal prototype employed three separate display elements per eye, prohibiting head-mounted configurations [Akeley et al. 2004]. As reviewed in Section 2.1, multifocal HMDs increasingly exploit time-multiplexed presentation, wherein a single high-refresh-rate display and a fast varifocal element sequentially address the image planes [Liu et al. 2010]. Despite wide investigation, multifocal displays continue to present numerous practical challenges. First, as established by MacKenzie et al. [2012], focal plane separation must be as close as 0.6 diopters to correctly stimulate accommodation. Thus, five focal planes are required to span a working range of 3.0 diopters (supporting virtual scenes extending from 33 cm to optical infinity). In practice, flickering is likely perceived with this many focal planes, due to the refresh rates of microdisplays currently used in HMD designs. Second, as investigated by Narain et al. [2015], the lateral spatial resolution of virtual objects presented between focal planes is restricted, demanding yet more planes to achieve the desired 3D resolution. Recently, Wu et al. [2016] proposed dynamically adapting focal plane separations to virtual content, effectively combining the varifocal and multifocal concepts to reduce the number of required image planes.

In this paper, we expand on the concept of an adaptive multifocal display, introducing *focal surface displays* in which a *spatially addressable* phase modulator is inserted into an otherwise conventional HMD. Following Figure 1, the phase modulator shapes focal surfaces to conform to the scene geometry, unlike multifocal displays with fixed, typically planar, focal surfaces. We produce a set of color images which are each mapped onto a corresponding focal surface. Visual appearance is rendered by tracing rays from the eye through the optics, and accumulating the color values for each focal surface. Our algorithm sequentially solves for first the focal surfaces, given the target depth map, and then the color images—full joint optimization is left for future work. Focal surfaces are adapted by nonlinear least squares optimization, minimizing the distance between the nearest depicted surface and the scene geometry. The color images, paired with each surface, are determined by linear least squares methods. Using databases of natural and rendered scenes, we demonstrate that focal surface displays depict more accurate retinal blur, with fewer multiplexed images, than prior multifocal displays, while maintaining high resolution throughout the user’s accommodative range. Through focal surface displays we aim to extend the technological development path beyond prior varifocal and multifocal concepts, opening a new point in the design tradespace of accommodation-supporting HMDs.

1.1 Contributions

Our primary technical contributions are:

- We introduce focal surface displays, capable of depicting near-correct focus cues in head-mounted displays, and assess capabilities relative to prior accommodation-supporting HMDs, including related multifocal architectures.
- We introduce an optimization framework that decomposes target focal stacks and depth maps into one or more pairs of focal surfaces and color images. Our pipelined approach

	resolution [†]	FOV [‡]	eye box [‡]	DOF	eye tracking required	adaptive optics	content-dependent optimization	image quality [‡]	retinal blur class
fixed focus [Cakmakci and Rolland 2006]	high	wide	wide	narrow	no	no	no	high	incorrect
monovision [Johnson et al. 2016; Konrad et al. 2016]	high	wide	wide	moderate	no	no	no	moderate	incorrect
varifocal [Dunn et al. 2017; Padmanaban et al. 2017]	high	wide	wide	wide	yes	yes	no	high	rendered
EDOF: Maxwellian view [von Waldkirch et al. 2004]	high	narrow	narrow	moderate	no	no	no	high	rendered
EDOF: focal sweep [von Waldkirch 2005]	moderate	narrow	narrow	moderate	no	optional	optional	low	rendered
fixed multifocal [Hu and Hua 2014; Narain et al. 2015]	moderate	moderate	narrow	moderate	yes	no	yes	moderate	near-correct
adaptive multifocal [Llull et al. 2015; Wu et al. 2016]	high	narrow	narrow	wide	yes	yes	yes	high	near-correct
retinal scanning [McQuade et al. 2003]	high	moderate	narrow	wide	no	yes	no	low	near-correct
light field: layered attenuators [Huang et al. 2015]	low	wide	moderate	wide	optional	no	yes	moderate	near-correct
light field: integral imaging [Lanman and Luebke 2013]	moderate	narrow	moderate	moderate	no	no	no	high	near-correct
holographic [Moon et al. 2014]	high	narrow	narrow	wide	no	yes	no	moderate	correct
focal surface displays	high	narrow	narrow	wide	yes	yes	yes	high	near-correct

Table 1. Accommodation-supporting displays are assessed relative to optical and perceptual criteria. [†]Resolution is listed according to theoretical upper bounds (e.g., diffraction limits). [‡]Field of view, eye box dimensions, and image quality depend on implementation choices: listed values correspond to the performance of prototypes in the cited publications, being indicative of current display technology limitations. Note that “moderate” resolution, field of view (FOV), eye box width, and depth of focus (DOF) are defined, respectively, as: 10–20 cycles per degree (cpd), 40–80 degrees, 5–10 mm, and 1–3 diopters. Excursions above or below these ranges are shaded green or red, respectively. Form factors are not compared, as most concepts are currently embodied by early stage prototypes not optimized for size or weight.

finds focal surfaces through nonlinear least squares optimization and color images by linear least squares methods.

- Through a first-order optical analysis, we describe the optimal construction of focal surface displays, assessing trade-offs between resolution, field of view, and depth of focus. Furthermore, we identify the benefit of extending the SLM phase modulation range to enable high-resolution display.
- We implement a binocular focal surface display prototype, employing one LCOS spatial light modulator and one OLED panel per eye. We assess its experimental performance in relation to geometric optical simulations.

2 RELATED WORK

Focal surface displays draw on insights spanning accommodation-supporting HMDs, goal-based caustics, as well as freeform and adaptive optics. Our prototype and development of this architecture is largely presented with regard to VR HMDs. However, as discussed in Section 6, there is a clear extension to certain augmented reality (AR) systems, particularly projector-based configurations. For reviews of existing VR and AR designs, consult Cakmakci and Rolland [2006] and Kress and Starner [2013], respectively.

2.1 Accommodation-Supporting Displays

As summarized in Table 1, any HMD can be evaluated relative to standard criteria, including resolution, field of view (FOV), and eye box dimensions. Today’s VR HMDs exhibit FOVs around 100 degrees with resolutions better than 5 cycles per degree (cpd). Emerging designs must ultimately support such specifications and beyond. Accommodation-supporting HMDs may further be evaluated in regard to their depth of focus (DOF) and the fidelity to which retinal blur is reproduced. Many designs require eye tracking, which introduces concerns about reliability that must be weighed against others. Additionally, emerging HMDs increasingly exploit adaptive optics, particularly tunable lenses (see Figure 2). Some schemes may leverage computational display concepts and can be judged on additional axes, including image quality (which may be limited due to compression artifacts) and the failure modes and computational

complexity resulting from content-dependent optimization. In this section, we review prior accommodation-supporting HMDs relative to these criteria, showing that focal surface displays expose a new, promising point in the design tradespace.

2.1.1 Monovision Displays. Marran and Schor [1997] provide a prior survey of accommodation-supporting HMDs. One configuration they assess is that of *monovision*, wherein the virtual image distance differs between the eyes. This configuration is inspired by a related optometric application by which presbyopia is addressed by placing the focus of one eye closer than the other. Recently, Johnson et al. [2016] and Konrad et al. [2016] assessed the performance of monovision HMDs. The former study found viewer comfort and visual performance did not improve, whereas the latter found some benefit. However, not all viewers prefer or eventually adapt to monovision, motivating the need for more widely applicable methods.

2.1.2 Varifocal Displays. Varifocal HMDs augment a conventional design with two components: an eye tracker and a variable focusing element. Eye tracking is used in a feedback system to dynamically set the tunable lens focus to match vergence, thus ensuring VAC is minimized. Shiwa et al. [1996] first demonstrated this concept using actuated lenses on an optical bench. Sugihara et al. [1998] created the first varifocal HMD, wherein the display translated rather than a lens. Liu et al. [2010] and Konrad et al. [2016] demonstrated varifocal displays using electronically tunable lenses. Recently, Dunn et al. [2017] and Padmanaban et al. [2017] presented varifocal displays with integrated eye tracking.

Varifocal displays may reduce VAC, but they cannot directly reproduce retinal blur. Gaze-contingent depth of field (DOF) rendering must be applied. Hillaire et al. [2008] and Mantiuk et al. [2011a] conclude that DOF blur is preferred with 2D displays. Duchowski et al. [2014] found that visual discomfort was reduced when viewing a stereoscopic display with gaze-contingent DOF blur, albeit with a statistically weak dislike for this blurring. Our interpretation of this result is that it highlights the limitations of rendered blur: latency and eye-tracking errors may create distracting artifacts, motivating the development of accommodation-supporting HMDs that support

near-correct, rather than rendered, retinal blur. Perceptual studies by Maiello et al. [2015] and Zannoli et al. [2016] have found that synthetically rendered blur may not assist depth perception to the same degree as near-correct retinal blur.

2.1.3 Accommodation-Invariant (EDOF) Displays. For HMDs, the analogue of a pinhole camera is a Maxwellian view: a point light source is focused on the viewer’s pupil, with an amplitude SLM modulating a focused image on the retina [Burns and Webb 2010]. Von Waldkirch et al. [2004] apply this principle to HMDs, showing a trade between DOF and resolution. Due to diffraction, DOF cannot extend above three diopters without restricting resolution below 30 cpd (i.e., 20/20 vision) [Jacobs et al. 1992]. Following Kramida [2016], FOV is limited due to restricted eye movement.

Maxwellian-view HMDs exhibit an accommodation-invariant response. In computational photography, this is known as extended depth of focus (EDOF) [Zalevsky 2010]. Von Waldkirch [2005] applied EDOF to HMDs, rapidly varying focus with a tunable lens; however, deconvolution was not considered and, as a result, image contrast was reduced. More recently, Huang et al. [2012] applied pre-filtering to a multilayer EDOF display, although contrast remained low. Even if image quality can be improved, accommodation-invariant HMDs still rely on rendered retinal blur.

2.1.4 Multifocal Displays. To our knowledge, Neil et al. [1997] proposed, and demonstrated, the first multifocal HMD. As they describe, the concept is preceded by decades of research into volumetric displays [Blundell and Schwartz 1999]. Rolland et al. [2000] proposed a closely related architecture, assessing that a 2.0-diopter DOF requires up to 27 planes. Even this may not be sufficient: measurements by Sprague et al. [2015] find an average 40-year-old or younger individual can accommodate in excess of 4.0 diopters.

