Towards Multifocal Displays with Dense Focal Stacks

JEN-HAO RICK CHANG, Carnegie Mellon University, USA
B. V. K. VIJAYA KUMAR, Carnegie Mellon University, USA
ASWIN C. SANKARANARAYANAN, Carnegie Mellon University, USA

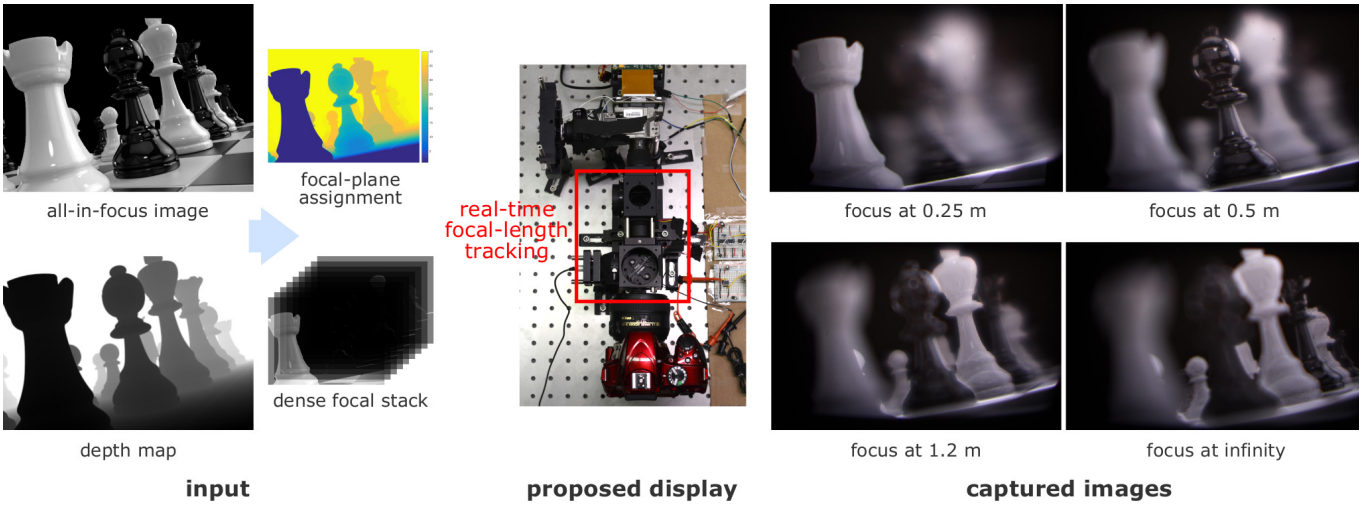


Fig. 1. Producing strong focusing cues for the human eye requires rendering scenes with dense focal stacks. This would require a virtual reality display that can produce thousands of focal planes per second. We achieve such a capability by exciting a focus-tunable lens with a high-frequency input and subsequently tracking the focal length at microsecond time resolution using an optical module. Using a lab prototype, we demonstrate that the high-speed tracking of the focal length, coupled with a high-speed display, can render a very dense set of focal stacks. Our system is capable of generating 1600 focal planes per second, which we use to render 40 focal planes per frame at 40 frames per second. Shown are images captured with a Nikon D3400 camera with a 50mm $f/2.8$ lens focused at different depths away from the tunable lens.

We present a virtual reality display that is capable of generating a dense collection of depth/focal planes. This is achieved by driving a focus-tunable lens to sweep a range of focal lengths at a high frequency and, subsequently, tracking the focal length precisely at microsecond time resolutions using an optical module. Precise tracking of the focal length, coupled with a high-speed display, enables our lab prototype to generate 1600 focal planes per second. This enables a novel first-of-its-kind virtual reality multifocal display that is capable of resolving the vergence-accommodation conflict endemic to today's displays.

CCS Concepts: • **Computing methodologies** → **Virtual reality**;

Additional Key Words and Phrases: focus-tunable lenses, multifocal displays, focus stacks

ACM Reference Format:
Jen-Hao Rick Chang, B. V. K. Vijaya Kumar, and Aswin C. Sankaranarayanan. 2018. Towards Multifocal Displays with Dense Focal Stacks. *ACM Trans. Graph.* 37, 6, Article 198 (November 2018), 13 pages. https://doi.org/10.1145/3272127.3275015

Authors' addresses: Jen-Hao Rick Chang, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, 15213, USA, rickchang@cmu.edu; B. V. K. Vijaya Kumar, Carnegie Mellon University, Pittsburgh, USA, kumar@ece.cmu.edu; Aswin C. Sankaranarayanan, Carnegie Mellon University, Pittsburgh, USA, saswin@andrew.cmu.edu.

1 INTRODUCTION

The human eye automatically changes the focus of its lens to provide sharp, in-focus images of objects at different depths. While convenient in the real world, for virtual or augmented reality (VR/AR) applications, this focusing capability of the eye often causes a problem that is called the vergence-accommodation conflict (VAC) [Hua 2017; Kramida 2016]. Vergence refers to the simultaneous movement of the two eyes so that a scene point comes into the center of the field of view, and accommodation refers to the changing of the focus of the ocular lenses to bring the object into focus. In the real world, these two cues act in synchrony. However, most commercial VR/AR displays render scenes by only satisfying the vergence cue, i.e., they manipulate the disparity of the images shown to each eye. But given that the display is at a fixed distance from the eyes, the corresponding accommodation cues are invariably incorrect, leading to a conflict between vergence and accommodation that can cause discomfort, fatigue, and distorted 3D perception, especially after long durations of usage [Hoffman et al. 2008; Vishwanath and Blaser 2010; Watt et al. 2005; Zannoli et al. 2016]. While many approaches have been proposed to mitigate the VAC, it remains one of the important challenges for VR and AR displays.

In this paper, we provide the design for a VR display that is capable of addressing the VAC by displaying content on a dense collection of depth or focal planes. The proposed display falls under the category

of multifocal displays, i.e., displays that generate content at different focal planes using a focus-tunable lens [Johnson et al. 2016; Konrad et al. 2016; Liu et al. 2008; Liu and Hua 2009; Llull et al. 2015; Love et al. 2009]. This change in focal length can be implemented in one of many ways; for example, by changing the curvature of a liquid lens [Optotune 2017; Varioptic 2017], the state of a liquid-crystal lens [Jamali et al. 2018a,b], the polarization of a waveplate lens [Tabiryan et al. 2015], or the relative orientation between two carefully designed phase plates [Bernet and Ritsch-Marte 2008]. The key distinguishing factor is that the proposed device displays a stack of focal planes that are an order of magnitude greater in number as compared to prior work, without any loss in the frame rate of the display. Specifically, our prototype system is capable of displaying 1600 focal planes per second, which can be used to display scenes with 40 focal planes per frame at 40 frames per second. As a consequence, we are able to render virtual worlds at a realism that is hard to achieve with current multifocal display designs.

To understand how our system can display thousands of focal planes per second, it is worth pointing out that the key factor that limits the depth resolution of a multifocal display is the operational speed of its focus-tunable lens. Focus-tunable liquid lenses change their focal length based on an input driving voltage; they typically require around 5 *m*s to settle onto a particular focal length. Hence, in order to wait for the lens to settle so that the displayed image is rendered at the desired depth, we can output at most 200 focal planes per second. For a display operating with 30-60 frames per second (fps), this would imply anywhere between three and six focal planes per frame, which is woefully inadequate.