MacKenzie et al. [2012] show that wider plane separations can correctly drive accommodation; however, maintaining high resolution between planes and extending DOF can only be achieved, currently, with additional adaptive optical elements [Wu et al. 2016]. Multifocal adaptive optics include ferroelectric liquid crystal (FLC) SLMs [Love et al. 2009; Neil et al. 1997], tunable lenses [Konrad et al. 2016; Liu et al. 2010; Wu et al. 2016], and deformable mirrors (DMs) [Hu and Hua 2014]. Focal surface displays leverage this trend for increasing electro-optic control, preparing for a future in which spatially-varying phase modulation is widely available.

Akeley et al. [2004] first considered the optimal presentation of imagery across multiple focal planes, introducing the “linear blending” algorithm. Ravikumar et al. [2011] assessed alternative algorithms, concluding that, of those available at the time, linear blending was preferred. More recently, Narain et al. [2015] introduced “optimized blending” to directly optimize the through-focus image, enhancing occluding, semi-transparent, and reflective objects. In this work, we generalize optimized blending to support adaptive focal surfaces.

2.1.5 Retinal Scanning Displays. Rather than using comparatively large screens, retinal scanning displays (RSDs) directly sweep a point of light across the viewer’s retina [Viirre et al. 1998]. McQuaide et al. [2003] modify RSDs to additionally modulate focus using a deformable mirror (DM). Unlike varifocal HMDs, focus can be adjusted—in theory—individually per pixel. This concept is a

precursor to focal surface displays; however, to our knowledge, it was never fully realized: deformable mirrors exhibit a modulation rate three orders of magnitude too slow for per-pixel focus control. Correspondingly, McQuaide et al. only demonstrate simple line images, albeit over a continuously-varying 3.0-diopter DOF.

Focal surface displays significantly differ from accommodation-supporting RSDs. First, we provide an optimization framework to tailor focal surfaces that respects the constraints of current phase SLM technology. Second, our framework allows multiple focal surfaces, yielding near-correct depictions of occlusions. Third, we leverage work on optimized blending for multifocal displays to account for limitations of focal surface control. Fourth, we demonstrate the first fully-realized embodiment with a binocular LCOS-based prototype capable of depicting natural scenes.

2.1.6 Light Field Displays. Volumetric displays inspired multifocal displays. Similarly, near-eye light field displays originate from the autostereoscopic community. Lanman and Luebke [2013] first applied integral imaging to VR HMDs, with a closely related AR HMD developed by Hua and Javidi [2014]. While depicting near-correct retinal blur, these prototypes exhibit low resolution, albeit while additionally depicting correct parallax across the eye box. Mai-mone et al. [2013] and Huang et al. [2015] introduced computational near-eye light field displays, for AR and VR, respectively, based on amplitude-only SLM stacks (i.e., multilayer LCDs). Following Table 1, such displays confront practical resolution limits due to diffraction and compression artifacts. Our multilayer focal surface display does not exhibit a similar limit due to the comparatively high fill factor and lack of color filters with LCOS panels.

2.1.7 Holographic Displays. Decades of research into direct-view holography has laid the groundwork for near-eye applications [Bove 2012]. Today’s digital holographic displays synthesize accurate wavefronts, and therefore correct retinal blur, by controlled illumination of a diffractive element. Moon et al. [2014] describe a recent holographic HMD, showing practical limits on FOV (less than 20 degrees), eye box dimensions (a few millimeters wide), and image quality (degraded due to speckle). Focal surface displays, which may incorporate similar phase modulators, fundamentally differ: incoherent illumination is produced by an emissive display, with subsequent modulation by a phase-only SLM that produces piecewise smooth modulations; furthermore, such displays require minimal modification to existing VR HMDs.

2.2 Caustics, Freeform Elements, and Adaptive Optics

Focal surface displays also trace their origin to recent progress in computational fabrication and adaptive optics. In a closely related work, Damberg et al. [2016] use a phase-only SLM to create a freeform adaptive lens for the purpose of high dynamic range (HDR) projection. Damberg et al. adapt prior research into *goal-based caustics*, wherein freeform lenses are fabricated to project images under controlled illumination [Papas et al. 2011; Yue et al. 2014]. Phase-only SLMs have been similarly adopted by the computational display community, with Glasner et al. [2014] and Levin et al. [2016] demonstrating their application to light-sensitive multi-view displays. To our knowledge, focal surface displays are the first application of phase SLMs to locally adapt the focus of an HMD.

3 FOCAL SURFACE DISPLAYS

A conventional VR HMD contains two primary optical elements: an eyepiece and an emissive display. This design delivers a single, fixed *focal surface*. As shown in Figure 3, a *focal surface display* adds a third element between the eyepiece and the display: a phase-modifying spatial light modulator (SLM). This SLM acts as a programmable lens with spatially varying focal length, allowing the virtual image of different display pixels to be formed at different depths. In this section, we present an optimization framework that decomposes a scene into one or more focal surfaces, and corresponding color images, to reproduce retinal blur consistent with natural scenes.

Inspired by related multifocal displays, we generalize our formulation to support multiple focal surfaces (as achieved by time multiplexing). The inputs to our algorithm are a depth map, representing the scene geometry, and a focal stack, modeling the variation of retinal blur with changes in accommodation. Both inputs are rendered from the perspective of the viewer's entrance pupil. The outputs are k phase functions ϕ_1, \dots, ϕ_k and color images c_1, \dots, c_k , to be presented by the SLM and underlying display, respectively. Ideally, we would jointly optimize the phase functions and color images. Because this results in a large, nonlinear problem, we introduce approximations that ensure the algorithm is computationally tractable. First, in Section 3.1, we decompose the target depth map into a set of smooth focal surfaces. Second, in Section 3.2, we optimize the phase functions to approximate these focal surfaces. Finally, in Section 3.3, we optimize the color images to reproduce the target focal stack.

While our formulation allows multiple focal surfaces, a single surface achieves similar retinal blur fidelity as prior multifocal displays. As with other computational displays, focal surface displays offer a trade-off between system complexity (the need for time multiplexing) and image quality (suppression of compression artifacts).

3.1 Approximating Depth Maps with Focal Surfaces

Given a target virtual scene, let $\hat{d}(\theta_x, \theta_y)$ be the depth (in diopters) along each viewing angle $(\theta_x, \theta_y) \in \Omega_\theta$, for *chief rays* passing through the center of the viewer's pupil and with Ω_θ being the discrete set of retinal image samples. If phase SLMs could render focal surfaces with arbitrary topology, then no further optimization would be required. As presented in Section 3.2, this is not the case: practically realizable focal surfaces are required to be smooth. Correspondingly, we develop the following method for decomposing a depth map into k smooth focal surfaces d_1, \dots, d_k .

For every viewing angle (θ_x, θ_y) we desire at least one focal surface $d_i(\theta_x, \theta_y)$ to be close to the target depth map $\hat{d}(\theta_x, \theta_y)$. If this occurs, then every scene element can be depicted with near-correct retinal blur, as light from the underlying display will appear to originate from the correct scene depth. (As established by Narain et al. [2015], optimized blending methods still benefit the rendition of occluding, semi-transparent, and reflective objects.) Given this goal, we formulate the following optimization problem.

$$\begin{aligned} \min_{d_1, \dots, d_k} & \sum_{(\theta_x, \theta_y) \in \Omega_\theta} \left(\min_i |\hat{d}(\theta_x, \theta_y) - d_i(\theta_x, \theta_y)| \right)^2 \\ \text{s.t. } & \left(\frac{\partial^2 d_i}{\partial x^2} \right)^2 + \left(\frac{\partial^2 d_i}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 d_i}{\partial y^2} \right)^2 < \epsilon \end{aligned} \quad (1)$$

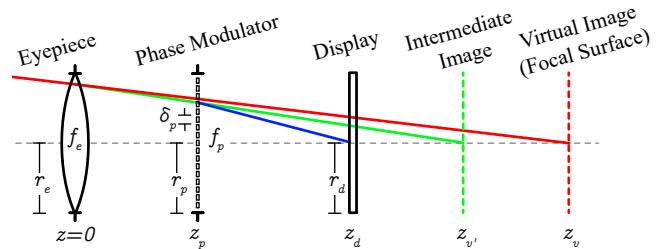


Fig. 3. A focal surface display is created by placing a phase-modulation element between an eyepiece and a display screen. This phase element and the eyepiece work in concert as a spatially programmable compound lens, varying the apparent virtual image distance across the viewer's field of view.

As analyzed in Section 3.2, synthesizing a focal surface using phase function ϕ may introduce some optical aberrations. Observationally, we find aberrations are minimized if the second derivatives of the focal surface are small. This observation is reflected by the bound constraints in our optimization problem. Note, however, that no explicit bound constraints are imposed on the optical powers d_i of the focal surfaces. This would appear to contradict our derivation of the minimum realizable focal length of a given phase SLM (see Section 3.2). Rather than adding these constraints directly, we simply truncate the target depth map \hat{d} to the realizable range.

We apply nonlinear least squares (NLS) to solve Equation 1, which has high-quality implementations and scales to large problem sizes [Agarwal and Others 2012]. Note that our objective involves the nonlinear residual $g_{\theta_x, \theta_y}(d) := \min_i |\hat{d}(\theta_x, \theta_y) - d_i(\theta_x, \theta_y)|$ for each pixel (θ_x, θ_y) . This residual is not differentiable, which is a problem for NLS. However, a close approximation is obtained by replacing the min with a “soft minimum” (soft-min), with the following definition [Cook 2010]:

$$\tilde{g}_{\theta_x, \theta_y}(d) = -t \log \sum_i e^{-|\hat{d}(\theta_x, \theta_y) - d_i(\theta_x, \theta_y)|/t}, \quad (2)$$

where t is a conditioning parameter to be tuned for a given application. Note that \tilde{g} is continuously differentiable and closely approximates g as $t \rightarrow 0$, with $|\tilde{g}(\theta_x, \theta_y) - g(\theta_x, \theta_y)| \leq t \log k$.¹

Applying Equation 2 to Equation 1, and re-expressing bound constraints as soft constraints, yields the following NLS problem:

$$\min_{d_1, \dots, d_k} \sum_{(\theta_x, \theta_y)} (\tilde{g}_{\theta_x, \theta_y}(d))^2 + \gamma \sum_{i, (\theta_x, \theta_y)} \|\partial^2 d_i(\theta_x, \theta_y)\|^2, \quad (3)$$

where $\partial^2 d_i(\theta_x, \theta_y)$ is the vector of second partial derivatives of d_i at (θ_x, θ_y) and γ is a weighting parameter. See Figures 4 and 5 for examples of applying this focal surface decomposition algorithm. As shown, locally adapted smooth focal surfaces offer an efficient representation of natural and artificially rendered depth maps.