The proposed display relies on the observation that, while focus-tunable lenses have long settling times, their frequency response is rather broad and has a cut-off upwards of 1000 Hz [Optotune 2017]. This suggests that we can drive the lens with excitations that are radically different from a simple step edge (i.e., a change in voltage). For example, we could make the lens sweep through its entire gamut of focal lengths at a high frequency simply by exciting it with a sinusoid or a triangular voltage of the desired frequency. If we can subsequently track the focal length of the lens in real-time, we can accurately display focal planes at any depth without waiting for the lens to settle. In other words, by driving the focus-tunable lens to periodically sweep the desired range of focal lengths and tracking the focal length at high-speed and in real-time, we can display numerous focal planes.

1.1 Contributions

This paper proposes the design of a novel multifocal display that produces three-dimensional scenes by displaying dense focal stacks. In this context, we make the following contributions:

- *High-speed focal-length tracking.* The core contribution of this paper is a system for real-time tracking of the focal length of a focus-tunable lens at microsecond-scale resolutions. We achieve this by measuring the deflection of a laser incident on the lens.
- *Design space analysis.* Displaying a dense set of focal planes is also necessary for mitigating the loss of spatial resolution due to the defocus blur caused by the ocular lens. To show this, we analytically derive the spatial resolution of the image formed on the retina when there is a mismatch between the focus of the eye and the depth at which the content is virtually rendered. This analysis justifies the need for AR/VR displays capable of a high focal-plane density.
- *Prototype.* Finally, we build a proof-of-concept prototype that is able to produce 40 8-bit focal planes per frame with 40 fps. This corresponds to 1600 focal planes per second — a capability that is an order of magnitude greater than competing approaches.

1.2 Limitations

In addition to limitations endemic to multifocal displays, the proposed approach has the following limitations:

- *Need for additional optics.* The proposed focal-length tracking device requires additional optics that increase its bulk.
- *Peak brightness.* Displaying a large number of focal planes per frame leads to a commensurate decrease in peak brightness of the display since each depth plane is illuminated for a smaller fraction of time. This is largely not a concern for VR displays, and can potentially be alleviated with techniques that redistribute light [Damberg et al. 2016].
- *Limitations of our prototype.* Our current proof-of-concept prototype uses a digital micromirror display (DMD) and, as a consequence, has low energy efficiency. The problem can be easily solved by switching to energy-efficient displays, like OLED, or laser-scanning projectors or displays that redistribute light to achieve higher peak brightness and contrast.

2 RELATED WORK

A typical VR display is composed of a convex eyepiece and a display unit. As shown in Figure 2a, the display is placed within the focal length of the convex lens in order to create a magnified virtual image. The distance $v > 0$ of the virtual image can be calculated by the thin lens formula:

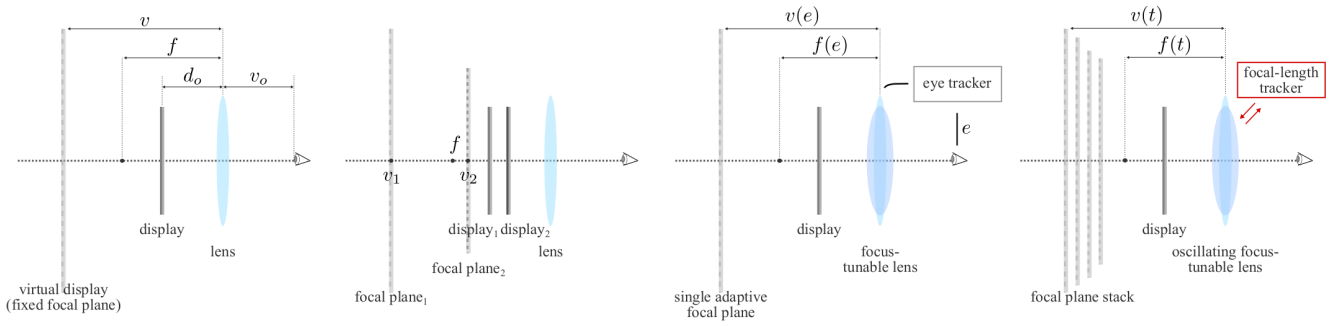
$$\frac{1}{d_o} + \frac{1}{-v} = \frac{1}{f}, \tag{1}$$

where d_o is the distance between the display and the lens, and f is the focal length. We can see that $\frac{1}{v}$ is an affine function of the optical power $(1/f)$ of the lens and the term $1/d_o$. By choosing d_o and f, the designer can put the virtual image of the display at the desired depth. However, for many applications, most scenes need to be rendered across a wide range of depths. Due to the fixed focal plane, these displays do not provide natural accommodation cues.

2.1 Accommodation-Supporting Displays

There have been many designs proposed to provide accommodation support. We concentrate on techniques most relevant to the proposed method, deferring a detailed description to [Kramida 2016] and [Hua 2017]; in particular, see Table 1 of [Matsuda et al. 2017].

2.1.1 Multifocal and Varifocal Displays. Multifocal and varifocal displays control the depths of the focal planes by dynamically adjusting f or d_o in (1). Multifocal displays aim to produce multiple focal planes at different depths for each frame (Figure 2b), whereas varifocal displays support only one focal plane per frame whose depth is dynamically adjusted based on the gaze of the user's eyes



(a) Typical VR display (b) Multifocal with fixed displays (c) Varifocal with a focus-tunable lens (d) Proposed display

Fig. 2. Typical VR displays have a fixed display with a fixed focal-length lens and thereby can output one focal plane at a fixed depth. Multifocal displays can produce multiple focal planes within a frame, using either multiple displays (shown above) or liquid lenses. Varifocal displays generate a single but adaptive focal plane using an eye tracker. The proposed display outputs dense focal plane stacks by tracking the focal-length of an oscillating focus-tunable lens. The depths of the focal planes are independent to the viewer, and thereby eye trackers are optional.

(Figure 2c). Multifocal and varifocal displays can be designed in many ways, including the use of multiple (transparent) displays placed at different depths [Akeley et al. 2004; Jannick P. Rolland 1999; Love et al. 2009], a translation stage to physically move a display or optics [Akşit et al. 2017; Shiwa et al. 1996; Sugihara and Miyasato 1998], deformable mirrors [Hu and Hua 2014], as well as a focus-tunable lens to optically reposition a fixed display [Johnson et al. 2016; Konrad et al. 2016; Lee et al. 2018; Liu et al. 2008; Padmanaban et al. 2017]. Varifocal focal displays show a single focal plane at any point in time, but they require precise eye/gaze-tracking at low latency. Multifocal displays, on the other hand, have largely been limited to displaying a few focal planes per frame due to the limited switching speed of translation stages and focus-tunable lenses. Concurrent to our work, Lee et al. [2018] propose a multifocal display that can also display dense focal stacks with a focus-tunable lens. However, their method can only display any given pixel at a single depth. This prohibits the use of rendering techniques [Akeley et al. 2004; Mercier et al. 2017; Narain et al. 2015] that require a pixel to be potentially displayed at many depths with different contents.

2.1.2 Light Field Displays. While multifocal and varifocal displays produce a collection of focal planes, light field displays aim to synthesize the light field of a 3D scene. Lanman and Luebke [2013] introduce angular information by replacing the eyepiece with a microlens array; Huang et al. [2015] utilize multiple spatial light modulators to modulate the intensity of light rays. While these displays fully support accommodation cues and produce natural defocus blur and parallax, they usually suffer from poor spatial resolution due to the space-angle resolution trade-off.