3.2 Synthesizing Focal Surfaces with Phase SLMs

Provided a set of focal surfaces d_i , the next stage in our pipeline requires solving for a set of phase functions ϕ_i to practically achieve

¹Note that when computing a soft-min, for numerical stability it is important to use the method described by Cook [2010], wherein the minimum value is subtracted before evaluating the exponential functions.

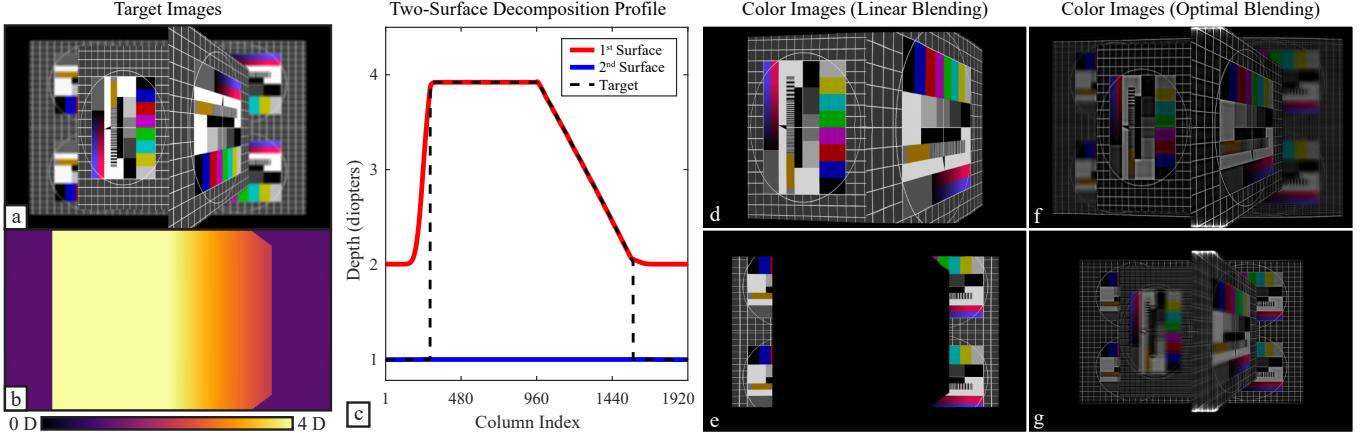


Fig. 4. A focal surface decomposition is presented above for a simple scene, containing: a background fronto-parallel plane at 1.0 diopters, a foreground fronto-parallel plane at 4.0 diopters, and a slanted plane spanning 2.0 to 4.0 diopters. (a) A single image from the target focal stack. (b) The target depth map. (c) A two-surface decomposition is compared to the target depth map for a profile taken along the middle row of the target imagery. (d,e) The color images associated with each focal surface are shown, using the linear blending method of Akeley et al. [2004]. Note that this decomposition produces clearly delineated “foreground” and “background” components. (f,g) The color images associated with each focal surface, using the optimized blending algorithm presented in Section 3.3. Note that, similar to Narain et al. [2015], the resulting color images present high spatial frequencies closer to the focal surface near where they occur in the target scene. We emphasize that, as with other multifocal displays reviewed in Section 2, time-multiplexed focal surface display reduces brightness, due to color image components being presented for a briefer duration than that occurring with a fixed-focus display mode.

them. To solve this problem, we first review the optical properties of phase SLMs and then present our phase optimization procedure.

3.2.1 Optical Properties of Phase SLMs. Variations in optical path length through a lens cause refraction. Similarly, differences in phase modulation across an SLM result in diffraction. Simulation of light propagation through a high-resolution SLM, via wave optics modeling, is currently computationally infeasible, but one can approximate these diffractive effects using geometric optics, similar to Glasner et al. [2014] and Damberg et al. [2016]. (Laude [1998] provides additional details regarding the operation of phase SLMs.) We denote SLM locations by (p_x, p_y) , with Ω_p being the discrete set of SLM pixel centers. Optical rays intersecting an SLM are redirected depending on the phase ϕ . For small angles (i.e., under the paraxial approximation), the deflection is proportional to the gradient of ϕ (see [Voelz 2011], Equation 6.1). If an incident ray has direction vector $(x, y, 1)$ and intersects the SLM at (p_x, p_y) , then the outgoing ray has direction vector

$$\left(x + \frac{\lambda}{2\pi} \frac{\partial \phi}{\partial x}(p_x, p_y), y + \frac{\lambda}{2\pi} \frac{\partial \phi}{\partial y}(p_x, p_y), 1 \right), \quad (4)$$

where λ is the illumination wavelength. Thus, if ϕ is a linear function, then the SLM operates as a prism, adding a constant offset to the direction of every ray. (Note that we assume monochromatic illumination in this derivation, with practical considerations for broadband illumination sources presented later in Section 6.1.) An SLM may also act as a thin lens (see [Voelz 2011], Equation 6.8) by presenting a quadratically varying phase as follows.

$$\phi(p_x, p_y) = -\frac{\pi}{\lambda f} (p_x^2 + p_y^2) \quad (5)$$

Note that these optical properties are local. The deflection of a single ray only depends on the first-order Taylor series of the phase (i.e., the phase gradient) around the point of intersection

with the SLM. Similarly, the change in focus of an ϵ -sized bundle of rays intersecting the SLM only depends on the second-order Taylor series. Specifically, if the Hessian of ϕ at a point (p_x, p_y) is given by

$$H_\phi(p_x, p_y) = -\frac{2\pi}{\lambda f} I, \quad (6)$$

where I is the 2×2 identity matrix, then the ϵ -sized neighborhood around (p_x, p_y) functions as a lens of focal length f (i.e., Equation 6 is the Hessian of Equation 5).

To this point, we have allowed the phase to be any real-valued function. In practice, an SLM will have a bounded range, typically from $[0, 2\pi]$. Phases outside this range are “wrapped”, modulo 2π . In addition, achievable phase functions are restricted by the Nyquist limit. The phase can change by no more 2π over a distance of $2\delta_p$, where δ_p is the SLM pixel pitch. Following Voelz [2011], these factors bound the minimum focal length f such that $|f| \geq \frac{2r_p \delta_p}{\lambda}$, where r_p is the radius of the SLM (taken diagonally).

3.2.2 Adapting Focal Surfaces with Phase SLMs. With this paraxial model of an SLM, we can determine a phase function ϕ to best realize a given target focal surface d . First, we must determine how the SLM focal length f_p (synthesized via Equation 5) affects a focal surface distance z_v . As indicated in Figure 3, the SLM acts within a focal surface display that is parameterized by the eyepiece distance ($z=0$), the SLM distance z_p , and the display distance z_d . Ignoring the eyepiece, the SLM produces an intermediate image of the display at distance $z_{v'}$. This intermediate image is transformed to a virtual image of the display, located at z_v , depending on the eyepiece focal length f_e . These relations are compactly summarized by application of the thin lens equation (see [Voelz 2011], Equation 7.1):

$$\frac{1}{f_p} = \frac{1}{z_{v'} - z_p} + \frac{1}{z_d - z_p} \text{ and } \frac{1}{f_e} = \frac{1}{z_v} - \frac{1}{z_{v'}}. \quad (7)$$

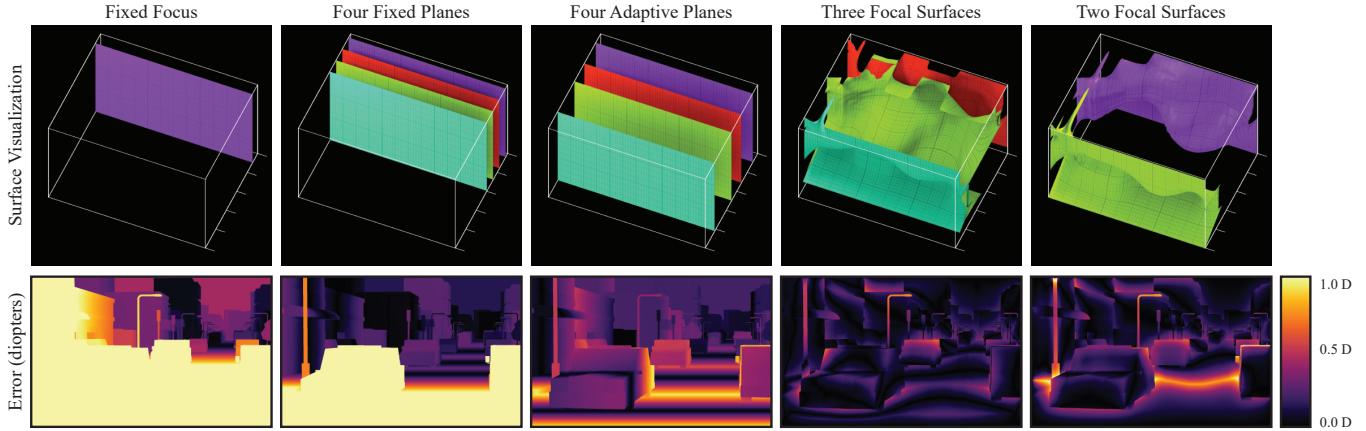


Fig. 5. Focal surface displays achieve lower depth map approximation errors, using less time multiplexing, than prior multifocal methods. As a result, such displays can support higher resolution image content (see Section 3.2). The upper row visualizes the optimized focal surfaces ranging from 0.0 to 5.0 diopters, abbreviated “D”. The lower row depicts the resulting depth map approximation errors in diopters. For a fixed focus design, the virtual image is positioned at 0.5 D. Following Narain et al. [2015], the fixed multifocal display employs four planes evenly spaced from 0.2 D to 2.0 D. The adaptive multifocal display and the focal surface display are optimized using k-means clustering, following Wu et al. [2016], and the methods in Sections 3.1 and 3.2 to position planes across a 5.0 D span, respectively. Focal surface displays show significantly fewer depth errors, with errors decreasing as more surfaces are used. (Source imagery courtesy Unity Asset Store publisher “VenCreations.”)