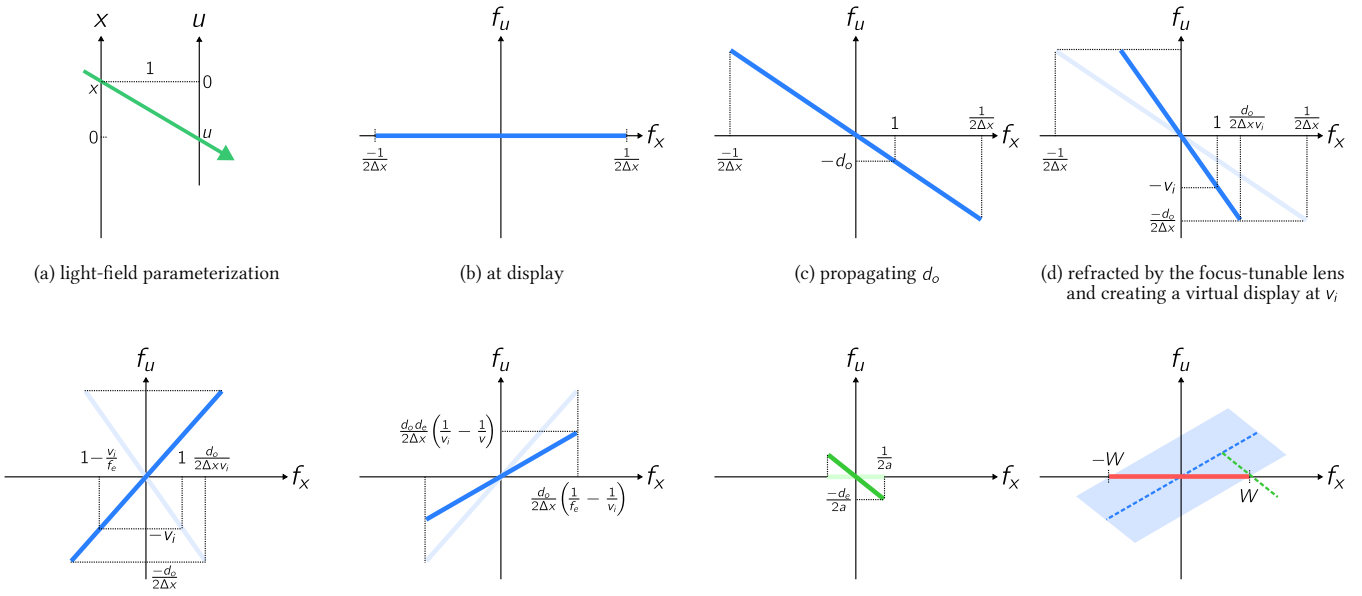
2.1.3 Other Types of Virtual Reality Displays. Other types of VR/AR displays have been proposed to solve the VAC. Matsuda et al. [2017] use a phase-only spatial light modulator to create spatially-varying lensing based on the virtual content and the gaze of the user. Maimone et al. [2017] utilize a phase-only spatial light modulator to create a 3D scene using holography. Similar to our work, Konrad et al. [2017] operate a focus-tunable lens in an oscillatory mode. Here, they use the focus-tunable lens to create a depth-invariant blur by using a concept proposed for extended depth of field imaging [Miau et al. 2013]. Intuitively, since the content is displayed at all focal planes, the VAC is significantly resolved. However, there is a loss of spatial resolution due to the intentionally introduced defocus blur.

2.2 Depth-Filtering Methods

When virtual scenes are rendered with few focal planes, there are associated aliasing artifacts as well as a reduction of spatial resolution on content that is to be rendered in between focal planes. Akeley et al. [2004] show that such artifacts can be alleviated using linear depth filtering, a method that is known to be quite effective [MacKenzie et al. 2010; Ravikumar et al. 2011]. However, linear depth filtering produces artifacts near object boundaries due to the inability of multifocal displays to occlude light. To produce proper occlusion cues with multifocal displays, Narain et al. [2015] propose a method that jointly optimizes the contents shown on all focal planes. By modeling the defocus blur of focal planes when an eye is focused at certain depths, they formulate a non-negative least-square problem that minimizes the mean-squared error between perceived images and target images at multiple depths. While this algorithm demonstrates promising results, the computational costs of the optimization are often too high for real-time applications. Mercier et al. [2017] simplify the forward model of Narain et al. [2015] and significantly improve the speed to solve the optimization problem. These filtering approaches are largely complementary to the proposed work, in that, they can be incorporated into the dense focal stacks produced by our proposed display.

3 HOW MANY FOCAL PLANES DO WE NEED?

A key factor underlying the design of multifocal displays is the number of focal planes required to support a target accommodation range. In order to be indistinguishable from the real world, a virtual world should enable human eyes to accommodate freely on arbitrary depths. In addition, the virtual world should have high spatial resolution anywhere within the target accommodation range. Simultaneously satisfying these two criteria for a large accommodation



(a) light-field parameterization

(b) at display

(c) propagating d_o

(d) refracted by the focus-tunable lens and creating a virtual display at v_i

(e) refracted by the lens in our eye, whose focal length is f_e

(f) propagating d_e to the retina

(g) aperture function propagated d_e to the retina

(h) cross-correlation of (f) and (g) results in the light field at the retina (shaded blue). Our eye sees the slice along f_x.

Fig. 3. Fourier transform of the 2-dimensional light field at each stage of a multifocal display. The display is assumed to be isotropic and has pixels of pitch Δx. (a) Each light ray in the light field is characterized by its intercepts with two parallel axes, x and u, which are separated by 1 unit, and the origin of the u-axis is relative to each individual value of x. (b) With no angular resolution, the light field spectrum emitted by the display is a flat line on f_x. We focus only on the central part ($|f_x| \leq \frac{1}{2\Delta x}$). (c) The light field propagates d_o to the tunable lens, causing the spectrum to shear along f_u. (d) Refraction due to the lens corresponds to shearing along f_x, forming a line segment of slope $-v_i$, where v_i is the depth of the focal plane. (e,f) Refraction by the lens in our eye and propagation d_e to the retina without considering the finite aperture of the pupil. (g) The spectrum of the pupil function propagates d_e to the retina. (h) The light field spectrum on the retina with a finite aperture is the 2-dimensional cross-correlation between (f) and (g). According to Fourier slice theorem, the spectrum of the perceived image is the slice along f_x, shown as the red line. The diameter of the pupil and the slope of (f), which is determined by the focus of the eye and the virtual depth v_i, determine the spatial bandwidth, W, of the perceived image.

range is very challenging, since it requires generating light fields of high spatial and angular resolution. In the following, we will show that displaying a dense focal stack is a promising step toward the ultimate goal of generating virtual worlds that can handle the accommodation cues of the human eye.

To understand the capability of a multifocal display, we can analyze its generated light field in the frequency domain. Our analysis, following the derivation in Wetzstein et al. [2011] and Narain et al. [2015], provides an upper-bound on the performance of a multifocal display, regardless of the depth filtering algorithm applied. It is also similar to that of Sun et al. [2017] with the key difference that we focus on the minimum number of focal planes required to retain spatial resolution within an accommodation range, as opposed to efficient rendering of foveated light fields.

3.1 Light-Field Parameterization and Assumptions

For simplicity, our analysis considers a flatland with two-dimensional light fields. In the flatland, the direction of a light ray is parameterized by its intercepts with two parallel axes, x and u, which are separated by 1 unit, and the origin of the u-axis is relative to each individual value of x such that u measures the tangent angle of a ray passing through x, as shown in Figure 3a. We model the human eye with a camera composed of a finite-aperture lens and a sensor plane d_e away from the lens, following the assumptions made in Mercier et al. [2017] and Sun et al. [2017]. We assume that the pupil of the eye is located at the center of the focus-tunable lens and is smaller than the aperture of the tunable lens. We assume that the display and the sensor emits and receives light isotropically. In other words, each pixel on the display uniformly emits light rays toward every direction and vice versa for the sensor. We also assume small-angle (paraxial) scenarios, since the distance d_o and the focal length of the tunable lens (or essentially, the depths of focal planes) are large compared to the diameter of the pupil. This assumption simplifies our analysis by allowing us to consider each pixel in isolation.

3.2 Light Field Generated by the Display

Since the display is assumed to emit light isotropically in angle, the light field created by a display pixel can be modeled as $\ell_d(x,u) = I\ \delta(x) * \text{rect}\left(\frac{x}{\Delta x}\right)$, where I is the radiance emitted by the pixel, $*$ represents two-dimensional convolution, and Δx is the pitch of the display pixel. The Fourier transform of $\ell_d(x,u)$ is $L_d(f_x,f_u) = \frac{I}{\Delta x}\text{sinc}(\Delta x f_x)$, which lies on the f_x axis, as shown in Figure 3b. We

only plot the central lobe of $\text{sinc}(\Delta x f_x)$ corresponding to $|f_x| \leq \frac{1}{2\Delta x}$, since this is sufficient for calculation of the half-maximum bandwidth of retinal images. In the following, we omit the constant $\frac{I}{\Delta x}$ for brevity.

3.3 Propagation from Display to Retina

Let us decompose the optical path from the display to the retina (sensor) and examine its effects in the frequency domain. After leaving the display, the light field propagates a distance d_o, gets refracted by the tunable lens, and by the lens of the eye where it is partially blocked by the pupil, whose diameter is a, and propagates a distance d_e to the retina where it finally gets integrated across angle. Propagation and refraction shears the spectrum of the light field along f_u and f_x, respectively, as shown in Figure 3(c,d,e). Before entering the pupil, the focal plane at depth v_i forms a segment of slope $-v_i$ within $|f_x| \leq \frac{d_o}{2v_i\Delta x}$, where $\frac{d_o}{v_i}$ is due to the magnification of the lens. For brevity, we show only the final (and most important) step and defer the full derivation to the appendix.

Suppose the eye focuses at depth $v = f_e d_e/(d_e - f_e)$, and the focus-tunable lens configuration creates a focal plane at v_i. The Fourier transform of the light field reaching the retina is

$$L_e(f_x, f_u) = L^{(v_i)}(f_x, f_u) \otimes A^{(d_e)}(f_x, f_u), \tag{2}$$

where $\otimes$ represents two-dimensional cross correlation, $L^{(v_i)}$ is the Fourier transform of the light field from the focal plane at v_i reaching the retina without aperture (Figure 3f), and $A^{(d_e)}$ is the Fourier transform of the aperture function propagated to the retina (Figure 3g). Depending on the virtual depth v_i, the cross correlation creates different extent of blur on the spectrum (Figure 3h). Finally, the Fourier transform of the image that is seen by the eye is simply the slice along f_x on L_e.

When the eye focuses at the focal plane ($v = v_i$), the spectrum lies entirely on f_x and the cross correlation with $A^{(d_e)}$ has no effect on the spectrum along f_x. The resulted retinal image has maximum spatial resolution $\frac{d_o}{2d_e\Delta x}$, which is independent of the depth of the focal plane v_i.

When the eye is not focused on the virtual depth plane, i.e., $v_i \neq v$, the cross correlation results in a segment of width

$$W = \frac{1}{2ad_e}\left(\left|\frac{1}{v} - \frac{1}{v_i}\right|\right)^{-1}$$

on the f_x-axis (Figure 3h). Note that $|L_e(\pm W, 0)| = \text{sinc}(0.5) \times \text{sinc}(0.5) \approx 0.4$, and thereby the half-maximum bandwidth of the spatial frequency of the perceived image is upper-bounded by W.

3.4 Spatial Resolution of Retinal Images

We can now characterize the spatial resolution of a multifocal display. Suppose the eye can accommodate freely on any depth v within a target accommodation range, $[v_a, v_b]$. Let $\mathcal{V} = \{v_1 = v_a, v_2, \ldots, v_n = v_b\}$ be the set of depth of the focal planes created by the multifocal display. When the eye focuses at v, the image formed on its retina has spatial resolution of

$$F_s(v) = \min\left\{\frac{d_o}{2d_e\Delta x}, \max_{v_i \in \mathcal{V}}\left(2ad_e\left|\frac{1}{v} - \frac{1}{v_i}\right|\right)^{-1}\right\}, \tag{3}$$

where the first term characterizes the inherent spatial resolution of the display unit, and the second term characterizes spatial resolution limited by accommodation, i.e. potential mismatch between the focus plane of the eye and the display. This bound on spatial resolution is a physical constraint caused by the finite display pixel pitch and the limiting aperture (i.e., the pupil) — even if the retina had infinitely-high spatial sampling rate. *Any post-processing methods including linear depth filtering, optimization-based filtering, and nonlinear deconvolution cannot surpass this limitation.*

3.5 Minimum Number of Focal Planes Needed

As can be seen in (3), the maximum spacing between any two focal planes in diopter determines $\min_{v \in [v_a, v_b]} F_s(v)$, the lowest perceived spatial resolution within the accommodation range. If we desire a multifocal display with spatial resolution across the accommodation range to be at least F, $F \leq \frac{d_o}{2d_e\Delta x}$, the best we can do with n focal planes is to have a constant inter-focal separation in diopter. This results in an inequality that

$$\left(\frac{2ad_e}{2n}\left(\frac{1}{v_a} - \frac{1}{v_b}\right)\right)^{-1} \geq F, \tag{4}$$

or equivalently

$$n \geq ad_e\left(\frac{1}{v_a} - \frac{1}{v_b}\right)F. \tag{5}$$

Thereby, increasing the number of focal planes n (and distributing them uniformly in diopter) is required for multifocal displays to support higher spatial resolution and wider accommodation range.

3.6 Relationship to Prior Work.

There are many prior works studying the minimum focal-plane spacing of multifocal displays. Rolland et al. [1999] compute the depth-of-focus based on typical acuity of human eyes (30 cycles per degree) and pupil diameter (4 mm) and conclude that 28 focal planes equally spaced by $\frac{1}{7}$ diopter are required to accommodate from 25 cm to ∞. Both theirs and our analyses share the same underlying principle — maintaining the minimum resolution seen by the eye within the accommodation range, and thereby provide the same required focal planes. By taking $a = 4$ mm, $d_e F = 30 \times \frac{180}{\pi}$, $v_a = 25$ cm, and $v_b = \infty$, we have $n \geq 27.5$, which concurs with their result. MacKenzie et al. [2012; 2010] measure accommodation responses of human eyes during usage of multifocal displays with different plane-separation configurations under linear depth filtering [Akeley et al. 2004]. Their results suggest that focal-plane separations as wide as 1 diopter can drive accommodation with insignificant deviation from the natural accommodation. However, it is also reported that smaller plane-separations provide more natural accommodation and higher retinal contrast — features that are desirable in any VR/AR display. By enabling dense focal stacks of focal-plane separation as small as 0.1 diopter, our prototype can simultaneously provide proper accommodation cues and display high-resolution images onto the retina.