By casting viewing ray (θ_x, θ_y) from the viewer’s pupil to the SLM, and then by applying Equation 7, a target focal length f_p can be assigned for each SLM pixel (p_x, p_y) to create a virtual image at the desired focal surface depth. To realize this focal length, Equation 6 requires a phase function ϕ with the Hessian

$$H_\phi(p_x, p_y) = -\frac{2\pi}{\lambda f(p_x, p_y)} I. \quad (8)$$

There may be no ϕ that exactly satisfies this expression. In fact, such a ϕ only exists when f is constant and ϕ is quadratic (i.e., the phase represents a uniform lens).² Since Equation 8 cannot be exactly satisfied, we solve the following linear least squares problem to obtain a phase function ϕ that is as close as possible:

$$\min_{\phi} \sum_{(p_x, p_y) \in \Omega_p} \|\hat{H}[\phi](p_x, p_y) - \frac{-2\pi}{\lambda f(p_x, p_y)} I\|_F^2, \quad (9)$$

where $\|\cdot\|_F^2$ is the Frobenius norm and where $\hat{H}[\cdot]$ is the discrete Hessian operator, given by finite differences of ϕ . Note that the phase function ϕ plus any linear function $a + bx + cy$ has the same Hessian H , so we additionally constrain $\phi(0, 0) = 0$ and $\nabla\phi(0, 0) = 0$.

3.2.3 Representing Natural Scenes. Applying focal surface displays requires answering a key question: can natural scenes be well approximated by smooth focal surfaces d_i , and, if so, how many surfaces are required to accurately reproduce retinal blur? Following Wu et al. [2016], we first consider the Middlebury 2014 dataset from

²Proof: Abbreviate partial derivatives by $\partial_{xx}\phi$, $\partial_{xy}\phi$, and $\partial_{yy}\phi$. If the Hessian is everywhere a multiple of the identity, then $\partial_{xy}\phi = 0$ everywhere. In particular, $\partial_x\phi$ only depends on x and hence $\partial_{xx}\phi$ also only depends on x (i.e., $\partial_{xx}\phi(x, y) = \partial_{xx}\phi(x, y')$ for all y, y'). By the same logic, $\partial_{yy}\phi$ only depends on y . Since the Hessian is everywhere a multiple of the identity, we have that $\partial_{xx}\phi(p_x, p_y) = \partial_{yy}\phi(p_x, p_y)$ for all p . Finally, pick a fixed (\hat{p}_x, \hat{p}_y) , for any (p_x, p_y) on the SLM we have $\partial_{xx}\phi(p_x, p_y) = \partial_{xx}\phi(p_x, \hat{p}_y) = \partial_{yy}\phi(p_x, \hat{p}_y) = \partial_{yy}\phi(\hat{p}_x, \hat{p}_y) = \partial_{xx}\phi(\hat{p}_x, \hat{p}_y)$, so $\partial_{xx}\phi$ is constant (and similarly for $\partial_{yy}\phi$).

Scharstein et al. [2014], containing 33 depth maps from real-world environments. In Figure 6, we compare our depth approximation error with prior fixed and adaptive multifocal displays. A *single focal surface, as produced by our method, more closely follows scene geometry than prior fixed-focus multifocal displays (with four planes) and adaptive multifocal displays (with three planes)*. In practice, two focal surfaces appear to be an effective representation, allowing occlusions, transparencies, and reflections to be captured, so long as two dominant surfaces are visible in each viewing direction. In this manner, our focal surface display technique significantly reduces the number of required surfaces and contributes to the practicality of time-multiplexed multifocal displays.

Relying solely on the Middlebury dataset could provide a misleading conclusion, as the depths in that collection only span an average range of 1.0 diopters. As a result, we created our own synthetically rendered database (see Supplementary Appendix S.A for details). Our database spans a range of 4.0 diopters, on average. Resulting depth approximation errors are shown in Figure 7. Note that focal surface displays continue to outperform prior multifocal displays.

3.2.4 Focusing Errors Limit Visual Acuity. Reducing the number of planes, as with prior multifocal displays, is often achieved by increasing their separation. As noted by Narain et al. [2015], this comes at the cost of reducing the maximum-supported resolution (measured in cycles per degree). For example, Narain et al. estimate that contrast falls below 50% for 11 cpd spatial frequencies, or higher, with a plane separation of 0.6 diopters. For context, that would imply that a conventional fixed-focus multifocal display could not achieve resolutions, throughout the supported accommodation range, exceeding more than twice that of modern VR HMDs.

Based on the statistics in Figure 6 and 7, both multifocal and focal surface displays should achieve focusing errors less than 0.12 diopters, if operated over an appropriate accommodation range

with a sufficient number of components. Following Kotulak and Schor [1986], with this fidelity of focus, such systems should drive accommodation correctly. In this circumstance, focusing errors can be directly translated to a spatial frequency (resolution) limit via the modulation transfer function (MTF) of the human eye. Narain et al. apply a similar analysis to assess contrast limits. In this paper, we apply the 35% through-focus MTF of the human eye, as estimated by Villegas et al. [2002], to convert focusing errors to spatial frequency limits in Figures 6 and 7. Note that, with focal surface displays, a significantly higher resolution limit is predicted, opening a path to high-resolution HMDs, unlike prior multifocal displays.

3.2.5 Additional Metrics for Focal Surface Optimization. The paraxial approximation was applied to the phase optimization in Equation 9. However, a different criterion could be employed: find the phase ϕ minimizing the distance between the minimum-spot-size measured focus and the true depth d , summed over all angles Ω_θ . This metric accounts for higher-order aberrations (as it is inspired by similar analysis performed by optical design software), although it does not account for scene content (one may not care what the focus is in regions of uniform color). This metric requires evaluating the forward rendering operator from Section 3.3 and, as a result, would again produce a large nonlinear optimization problem—motivating our adoption of the paraxial model that, in practice, produces accurate focal surfaces. Efficiently leveraging the minimum-spot-size metric is a promising path for future work.

3.3 Optimized Blending with Focal Surfaces

Having determined k phase functions ϕ_i , corresponding to focal surfaces d_i , the last stage in our pipeline determines color images c_i , shown on the underlying display, to reproduce the target focal stack. This focal stack is represented by a set of l retinal images r_1, \dots, r_l . For this purpose, we generalize the optimized blending algorithm of Narain et al. [2015]. In this section, we first describe a ray-traced model of retinal blur. Afterward, this model is applied to evaluate the forward and adjoint operators required to solve the linear least squares problem representing optimized blending.

3.3.1 Modeling Retinal Blur with Ray Tracing. An optical ray is traced through our system under a geometric optics model. Following Figure 3, each ray originates at a point within the viewer’s pupil. The ray then passes through the front and back of the eyepiece, the SLM, and then impinges on the display. At the eyepiece surfaces, rays are refracted using the radius of curvature of the lens, its optical index, and the paraxial approximation. Equation 4 models light transport through the SLM. Each ray is assigned the color interpolated at its coordinate of intersection with the display. We denote locations on the display by (q_x, q_y) and the set of display pixel centers by Ω_q . Note that any rays that miss the bounds of the eyepiece, SLM, or display are culled (i.e., are assigned a black color).

To model retinal blur, we accumulate rays that span the viewer’s pupil, which we sample using a Poisson distribution. In this manner, we approximate the viewer’s eye as an ideal lens focused at a depth z which changes depending on the viewer’s accommodative state. For each chief ray (θ_x, θ_y) and depth z , we sum across a bundle of rays $R_{\theta_x, \theta_y, z}$ from the Poisson-sampled pupil. This produces an estimate of the retinal blur when focused at a depth z . We define these

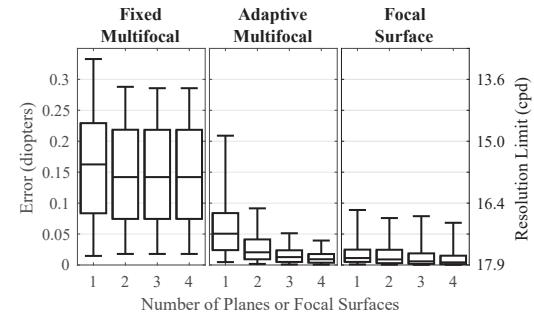


Fig. 6. Focal surface displays represent natural scene depths with few image components. Box plots compare the depth map errors $g_{\theta_x, \theta_y}(d)$ using the denoted methods with the Middlebury 2014 dataset [Scharstein et al. 2014]. The bottom and top of the whiskers indicate the 5th and 95th percentiles, respectively. The bottom, middle, and top of the boxes represent the 1st quartile, the median, and the 3rd quartile, respectively. Focal surface displays produce fewer depth errors, especially when fewer planes are used.

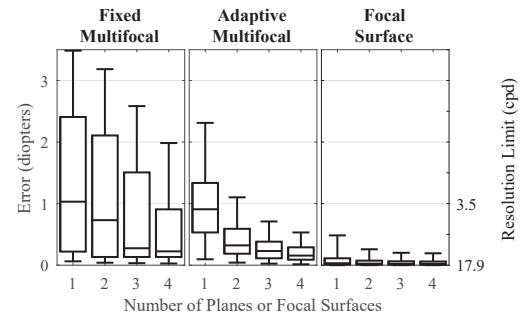


Fig. 7. We repeat the assessment of Figure 6, but with the database of rendered scenes described in Section 3.2. Note that the trends are repeated, but, due to the larger depth ranges in this database, additional virtual image surfaces are required with prior fixed and adaptive multifocal displays.

preceding steps as the *forward operator* $r = A_{z, \phi}(c)$, which accepts a phase function ϕ and color image c and predicts the perceived retinal image r when focused at a distance z .

3.3.2 Depicting Focal Stacks with Optimized Blending. For a fixed phase function ϕ and accommodation depth z , the forward operator $A_{z, \phi}(c)$ is linear in the color image c . The rendering operators $A_{z, \phi_i}(c_i)$ combine additively, so our combined forward operator, representing viewing of multiple-component focal surface displays, is $A_z(c_1, \dots, c_k) = \sum_i A_{z, \phi_i}(c_i)$. We can concatenate the forward renders for multiple accommodation depths z_1, \dots, z_l to estimate the reconstructed focal stack, with corresponding linear operator $A = [A_{z_1}; \dots; A_{z_l}]$. The forward operator, for a given set of color images c , gives the focal stack r that would be produced on the retina—minimizing $\|Ac - r\|^2$ gives the color image best approximating the desired focal stack. We have already given an efficient algorithm for computing $A_{z, \phi}$. Its transpose, mapping retinal image samples to display pixels, can be similarly evaluated with ray tracing operations with accumulation in the color image c rather than the retinal image r . In conclusion, these forward and adjoint operators are applied with an iterative least squares solver. (For implementation details, see Section 5.2.) Results of our full optimization pipeline are shown in Figure 8 and in the supplementary video.

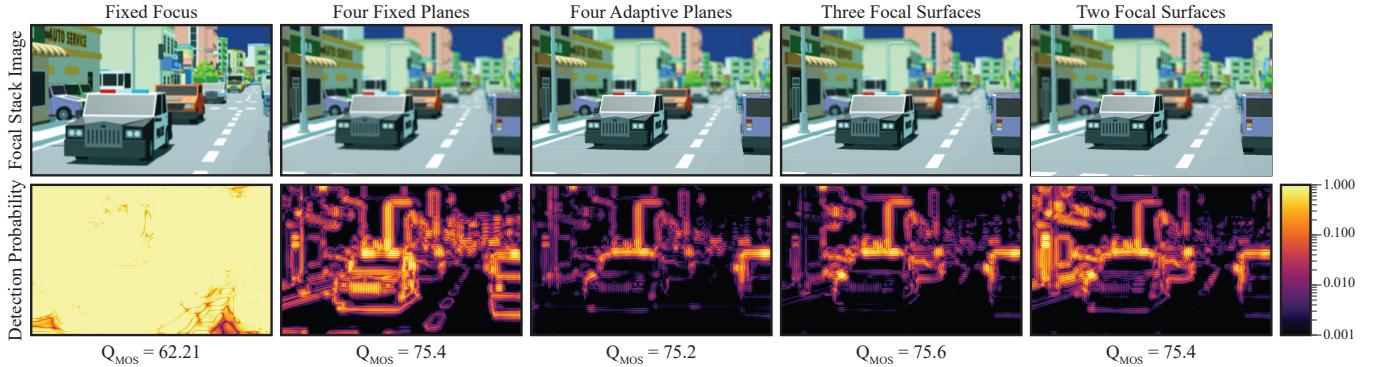


Fig. 8. Focal surface displays depict near-correct retinal blur with fewer virtual image surfaces than prior multifocal architectures. Following Figures 5–7, focal surface displays produce virtual images that more closely align with the scene geometry. As a result, sharply focused imagery can be obtained throughout the scene, reducing focusing errors occurring with prior fixed and adaptive multifocal displays. In this figure, we quantitatively assess the focal stack reproduction error following the method of Narain et al. [2015]: the lower row depicts the maximum per-pixel probability of detecting a difference between the target and reconstructed focal stacks, as quantified using the HDR-VDP-2 metric [Mantiuk et al. 2011b]. The corresponding quality predictor of the mean opinion score (MOS) is listed along the bottom. Note that focal surface displays achieve similar fidelity as prior adaptive multifocal displays, although with fewer virtual image surfaces. (Source imagery courtesy Unity Asset Store publisher “VenCreations.”)

4 DESIGNING FOCAL SURFACE DISPLAYS

In designing standard VR HMDs, there is a direct trade-off between field of view and resolution, which is largely determined by the placement of the display, the size and resolution of this display, and by the focal length of the eyepiece. For focal surface displays, there is a similar trade-off between the position of the SLM and the depth of focus (i.e., the supported accommodation range). In this section, we evaluate these trade-offs in terms of three metrics: field of view, depth of focus (DOF), and the degree of optical aberrations.

The field of view of a focal surface display is limited by the smaller of the display, the SLM, or the eyepiece (as appearing to the viewer). Wide eyepieces and displays are commonly available, so SLM dimensions currently limit the FOV. Ignoring variation with eye relief, the FOV is given by the angle subtended by the magnified SLM, or $2 \arctan \frac{r_p}{z_p}$, where r_p and z_p are the SLM radius and distance from the eyepiece, respectively. Thus, FOV is maximized by moving the SLM closer to the eyepiece.

Following Section 3.2, the SLM focal length is bounded such that $|f_p| \geq \frac{2r_p\delta_p}{\lambda}$. Substituting this range into Equation 7 gives a nonlinear expression mapping SLM focal length f_p and position z_p to the virtual image depth, and, as such, bounds the depth of focus. The resulting trade-off between DOF and the system design parameters is illustrated in Figure 9: depth of focus for a given lens position is the difference in contour values between the red constraints. From this analysis, we conclude that DOF increases as we move the SLM closer to the eyepiece.

Our final design metric is to minimize optical aberrations. As presented in Section 3.2, our method for generating phase functions optimizes phase curvature within small neighborhoods (since it is based on the discrete Hessian operator, which we evaluate using a 3×3 window). To estimate focus at angle (θ_x, θ_y) using our more accurate minimum-spot-size metric, we cast all rays in the bundle $R_{\theta_x, \theta_y, z}$ leaving the pupil. These rays intersect the SLM in a connected region P (i.e., the “circle of confusion”). The extent to which

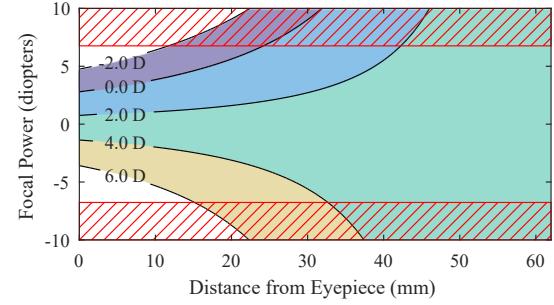
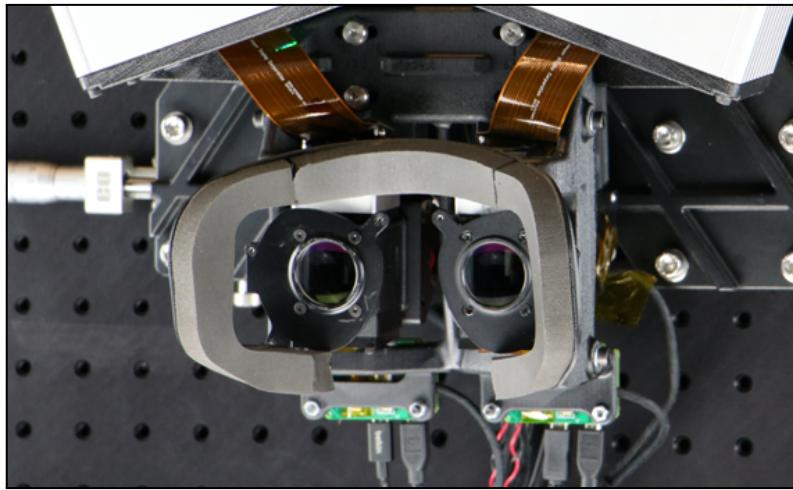


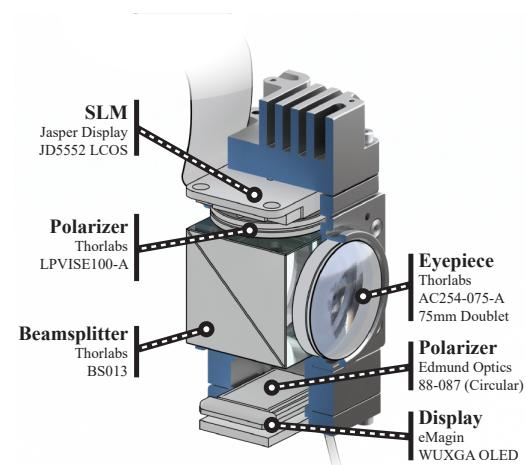
Fig. 9. The accommodation range of a focal surface display depends critically on the SLM placement. Here we denote, via the labeled plot contours, the virtual image distance z_v achieved with an SLM, when used to represent a lens of focal length f_p and positioned a distance z_p from the eyepiece. Red lines indicate focal lengths beyond the dynamic range of the SLM. Note that these numbers correspond with the prototype described in Section 5.1.

the rays intersect at a single point on the display depends on how close to quadratic the phase function is throughout all of P , not just the Hessian at a single point. It is easier to achieve this condition if the circle of confusion (i.e., P) is small, because the second-order Taylor series (i.e., the Hessian) is a better approximation in a small neighborhood. The size of P is linearly proportional to the distance between the SLM and the display. We conclude that, for aberration control, we desire the SLM to be as close to the display as possible.

In summary, minimizing aberrations encourages moving the display in the opposite direction as required to increase DOF and FOV. As with all optical systems, the designer must balance between these trade-offs. For our prototype, we positioned the SLM as close to the display as possible, while supporting accommodation from 0.0 to 4.0 diopters. In practice, the hardware constrains the SLM position due to the volume occupied by the beam splitter. Similarly, selecting from catalog lenses and SLMs limits the focal length, the SLM pixel pitch, and the SLM dimensions. Thus, only certain points in this



(a) Construction of the Prototype



(b) Arrangement of the Optical Components

Fig. 10. Our binocular focal surface display prototype incorporates commodity optical and mechanical components, as well as 3D-printed support brackets. (a) The prototype is mounted to an optical breadboard to support the comparatively large LCOS driver electronics. A chin rest is used to position the viewer within the eye box. (b) A cutaway of the prototype exposes the arrangement of the optical components.

design tradespace were readily accessible. However, the DOF of our prototype remains comparable to prior accommodation-supporting display prototypes, as summarized in Table 1.

5 IMPLEMENTATION AND RESULTS

A prototype is necessary to demonstrate the fundamental concepts presented in the preceding sections, as well as to identify practical limitations encountered with current-generation phase modulation hardware. In this section, we describe our hardware and software choices, and we evaluate the resulting experimental performance.