3.7 Maximum Number of Focal Planes Needed

At the other extreme, if we have a sufficient number of focal planes, the limiting factor becomes the pixel pitch of the display unit. In

this scenario, for a focal plane at virtual depth v_i, the retinal image of an eye focuses on v will have maximal spatial resolution $\frac{d_o}{2d_e\Delta x}$ if

$$\left|\frac{1}{v}-\frac{1}{v_i}\right| \leq \frac{\Delta x}{ad_o}.$$

In other words, the depth-of-field of a focal plane — defined as the depth range that under focus provides the maximum resolution — is $\frac{2\Delta x}{ad_o}$ diopters. Since the maximum accommodation range of the multifocal display with a convex tunable lens is $\frac{1}{d_o}$ diopter, we need at least $\frac{a}{2\Delta x}$ focal planes to achieve the maximum spatial resolution of the multifocal display across the maximum supported depth range, or $\frac{D_o a d_o}{2\Delta x}$ focal planes for a depth range of D_o. For example, our prototype has $\Delta x = 13.6$ um, $d_o = 7$ cm, and pupil diameter $a = 4$ mm, it would require 147 focal planes for the maximum possible depth range of $d_o = 7$ cm to infinity or $D_o = 14.3$ diopters to reach the resolution upper-bound. For a shorter working range of 25 cm to infinity, or 4 diopters, it would require 41 focal planes.

4 GENERATING DENSE FOCAL STACKS

We now have a clear goal — designing a multifocal display supporting a very dense focal stack, which enables display high-resolution images across a wide accommodation range. The key bottleneck for building multifocal displays with dense focal stacks is the settling time of the focus-tunable lens. The concept described in this section outlines an approach to mitigate this bottleneck and provides a design template for displaying dense focal stacks.

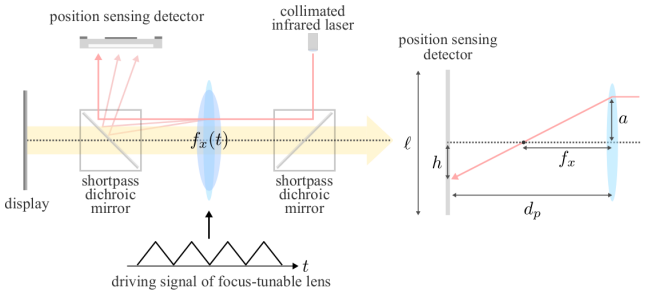
4.1 Focal-Length Tracking

The centerpiece of our proposed work is the idea that we do not have to wait for the focus-tunable lens to settle at a particular focal length. Instead, if we constantly drive the lens so that it sweeps across a range of focal lengths, and subsequently track the focal length in real time, we can display the corresponding focal plane without waiting for the focus-tunable lens to settle. This enables us to display as many focal planes as we want, as long as the display supports the required frame rate.

While the optical power of focus-tunable lenses is controlled by an input voltage or current, simply measuring these values only provides inaccurate and biased estimates of the focal length. This is due to the time-varying transfer functions of tunable lenses, which are known to be sensitive to operating temperature and irregular motor delays. Instead, we propose to estimate the focal length by probing the tunable lens optically. This enables robust estimations that are invulnerable to the unexpected factors.

In order to measure the focal length, we send a collimated infrared laser beam through the edge of the focus-tunable lens. Since the direction of the outgoing beam depends on the focal length, the laser beam changes direction as the focal length changes. There are many approaches to measure this change in direction, including using a one-dimensional pixel array or an encoder system. In our prototype, we use a one-dimensional position sensing detector (PSD) to enable fast and accurate measurement of the location. The schematic is shown in Figure 4a.

The focal length of the laser is estimated as follows. We first align the laser so that it is parallel to the optical axis of the focus-tunable lens. After deflection by the lens, the beam is incident on a spot on the PSD whose position, as shown in Figure 4b, is given as



(a) Schematic of focal-length tracking (b) Optical layout

Fig. 4. (a) The focal-length tracking system is composed of two shortpass dichroic mirrors and a position sensing detector. The dichroic mirror allows visible light to pass through but reflects the infrared light ray emitted from the collimated laser. (b) The position of the laser spot on the position sensing detector is an affine function of the optical power of the lens.

$$h = a\left(\frac{d_p}{f_x}-1\right), \tag{6}$$

where f_x is the focal length of the lens, d_p is the distance measured along the optical axis between the lens and the PSD, and h is the distance between the optical center of the lens and the spot the laser is incident on. Note that the displacement h is an affine function of the optical power of the focus-tunable lens.

We next discuss how the location of the spot is estimated from the PSD outputs. A PSD is composed of a photodiode and a resistor distributed throughout the active area. The photodiode has two connectors at its anode and a common cathode. Suppose the total length of the active area of the PSD is ℓ. When a light ray reaches a point at h on the PSD, the generated photocurrent will flow from each anode connector to the cathode with amount inversely proportional to the resistance in between. Since resistance is proportional to length, we have the ratio of the currents in the anode and cathode as

$$\frac{i_1}{i_2}=\frac{R_2}{R_1}=\frac{\frac{\ell}{2}-h}{\frac{\ell}{2}+h}, \text{ or } h=\frac{\ell}{2}\frac{i_2-i_1}{i_2+i_1}. \tag{7}$$

Combining (7) and (6), we have

$$\frac{1}{f_x}=\frac{\ell}{2ad_p}r+\frac{1}{d_p}, \text{ where } r=\frac{i_2-i_1}{i_2+i_1}. \tag{8}$$

As can be seen, the optical power of the tunable lens $\frac{1}{f_x}$ is an affine function of r. With simple calibration (to get the two coefficients), we can easily estimate the value.

4.2 The Need for Fast Displays

In order to display multiple focal planes within one frame, we also require a display that has a frame rate greater than or equal to the focal-plane display rate. To achieve this, we use a digital micromirror device (DMD)-based projector as our display. Commercially available DMDs can easily achieve upwards of 20,000 bitplanes per second. Following the design in [Chang et al. 2016], we modulate the intensity of the projector's light source to display 8-bit images;

this enables us to display each focal plane with 8-bits of intensity and generate as many as $20,000/8 \approx 2,500$ focal planes per second.

4.3 Design Criteria and Analysis

We now analyze the system in terms of various desiderata and the system configurations required to achieve them.

4.3.1 Achieving a Full Accommodation Range. A first requirement is that the system be capable of supporting the full accommodation range of typical human eyes, i.e., generate focal planes from 25 cm to infinity. Suppose the optical power of the focus-tunable lens ranges from $D_1 = \frac{1}{f_1}$ to $D_2 = \frac{1}{f_2}$ diopter. From (1), we have

$$\frac{1}{-v(t)} = \frac{1}{f_x(t)} - \frac{1}{d_o} = -\left(\frac{1}{d_o} - D_x(t)\right), \tag{9}$$

where d_o is the distance between the display unit and the tunable lens, $v(t)$ is the distance of the virtual image of the display unit from the lens, $f_x(t) \in [f_2, f_1]$ is the focal length of the lens at time t, and $D_x(t) = \frac{1}{f_x(t)}$ is the optical power of the lens in diopter. Since we want $v(t)$ to range from 25 cm to infinity, $1/v(t)$ ranges from 4 m^{-1} to 0 m^{-1}. Thereby, we need

$$4 - D_1 \leq \frac{1}{d_o} \leq D_2.$$

An immediate implication of this is that $D_2 - D_1 \geq 4$, i.e., to support the full accommodation range of a human eye, we need a focus-tunable lens whose optical power spans at least 4 diopters. We have more choice over the actual range of focal lengths taken by the lens. A simple choice is to set $1/f_2 = D_2 = 1/d_o$; this ensures that we can render focal planes at infinity; subsequently, we choose f_1 sufficiently large to cover 4 diopters. By choosing a small value of f_2, we can have a small d_o and thereby achieve a compact display.