5.1 Hardware

Our prototype largely uses off-the-shelf optical and mechanical components, augmented with a handful of 3D-printed parts. The optical path begins, as shown in Figure 10b, with an eMagin WUXGA 1920×1200 60 Hz color OLED display, addressed via an MRA Digital HDMI driver board. The OLED is covered with an Edmund Optics 88-087 left-handed circular polarizer to suppress stray light reflections. Illumination from the display next encounters a Thorlabs 50:50 non-polarizing beamsplitter. The light reflected by the beamsplitter immediately impinges on a “beam dump” (i.e., a felt-covered, light-absorbing surface). Note that an eye tracking camera could be fitted to this side of the beamsplitter, as it allows imaging of viewer’s pupil in a manner that bypasses the phase modulator. The transmitted path through the beamsplitter contains the phase modulator, a Jasper Display JD5552 1920×1080 60 Hz reflective LCOS SLM, addressed via the driver board supplied in the Jasper Display JD9554 Educational Kit. To operate this SLM in a phase-modulation mode, a Thorlabs LPVISE100-A polarizer is affixed in front of the SLM. The phase-modulated illumination propagates back through the beamsplitter, with the reflected path passing through a Thorlabs 75 mm lens (the eyepiece) and on to the viewer. The transmitted path returns towards the OLED, with the previously introduced

circular polarizer acting as an optical isolator and suppressing the return reflection. This entire assembly was duplicated, in a mirrored fashion, to enable binocular viewing, with each side mounted to a translation stage to adjust the interaxial distance (IAD) and with an optical breadboard supporting the LCOS drivers. A photograph of the assembled prototype is shown in Figure 10a.

Given the design considerations and the practical SLM limitations presented in Sections 4 and 6.1, respectively, the prototype has a measured DOF spanning 0.75–4.0 diopters (slightly less than the design specifications), assuming an eye relief of 10 mm. The field of view, limited by the size of the SLM, is 18° diagonally.

The HDMI inputs for the OLEDs and SLMs are connected to a host computer containing a pair of NVIDIA GTX Titan X (Maxwell) graphics cards with a 3.4 GHz Intel Core i7-3770 processor and 16 GB RAM. This computer was also used to run the focal surface decomposition, blending, and other rendering algorithms.

5.2 Software

5.2.1 Rendering. The forward rendering model from Section 3.3 was implemented using NVIDIA OptiX. Our scene database was rendered using Unity 5.5, assuming an ideal circular pupil. Focal stacks were evaluated offline with an accumulation buffer.

5.2.2 Optimization. Focal surface decomposition is optimized using a cost function following Section 3.2, as implemented in C/C++ with Ceres Solver [Agarwal and Others 2012]. The LBFGS algorithm [Nocedal 1980] was selected for iterative gradient descent. Depth map decompositions were evaluated on 192×108 downsampled images, with an average run time of 2.4 seconds (for three image components). Phase function optimization at the native SLM resolution took about 46 seconds per focal surface. Our optimized blending algorithm, again with three planes, took an average of 42 minutes (with 30 iterations), comparable to the run time reported by Narain et al. [2015]. In contrast, linear blending required 17 seconds.



Fig. 11. Our prototype focal surface display achieves high resolution with near-correct retinal blur. Photographs of the prototype are shown in the first three columns, as taken by focusing the camera at the indicated distances. The last two columns depict the corresponding optimization outputs, including the phase functions and the color images. Note that optimized blending is applied with three time-multiplexed focal surfaces. The phase functions are wrapped assuming a wavelength of 532 nm. Note that the results for the lower scene employ linear blending, following Akeley et al. [2004]. See the supplementary video for the full focal stack results. (Source imagery courtesy Unity Asset Store publisher “VenCreations” and Thomas Guillou.)

5.2.3 Calibration. Operation of a focal surface display requires understanding the alignment of optical components. Errors in assembly manifest as displacements in the focal surfaces, requiring calibration. For this purpose, we first employ a calibrated varifocal camera, using a Varioptic Caspian C-C-39N0-250-R33 tunable lens. With this camera, we measure the location of the rendered focal surfaces and, thereby, refine our estimates of the system parameters. Second, we position the camera so that it is located at the rendered center of projection. Third, we measure and correct transverse chromatic aberration using controlled illumination patterns.

5.3 Experimental Results

Experimental results are reported in Figure 11. In these examples, we apply time-multiplexed presentation with three focal surfaces, similar to prior simulations. We emphasize that color fields were displayed simultaneously in all cases (see Section 6.1 for details).

Our prototype addresses a key question: does diffraction degrade image quality to an extent prohibiting practical applications? To this end, we measured the modulation transfer function (MTF). Following Figure 13, MTF was assessed by displaying a series of sinusoids at a given focal distance, focusing a varifocal camera to that distance, and measuring the average contrast over the FOV. As predicted in prior sections, focal surface displays support high-resolution imagery. Specifically, our prototype achieves a resolution better than 5 cycles per degree throughout the accommodation range. As a result, our prototype is on par with modern VR HMDs, and considerably better in the center of its depth of focus.

Higher resolutions (exceeding 20 cpd) are possible when the SLM is used with longer focal lengths, as occurring for system focus near 3.0 diopters. In our prototype, the SLM creates shorter focal lengths as the focus approaches 1.0 and 4.0 diopters, resulting in reduced contrast (see Figure 13). The SLM may also exhibit chromatic aberration, further reducing contrast. Critically, diffraction-related issues often prohibit layered displays from achieving high resolutions (see Table 1). Focal surface displays are not similarly hindered. However, practical SLMs support finite, discrete phase modulation, typically limited to a range of 2π . Large phase gradients, as occurring with short focal lengths, produce quantization artifacts and frequent phase resets, resulting in unwanted energy in higher-order diffraction modes and stray light [Laude 1998]. These effects reduce contrast, as shown in Figure 11. Thus, we observe a key direction for future work: extending the phase modulation range beyond 2π to allow higher resolutions and sharper variations in focal surfaces.

6 DISCUSSION

6.1 Addressing Limitations

Focal surface displays have been shown to achieve high-fidelity depictions of natural scenes. We now turn our attention to discussing the current and future practicality of this concept. As with any computational display, one must jointly consider issues regarding optical hardware, display technology, and optimization algorithms.

6.1.1 Supporting Multiple Focal Surfaces. The primary motivation for pursuing focal surface displays over simpler multifocal designs is to reduce the number of multiplexed images. Llull et

al. [2015] apply a 400 Hz tunable lens to achieve a 60 Hz multifocal display. We use a 60 Hz SLM, but this is not a fundamental limitation: Jasper Display JD4552 and HOLOEYE LETO support 720 Hz and 180 Hz, respectively. In terms of image quality, single focal surfaces arguably perform competitively. However, we strive to depict near-correct retinal blur, particularly at occlusions. As such, designs with two focal surfaces appear a viable, and practically realizable, first step toward accommodation-supporting HMDs.

6.1.2 Resolving Phase Modulation Issues. Our use of phase SLMs is related to earlier work on dynamic freeform lensing. As previously assessed by Damberg et al. [2016], using LCOS panels in imaging systems presents two primary concerns: stray light and chromatic aberration. We discuss each in turn.

As discussed in Section 5.3, stray light may result from inefficiencies of the phase SLM. However, LCOS phase SLMs are routinely applied with adaptive optics, including for retinal imaging and aberration correction. As such, LCOS panels have already benefited from extended research into suppressing stray light. A full assessment of these effects is beyond the scope of this work. However, our MTF measurements in Figure 13, as well as the experimental results, support that high-resolution imagery can be created.

Following Equation 5, the effective SLM focal length is wavelength dependent. As a result, the LCOS panel may introduce transverse and axial chromatic aberrations. While the former can be digitally corrected by warping displayed images, the latter cannot and manifests as focusing artifacts. The conventional solution is field sequential color presentation. However, our goal is to reduce time multiplexing and, as a result, we aim for field simultaneous color presentation. We emphasize that Laude [1998], Márquez et al. [2006], and Fernandez et al. [2010] each report the successful operation of phase-only SLMs as focusing elements using polychromatic and broadband illumination. As summarized in Figure 14, we measure an average axial chromatic aberration (ACA) of less than 0.25 diopters over the supported accommodation range. Simulations depicted in Figure 12 indicate modest benefits, in terms of minimizing color fringing, by employing field sequential color (i.e., by using separately optimized phase functions for each color channel). Note that ACA is predicted with the geometric optics simulations, due to the dispersion introduced by Equation 4.

Our simulations apply the geometric optics model from Section 3, which does not predict all experimental artifacts. First, we do not model wave optics effects, including stray light due to phase quantization and phase resets. As a result, the experimentally measured contrast loss, as reported in Figure 13, is not reproduced in the simulations. Second, the ACA of the physical SLM differs slightly from our model. Third, the calibration procedure in Section 5.2.3 does not account for vignetting and all sources of misalignment, introducing multifocal blending artifacts near the periphery.

While experimental results do not yet attain the quality of our geometric optics model, field simultaneous color and mitigation of stray light appear realizable with practical SLMs, particularly by applying phase modulation exceeding the 2π range of our prototype, as described by Fernandez et al. [2010]. We emphasize that all our experimental results, except for those in Figure 12, were captured while displaying all color fields simultaneously.