4.3.2 Field-of-View. The proposed display shares the same field-of-view and eye box characteristics with other multifocal displays. The field-of-view will be maximized when the eye is located right near the lens. This will results in a field-of-view of $2\,\mathrm{atan}\left(\frac{H}{2d_o}\right)$, where H is the height (or width) of the physical display (or its magnification image via lensing). When the eye is further away from the lens, the numerical aperture will limit the extent of the field-of-view. Since the apertures of most tunable lenses are small (around 1 cm in diameter), we would prefer to put the eye as close as the lens as possible. This can be achieved by embedding the dichroic mirror (the right one in Figure 4a) onto the rim of the lens. For our prototype that will be described in Section 5, we use a $4f$ system to relay the eye to the aperture of the focus-tunable lens. Our choice of the $4f$ system enables a 45-degree field-of-view, limited by the numerical aperture of the lens in the $4f$ system.

There are alternate implementations of focus tunable lenses that have the potential for providing larger apertures and hence, displays with larger field of views. Bernet and Ritsh-Marte [2008] design two phase plates that produce the phase function of a lens whose focal length is determined by the relative orientation of the plates; hence, we could obtain a large aperture focus tunable lens by rotating one of the phase plates. Other promising solutions to enable large-aperture tunable lensing include the Fresnel and Pancharatnam-Berry liquid crystal lenses [Jamali et al. 2018a,b] and tunable metasurface doublets [Arbabi et al. 2018]. In all of these cases, our tracking method could be used to provide precise estimates of the focal length.

4.3.3 Eye Box. The eye box of multifocal displays are often small, and the proposed display is no exception. Due to the depth difference of focal planes, as the eye shifts, contents on each focal plane shift by different amounts, with the closer ones traverse more than the farther ones. This will leave uncovered as well as overlapping regions at depth discontinuities. Further, the severity of the artifacts depends largely on the specific content being displayed. In practice, we observe that these artifacts are not distracting for small eye movements in the order of few millimeters. This problem can be solved by incorporating an eye tracker, as in Mercier et al. [2017].

4.4 Reduced Maximum Brightness and Energy Efficiency

Key limitations of our proposed design are the reduction in maximum brightness and, depending on the implementation, the energy efficiency of the device. Suppose we are displaying n focal planes per frame and T frames per second. Each focal plane is displayed for $\frac{T}{n}$ second, which is n-times smaller compared to typical VR displays with one focal plane. For our prototype, we use a high power LED to compensate for the reduction in brightness. Further, brightness of the display is not a primary concern since there are no competing ambient lights sources for VR displays.

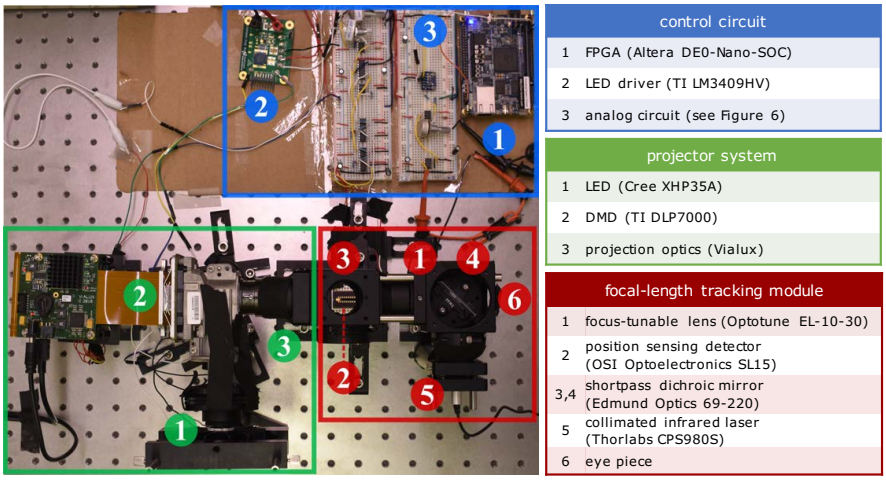
Energy efficiency of the proposed method also depends on the type of display used. For our prototype, since we use a DMD to spatially modulate the intensity at each pixel, we waste $\frac{n-1}{n}$ of the energy. This can be completely avoided by adopted by using OLED displays, where a pixel can be completely turned off. An alternate solution is to use a phase spatial light modulator (SLM) [Damberg et al. 2016] to spatially redistribute a light source so that each focal plane only gets illuminated at pixels that need to be displayed; a challenge here is the slow refresh rate of the current crop of phase SLMs. Another option is to use a laser along with a 2D galvo to selectively illuminate the content at each depth plane; however, 2D galvos are often slow when operated in non-resonant modes.

5 PROOF-OF-CONCEPT PROTOTYPE

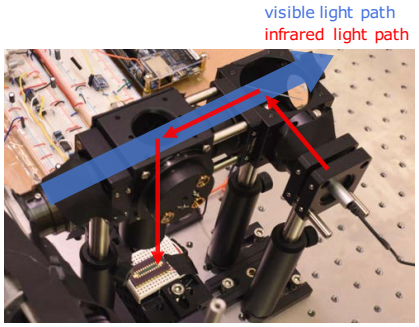
In this section, we present a lab prototype that generates a dense focal stack using high-speed tracking of the focal length of a tunable lens and a high-speed display.

5.1 Implementation Details

The prototype is composed of three functional blocks: the focus-tunable lens, the focal-length tracking device, and a DMD-based projector. All the three components are controlled by an FPGA (Altera DE0-nano-SOC). The FPGA drives the tunable lens with a digital-to-analog converter (DAC), following Algorithm 1. Simultaneously, the FPGA reads the focal-length tracking output with an analog-to-digital converter (ADC) and uses the value to trigger the projector to display the next focal plane. Every time a focal plane has been displayed, the projector is immediately turned off to avoid blur caused by the continuously changing focal-length configurations. A photo of the prototype is shown in Figure 5. In the following, we will introduce each component in detail.



(a) Photograph of the prototype and its component list



(b) Light path for infrared and visible light

Fig. 5. The prototype is composed of a projector, the proposed focal-length tracking module, and the control circuits. (b) The two shortpass dichroic mirrors allow visible light to pass through and reflect infrared. The enables us to create individual light path for each of them.

5.1.1 Calibration. In order to display focal planes at correct depths, we need to know the corresponding PSD tracking outputs. From equations (8) and (9), we have

$$\frac{1}{v(t)} = \frac{1}{d_o} - \frac{1}{d_p} - \frac{\ell}{2ad_p} r(t) = \alpha + \beta r(t). \tag{10}$$

Thereby, we can estimate the current depth $v(t)$ if we know α and β, which only requires two measurements to estimate. With a camera focused at $v_a = 25$ cm and $v_b = \infty$, we get the two corresponding ADC readings r_a and r_b. The two points can be accurately measured, since the depth-of-field of the camera at 25 cm is very small, and infinity can be approximated as long as the image is far away. Since (10) has an affine relationship, we only need to divide $[r_a, r_b]$ evenly into the desired number of focal planes.

5.1.2 Control Algorithm. The FPGA follows Algorithm 1 to coordinate the tunable lens and the projector. On a high level, we drive the tunable lens with a triangular wave by continuously increasing/decreasing the DAC levels. We simultaneously detect the PSD's DAC reading r to trigger the projection of focal planes. When the last/first focal plane is displayed, we switch the direction of the waveform. Note that while Algorithm 1 is written in serial form, every module in the FPGA runs in parallel.