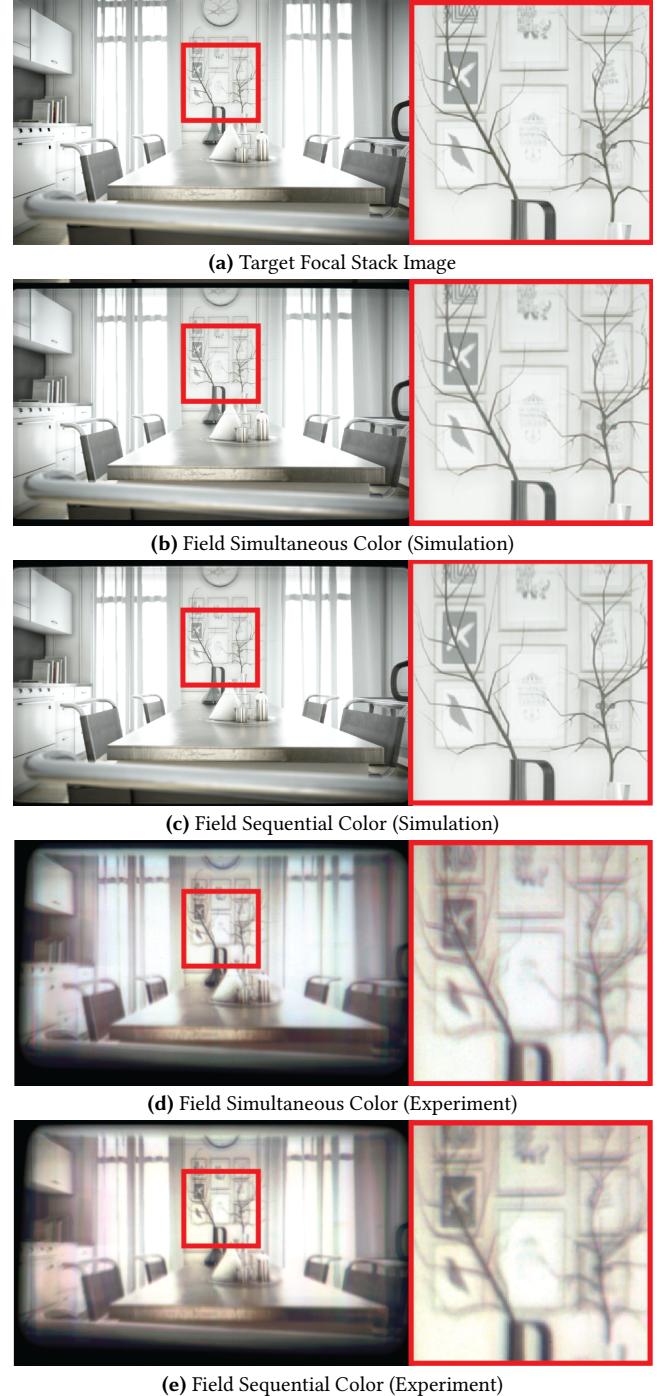


Fig. 12. To minimize time multiplexing, focal surface displays should operate in a field simultaneous color mode. Following Section 6.1.2, artifacts due to axial chromatic aberration (ACA) may appear in this mode. (a) A target focal stack image. (b,c) Simulations comparing field simultaneous and field sequential modes, using the geometric optics model from Section 3. (d,e) Corresponding experimental results. Note that the contrast of experimental results differs from simulations due to stray light and misalignments that cannot be predicted without more accurate wave optics modeling and calibration, respectively. (Source imagery courtesy Ruggiero Corridori.)

6.1.3 Optimizing Algorithm Performance. The algorithms that drive our prototype are not yet suitable for interactive content. A promising direction for future work is to explore efficient depth decomposition and optimized blending frameworks. In terms of the latter, the optimized blending algorithm of Narain et al. [2015] poses a more significant hurdle, with reported minute-long run times. However, linear blending could be adopted to approach real-time refresh rates, albeit with diminished retinal blur fidelity.

6.1.4 Enabling Practical Applications. Here we turn our attention first to practical VR applications, and then to AR. Our prototype is not yet wearable, due to the large LCOS drivers. This is not a fundamental limitation, as attested by commercial pico projectors. However, VR applications do confront a current roadblock: LCOS panels are smaller than modern VR optics. As such, the field of view remains limited. Increasing the FOV requires three changes: using a shorter focal length eyepiece, eliminating the beamsplitter and replacing the reflective LCOS with a transmissive one, and reducing the overall optical stack height. Even if these measures were taken, a larger SLM would be required. Practical VR applications will require custom SLMs. However, we emphasize most accommodation-supporting HMDs are similarly technologically limited to narrow FOVs, as surveyed in Table 1.

Focal surface displays currently appear to be a forward-looking architecture requiring further maturation of SLM technology. While our prototype modifies a conventional VR architecture, largely due to the accessibility of catalog eyepieces, we believe focal surface displays can be equally applied to AR devices. Specifically, those that substitute a projector and a combiner for the display and eyepiece. This configuration is a natural direction for focal surface displays: larger SLMs (our primary limitation) would not be required, as existing models would easily fit into a miniature projector. As such, focal surface displays continue the legacy of retinal scanning displays, providing a viable path to address refresh rate and multivalued depth limitations encountered by McQuaide et al. [2003].

6.2 Future Work

Immediate extensions to this work include upgrading to wave optics modeling, generalizing to non-smooth focal surfaces, and exploring alternative depth map decompositions (e.g., those that penalize all focal surfaces, rather than just the closest.) However, the future work for focal surface displays largely overlaps with that required for all multifocal displays. As presented in Section 3, focal surface displays are a form of fixed-viewpoint volumetric display: rendering, optimization, and viewing are all assumed to occur relative to the viewer’s entrance pupil. It is worth noting that Maxwellian view, retinal scanning, and other extended depth of focus concepts also share this assumption. A promising direction is to determine whether, through hardware or algorithms, eye movement can be supported. With eye tracking, focal surface displays may be driven in a gaze-contingent manner, similar to varifocal concepts. There is also an opportunity to leverage concepts from near-eye light field displays, rendering imagery to support limited eye movement. In this manner, we believe the challenges and research directions for all accommodation-supporting displays are closely tied.

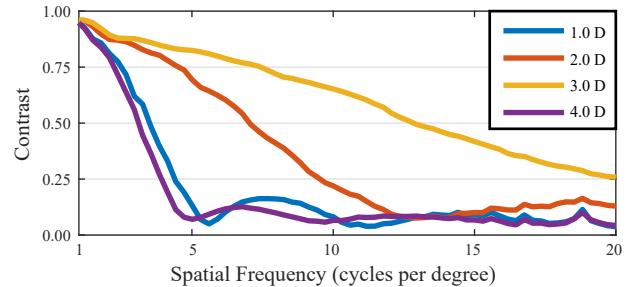


Fig. 13. The measured modulation transfer function (MTF) of our prototype confirms high resolution is achieved. Following Section 5.3, the MTF was measured as the system varies focus from 0.0 to 4.0 diopters. Contrast loss is expected as the SLM synthesizes shorter focal lengths, due to the increased stray light from phase quantization and phase resets (see Section 5.3).

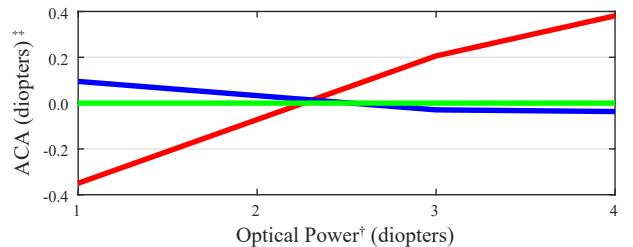


Fig. 14. The measured axial chromatic aberration (ACA) of our prototype is less than that of the typical human eye [Fernandez et al. 2013], confirming that focal cues are correctly rendered with field simultaneous color presentation, in spite of polychromatic illumination. [†]The SLM optical power was optimized, following Equation 5, for $\lambda = 532$ nm. [‡]ACA is reported as the apparent optical distance in diopters, measured relative to the green channel. Focal distances are measured using a varifocal camera and a depth-from-focus metric (i.e., maximizing contrast for a high-frequency pattern).

7 CONCLUSION

Focal surface displays continue down the path set by varifocal and multifocal concepts, further customizing virtual images to scene content. We have demonstrated that emerging phase-modulation SLMs are well-prepared to realize this concept, having benefited from decades of research into closely-related adaptive imaging applications. We have demonstrated high-resolution focal stack reproductions with a proof-of-concept prototype, as well as presented a complete optimization framework addressing the joint focal surface and color image decomposition problems. By unifying concepts in goal-based caustics, retinal scanning displays, and other accommodation-supporting HMDs, we hope to inspire other researchers to leverage emerging display technologies that may address vergence-accommodation conflict in HMDs.

ACKNOWLEDGMENTS

We thank the reviewers for their helpful feedback. We also thank Nicholas Trail for encouraging the authors to pursue the topic of local focal surface adaptation and for assisting with the optical hardware design. We recognize Ryan Ebert for the design and assembly of the prototype, refining an earlier testbed developed by Nathan Matsuda. We also appreciate the feedback provided by Heeyoon Lee, Wanli Chi, and Oliver Cossairt.