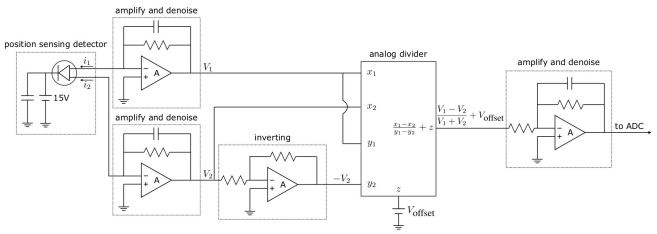
The control algorithm is simple yet robust. It is known that the transfer function of the tunable lens is sensitive to many factors, including device temperature and unexpected motor delay and errors [Optotune 2017]. In our experience, even with the same input waveform, we observe different offsets, peak-to-peak values on the PSD output waveform for each period. Since the algorithm does not drive the tunable lens with fixed DAC values and instead directly detect the PSD output (i.e., the focal length of the tunable lens), it is robust to these unexpected factors. However, the robustness comes with a price. Due to the motor delay, the peak-to-peak value $r_{\max} - r_{\min}$ is often a lot larger than $r_n - r_1$. This causes the frame rate of the prototype (1600 focal planes per second, or 40 focal planes per frame at 40 fps) to be lower than the highest display frame rate (2500 focal planes per second).

ALGORITHM 1: Tunable-lens and focal-plane control

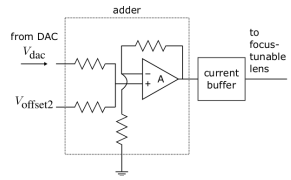
Data: n target PSD triggers $r_1, \ldots, r_n$
Input: PSD ADC reading r
Output: Tunable-lens DAC level L, projector display control signal
Initialize $L = 0$, $\Delta L = 1$, $i = 1$
repeat
  $L \leftarrow L + \Delta L$
  if $|r - r_i| \le \Delta r$ **then**
    Display focal plane i and turn it off when finished.
    $i \leftarrow i + \Delta L$
    if $\Delta L == 1$ *and* $i > n$ **then**
      Change triangle direction to down: $\Delta L \leftarrow -1$, $i \leftarrow n$
    else if $\Delta L == -1$ *and* $i < 1$ **then**
      Change triangle direction to up: $\Delta L \leftarrow +1$, $i \leftarrow 1$
  end
until *manual stop*;

Note that since 40 fps is close to the persistence of vision, our prototype sometimes leads to flickering. However, the capability of the proposed device is to increase the number of focal planes per second and as such we can get higher frame rate by trading off the focal planes per frame. For example, we can achieve 60 fps by operating at 26 focal planes per frame.

5.1.3 Focus-Tunable Lens and its Driver. We use the focus-tunable lens EL-10-30 from Optotune [Optotune 2017]. The optical power of the lens ranges from approximately 8.3 to 20 diopters and is an affine function of the driving current input from 0 to 300 mA. We use a 12-bit DAC (MCP4725) with a current buffer (BUF634) to drive the lens. The DAC provides 200 thousand samples per second, and the current buffer has a bandwidth of 30 MHz. This allows us to faithfully create a triangular input voltage up to several hundred Hertz. The circuit is drawn in Figure 6b.



(a) Analog circuit for processing focal-length tracking



(b) Analog circuit for driving focus-tunable lens

Fig. 6. Analog circuits used in the prototype. All the operational amplifiers are TI OPA-37, the analog divider is TI MPY634, and the current buffer is TI BUF634. All denoising RC circuits have cutoff frequency at 47.7 kHz.

5.1.4 Focal-Length Tracking and Processing. The focal-length tracking device is composed of a one-dimensional PSD (SL15 from OSI Optoelectronics), two 800 *n*m dichroic short-pass mirrors (Edmundoptics #69-220), and a 980 *n*m collimated infrared laser (Thorlabs CPS980S). We drive the PSD with a reverse bias voltage of 15 V. This enables us to have 15 *u*m precision on the PSD surface and rise time of 0.6 *u*s. Across the designed accommodation range, the laser spot traverses within 7 *m*m on the PSD surface, which has a total length 15 *m*m. This allows us to accurately differentiate up to 466 focal-length configurations.

The analog processing circuit has three stages — amplifier, analog calculation, and an ADC, as shown in Figure 6a. We use two operational amplifiers (TI OPA-37) to amplify the two output current of the PSD. The gain-bandwidth of the amplifiers are 45 MHz, which can fully support our desired operating speeds. We also add a low-pass filter with a cut-off frequency of 47.7 kHz at the amplifier, as a denoising filter. The computation of $r(t)$ is conducted with two operational amplifiers (TI OPA-37) and an analog divider (TI MPY634). We use a 12-bit ADC (LTC2308) with a rate of 200 thousand samples per second to port the analog voltage to the FPGA.

Overall, the latency of the focal-length tracking circuit is ~ 20 *u*s. The bottleneck is the low-pass filter and the ADC; rest of the components have time responses in nanoseconds. Note that in 20 *u*s the focal length of the tunable lens changes by 0.01 diopters — well below the detection capabilities of the eye [Campbell 1957]. Also, the stability of the acquired focal stack (which took a few hours to capture) indicates that the latency was either minimal or at least predictable and can be dealt with by calibration.

5.1.5 DMD-based Projector. The projector is composed of a DLP-7000 DMD from Texas Instruments, projection optics from Vialux, and a high-power LED XHP35A from Cree. We control the DMD with a development module Vialux V-7000. We update the configuration of micro-mirrors every 50 *u*s. Following Chang et al. [2016], we use pulse-width modulation, performed through a LED driver

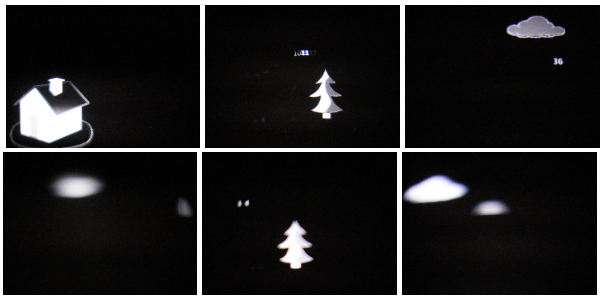


Fig. 7. Example images that are captured in burst shooting mode with a $f/4$ lens, exposure time equal to 0.5 *m*s, and ISO equal to 12, 800. Note that in order to capture a single focal plane, we need exposure time of 0.2 *m*s. Thereby, these images are composed of at most 3 focal planes.

(TI LM3409HV), to change the intensity of the LED concurrently with the update of micro-mirrors. This enables us to display at most 2500 8-bit images per second.

For simplicity, we preload each of the 40 focal planes onto the development module. Each focal stack requires 40×8 = 320 bitplanes, and thereby, we can store up to 136 focal stacks on the module. The lack of video-streaming capability needs further investigation to make it practical; it could potentially be resolved by using the customized display controller in [Lincoln et al. 2017, 2016] that is capable of displaying bitplanes with 80 *u*s latency. This would enable us to display 1562 8-bit focal planes per second. We also note that whether we use depth filtering or not, the transmitted bitplanes are sparse since each pixel has content, at best, at a few depth planes. Thereby, we do not need to transmit the entire 320 bitplanes.

Note that we divide the 8 bitplanes of each focal planes into two groups of 4 bitplanes, and we display the first group when the triangular waveform is increasing, and the other at the downward waveform. From the results that will be presented in Section 6, we can see that the images of the two groups align nicely. This demonstrates the high accuracy of the focal-length tracking.

As a quick verification of the prototype, we used the burst mode on the Nikon camera to capture multiple photographs at an aperture of $f/4$, ISO 12,800 and an exposure time of 0.5 *m*s. Figure 7 shows six examples of displayed focal planes. Since a single focal plane requires an exposure time of 50×4 = 0.2 *m*s, the captured images are composed of at most 3 focal planes.