REFERENCES

- Sameer Agarwal and Others. 2012. Ceres Solver. <http://ceres-solver.org>. (2012).
- Kurt Akeley, Simon J. Watt, Ahna R. Girshick, and Martin S. Banks. 2004. A Stereo Display Prototype with Multiple Focal Distances. *ACM Trans. Graph.* 23, 3 (2004).
- Barry Blundell and Adam Schwartz. 1999. *Volumetric Three-Dimensional Display Systems*. Wiley-IEEE Press.
- V. Michael Bove. 2012. Display Holography's Digital Second Act. *Proc. IEEE* 100, 4 (2012), 918–928.
- Stephen A. Burns and Robert H. Webb. 2010. Optical Generation of the Visual Stimulus. In *Handbook of Optics, Third Edition Volume III*, Michael Bass (Ed.). McGraw-Hill.
- Ozan Cakmakci and Jannick Rolland. 2006. Head-Worn Displays: A Review. *Journal of Display Technology* 2, 3 (2006), 199–216.
- John D. Cook. 2010. How to compute the soft maximum. <http://www.johndcook.com/blog/2010/01/20/>. (2010).
- Gerwin Damberg, James Gregson, and Wolfgang Heidrich. 2016. High Brightness HDR Projection Using Dynamic Freeform Lensing. *ACM Trans. Graph.* 35, 3 (2016).
- Andrew T. Duchowski and Others. 2014. Reducing Visual Discomfort of 3D Stereoscopic Displays with Gaze-contingent Depth-of-field. In *ACM Symposium on Applied Perception*. 39–46.
- David Dunn, Cary Tippets, Kent Torell, Petr Kellnhofer, Kaan Akşit, Piotr Didyk, Karol Myszkowski, David Luebke, and Henry Fuchs. 2017. Wide Field of View Varifocal Near-Eye Display using See-through Deformable Membrane Mirrors. *IEEE TVCG* 23, 4 (2017), 1322–1331.
- Enrique J. Fernández, Pedro M. Prieto, and Pablo Artal. 2010. Adaptive optics binocular visual simulator to study stereopsis in the presence of aberrations. *J. Optical Society of America A* 27, 11 (2010). DOI: <https://doi.org/10.1364/JOSAA.27.000A48>
- Enrique J. Fernandez, Pedro M. Prieto, Emmanuel Chirre, and Pablo Artal. 2013. Performance of a 6-Pi liquid crystal on silicon (LCoS) spatial light modulator under white light illumination for visual applications. *Imaging and Applied Optics*.
- Daniel Glasner, Todd Zickler, and Anat Levin. 2014. A Reflectance Display. *ACM Trans. Graph.* (2014).
- Sébastien Hillaire, Anatole Lécuyer, Rémi Cozot, and Géry Casiez. 2008. Using an Eye-Tracking System to Improve Camera Motions and Depth-of-Field Blur Effects in Virtual Environments. In *IEEE Virtual Reality*. 47–50.
- David M. Hoffman, R. Girshick, Kurt Akeley, and Martin S. Banks. 2008. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision* 8, 3 (2008), 33.
- Xinda Hu and Hong Hua. 2014. High-resolution optical see-through multi-focal-plane head-mounted display using freeform optics. *Optics Express* 22, 11 (2014).
- Hong Hua and Bahram Javidi. 2014. A 3D integral imaging optical see-through head-mounted display. *Optics Express* 22, 11 (2014).
- Fu-Chung Huang, Kevin Chen, and Gordon Wetzstein. 2015. The Light Field Stereoscope: Immersive Computer Graphics via Factored Near-eye Light Field Displays with Focus Cues. *ACM Trans. Graph.* 34, 4, Article 60 (2015), 12 pages.
- Fu-Chung Huang, Douglas Lanman, Brian Barsky, and Ramesh Raskar. 2012. Correcting for Optical Aberrations using Multilayer Displays. *ACM Trans. Graph.* 31, 6 (2012).
- Robert J. Jacobs, Ian L. Bailey, and Mark A. Bullimore. 1992. Artificial pupils and Maxwellian view. *Applied Optics* 31, 19 (1992).
- Paul V. Johnson, Jared A. Parnell, Joohwan Kim, Christopher Saunter, Martin S. Banks, and Gordon D. Love. 2016. Assessing Visual Discomfort using Dynamic Lens and Monovision Displays. *Imaging and Applied Optics* 2016.
- Robert Konrad, Emily A. Cooper, and Gordon Wetzstein. 2016. Novel Optical Configurations for Virtual Reality: Evaluating User Preference and Performance with Focus-tunable and Monovision Near-eye Displays. *ACM CHI* (2016), 1211–1220.
- John C. Kotulak and Clifton M. Schor. 1986. The accommodative response to subthreshold blur and to perceptual fading during the Troxler phenomenon. *Perception* 15, 1 (1986), 7–15.
- Gregory Kramida. 2016. Resolving the Vergence-Accommodation Conflict in Head-Mounted Displays. *IEEE TVCG* 22, 7 (2016), 1912–1931.
- Bernard Kress and Thad Starner. 2013. A review of head-mounted displays (HMD) technologies and applications for consumer electronics. *SPIE* 8720 (2013).
- Douglas Lanman and David Luebke. 2013. Near-Eye Light Field Displays. *ACM Trans. Graph.* 32, 6, Article 220 (2013), 10 pages.
- Vincent Laude. 1998. Twisted-nematic liquid-crystal pixelated active lens. *Optics Communications* 153 (1998), 134–152.
- Anat Levin, Haggai Maron, and Michal Yarom. 2016. Passive light and viewpoint sensitive display of 3D content. In *IEEE Int. Conf. on Computational Photography*.
- Sheng Liu, Hong Hua, and Dewen Cheng. 2010. A Novel Prototype for an Optical See-Through Head-Mounted Display with Addressable Focus Cues. *IEEE TVCG* 16, 3 (2010), 381–393.
- Patrick Llull, Noah Bedard, Wanmin Wu, Ivana Tošić, Kathrin Berkner, and Nikhil Balram. 2015. Design and optimization of a near-eye multifocal display system for augmented reality. *Imaging and Applied Optics*.
- Gordon D. Love, David M. Hoffman, Philip J.W. Hands, James Gao, Andrew K. Kirby, and Martin S. Banks. 2009. High-speed switchable lens enables the development of a volumetric stereoscopic display. *Optics Express* 17, 18 (2009).
- Kevin J. MacKenzie, Ruth A. Dickson, and Simon J. Watt. 2012. Vergence and accommodation to multiple-image-plane stereoscopic displays: “real world” responses with practical image-plane separations? *Journal of Electronic Imaging* 21, 1 (2012).
- Guido Maiello, Manuela Chessa, Fabio Solari, and Peter J. Bex. 2015. The (In)Effectiveness of Simulated Blur for Depth Perception in Naturalistic Images. *PLOS ONE* 10, 10 (2015).
- Andrew Maimona and Henry Fuchs. 2013. Computational augmented reality eyeglasses. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 29–38.
- Rados Mantiuk, Bartosz Bazylik, and Anna Tomaszecka. 2011a. Gaze-Dependent Depth-of-field Effect Rendering in Virtual Environments. In *Serious Games Development and Applications*. Springer-Verlag, 1–12.
- Rafal Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich. 2011b. HDR-VDP-2: A Calibrated Visual Metric for Visibility and Quality Predictions in All Luminance Conditions. *ACM Trans. Graph.* 30, 4, Article 40 (2011).
- Andrés Márquez, Claudio Lemmi, Juan Campos, and María J. Yzuel. 2006. Achromatic diffractive lens written onto a liquid crystal display. *Optics Letters* 31, 3 (2006).
- Lynn Marran and Clifton Schor. 1997. Multiaccommodative Stimuli in VR Systems: Problems & Solutions. *Human Factors* 39, 3 (1997), 382–388.
- Sarah C. McQuaide, Eric J. Seibel, John P. Kelly, Brian T. Schwengeler, and Thomas A. Furness III. 2003. A retinal scanning display system that produces multiple focal planes with a deformable membrane mirror. *Displays* 24, 2 (2003).
- Eunkyoung Moon, Myeongjae Kim, Jinyoung Roh, Hwi Kim, and Joonku Hahn. 2014. Holographic head-mounted display with RGB light emitting diode light source. *Optics Express* 22, 6 (2014), 6526–6534.
- Rahul Narain, Rachel A. Albert, Abdullah Bulbul, Gregory J. Ward, Martin S. Banks, and James F. O’Brien. 2015. Optimal Presentation of Imagery with Focus Cues on Multi-plane Displays. *ACM Trans. Graph.* 34, 4, Article 59 (2015).
- Mark A. A. Neil, Edward G. S. Paige, and Leon O. D. Sucharov. 1997. Spatial-light-modulator-based three-dimensional multiplanar display. *SPIE* 3012 (1997), 337–341.
- Jorge Nocedal. 1980. Updating quasi-Newton matrices with limited storage. *Math. Comp.* 35, 151 (1980), 773–782.
- Nitish Padmanaban, Robert Konrad, Tal Stramer, Emily A. Cooper, and Gordon Wetzstein. 2017. Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *Proc. of the National Academy of Sciences* 114, 9 (2017).
- Marios Papas, Wojciech Jarosz, Wenzel Jakob, Szymon Rusinkiewicz, Wojciech Matusik, and Tim Weyrich. 2011. Goal-based Caustics. *Computer Graphics Forum* 30, 2 (2011).
- Eli Peli. 1999. Optometric and perceptual issues with head-mounted displays. In *Visual Instrumentation: Optical Design and Principles*, P. Mouroulis (Ed.). McGraw-Hill.
- Sowmya Ravikumar, Kurt Akeley, and Martin S. Banks. 2011. Creating effective focus cues in multi-plane 3D displays. *Optics Express* 19, 21 (2011).
- Jannick P. Rolland, Myron W. Krueger, and Alexei Goon. 2000. Multifocal Planes Head-Mounted Displays. *Applied Optics* 39 (2000), 3209–3215.
- Daniel Scharstein, Heiko Hirschmüller, York Kitajima, Greg Krathwohl, Nera Nesic, Xi Wang, and Porter Westling. 2014. High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth. In *GCPR (LNCS)*, Vol. 8753. Springer.
- Takashi Shibata, Joohwan Kim, David M. Hoffman, and Martin S. Banks. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision* 11, 8 (2011), 11.
- Shinichi Shiwa, Katsuyuki Omura, and Fumio Kishino. 1996. Proposal for a 3-D display with accommodative compensation: 3DDAC. *J. Soc. Information Display* 4, 4 (1996).
- William W. Sprague, Emily A. Cooper, Ivana Tošić, and Martin S. Banks. 2015. Stereopsis is adaptive for the natural environment. *Science Advances* 1, 4 (2015).
- Toshiaki Sugihara and Tsutomu Miyasato. 1998. System development of fatigue-less HMD system 3DDAC (3D Display with Accommodative Compensation): System implementation of Mk.4 in light-weight HMD. *Image Engineering* 97, 467 (1998).
- Erik S. Viirre, Homer Pryor, Satoru Nagata, and Thomas A. Furness III. 1998. The virtual retinal display. In *Proceedings of Medicine Meets Virtual Reality*. 252–257.
- Eloy A. Villegas, Concepción González, Bernard Bourdoncle, Thierry Bonnin, and Pablo Artal. 2002. Correlation between optical and psychophysical parameters as a function of defocus. *Optometry & Vision Science* 79, 1 (2002), 60–67.
- David G. Voelz. 2011. *Computational Fourier Optics: A MATLAB Tutorial*. SPIE Press.
- Marc von Waldkirch. 2005. *Retinal projection displays for accommodation-insensitive viewing*. Ph.D. Dissertation, ETH Zurich.
- Marc von Waldkirch, Paul Lukowicz, and Gerhard Tröster. 2004. Spectacle-Based Design of Wearable See-Through Display for Accommodation-Free Viewing. In *Pervasive Computing*.
- Wanmin Wu, Patrick Llull, Ivana Tošić, Noah Bedard, Kathrin Berkner, and Nikhil Balram. 2016. Content-adaptive focus configuration for near-eye multi-focal displays. In *IEEE Multimedia and Expo*.
- Yonghao Yue, Kei Iwasaki, Bing-Yu Chen, Yoshinori Dobashi, and Tomoyuki Nishita. 2014. Poisson-based Continuous Surface Generation for Goal-Based Caustics. *ACM Trans. Graph.* 33, 3 (2014), 7.
- Zeev Zalevsky. 2010. Extended depth of focus imaging: A review. *Journal of Photonics for Energy* (2010).
- Marina Zannoli, Gordon D. Love, Rahul Narain, and Martin S. Banks. 2016. Blur and the perception of depth at occlusions. *Journal of Vision* 16, 6 (2016), 17.