6 EXPERIMENTAL EVALUATIONS

We showcase the performance of our prototype on a range of scenes designed carefully to highlight the important features of our system. The supplemental material has video illustrations that contain full camera focus stacks of all results in this section.

6.1 Focal-Length Tracking

To evaluate the focal-length tracking module, we measure the input signal to the focus-tunable lens and the PSD output *r* from an Analog Discovery oscilloscope. The measurements are shown in Figure 8. As can be seen, the output waveform matches that of the input. The high bandwidth of the PSD and the analog circuit enables us to track the focal length robustly in real-time. From the figure, we can also observe the delay of the focus-tunable lens (~ 3 *m*s).

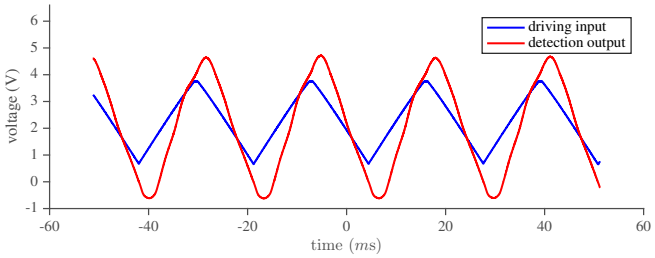


Fig. 8. Measurements of the input signal to the tunable lens and the output of the PSD after analog processing. The output waveform matches that of the input. This shows that the proposed focal-length tracking is viable.

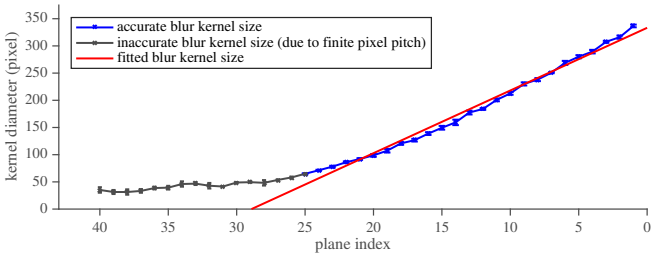


Fig. 9. Measured blur kernel diameter by a camera focusing at infinity (plane 40). Due to the finite pixel pitch, the estimation becomes inaccurate when the spot size is too small (when the spots are displayed on focal planes close to infinity). When the blur kernel size can be accurately estimated, they fit nicely as a linear segment. This indicates the depth of the focal planes are distributed uniformly in diopter.

6.2 Depths of Focal Planes

As stated previously, measuring the depth of the displayed focal planes is very difficult. Thereby, we use a method similar to depth-from-defocus to measure their depths. When a camera is focusing at infinity, the defocus blur kernel size will be linearly dependent on the depth of the (virtual) object in diopter. This provides a method to measure the depths of the focal planes.

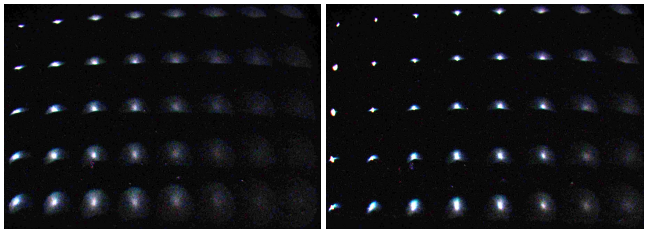
For each of the focal plane, we display a 3×3 pixels white spot at the center, capture multiple images of various exposure time, and average the images to reduce noise. We label the diameter of the defocus blur kernels and show the results in Figure 9. As can be seen, when the blur-kernel diameters can be accurately estimated, i.e., largely defocus spots on closer focal planes, the values fit nicely to a straight line, indicating the depths of focal planes are uniformly separated in diopter. However, as the displayed spot size as a spot come into focus, the estimation of blur kernel diameters becomes inaccurate since we cannot display an infinitesimal spot due to the finite pixel pitch of the display. Since there were no special treatments to individual planes in terms of system design or algorithm, we expect these focal planes to be placed accurately as well.

6.3 Characterizing the System Point-Spread Function

To characterize our prototype, we measure its point spread function with a Nikon D3400 using a 50 mm $f/1.4$ prime lens. We display a static scene that is composed of 40 3×3 spots with each spot at a different focal plane. Using the camera, we capture a focal stack of 169 images ranging from 0 to 4 diopters away from the focus-tunable lens. For improved contrast, we remove the background and noise due to dust and scratches on the lens by capturing the same focal stack with no spot shown on the display. Figure 10 shows the point spread function of the display at four different focus settings, and a video of this focal stack is attached in the supplemental material. The result shows that the prototype is able to display the spots at 40 depths concurrently within a frame, verifies the functionality of the proposed method. The shape and the asymmetry of the blur kernels can be attributed to the spherical aberration of the focus-tunable lens as well as the throw of the projection lens on the DMD.

6.4 Benefits of Dense Focal Stacks

To evaluate the benefit provided by dense focal stacks, we simulate two multifocal displays, one with 4 focal planes and the other with 40 focal planes. The 40 focal planes are distributed uniformly in diopter



(a) Camera focuses at 25 cm (b) Camera focuses at 30 cm

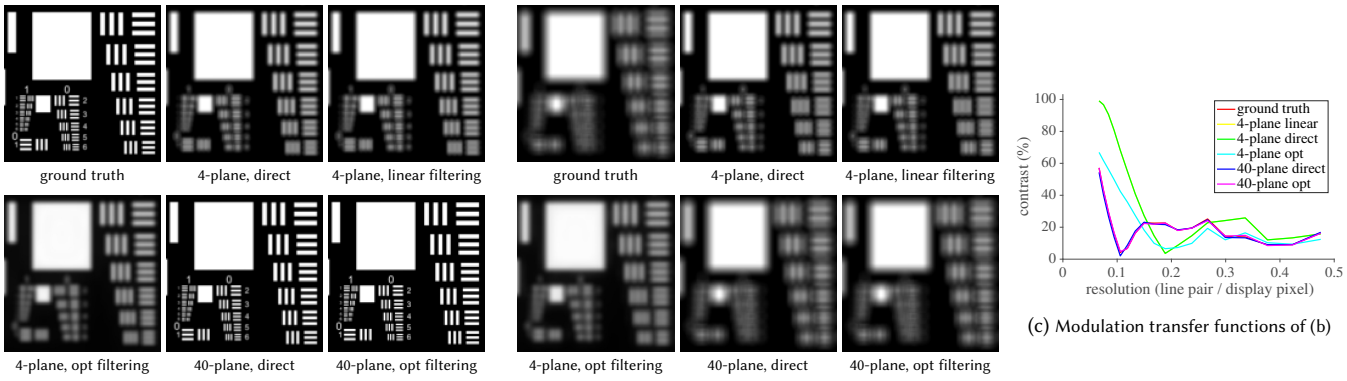
(c) Camera focuses at ~ 1.2 m (d) Camera focuses at infinity

Fig. 10. Measured point spread function of the prototype. Each of the 40 points is on a different focal plane — the top-left is closest to the camera and the bottom-right is farthest. For better visualization, we multiply the image by 10 and filter the image with a 4×4 median filter. The results show that the prototype is able to produce 40 distinct focal planes.

from 0 to 4 diopters, and the 4-plane display has focal planes at the depth of the 5th, 15th, 25th, and 35th focal planes of the 40-plane display. The scene is composed of 28 resolution charts, each at a different depth from 0 to 4 diopters (please refer to the supplemental material for figures of the entire scene). The dimension of the scene is 1500×2000 pixels.

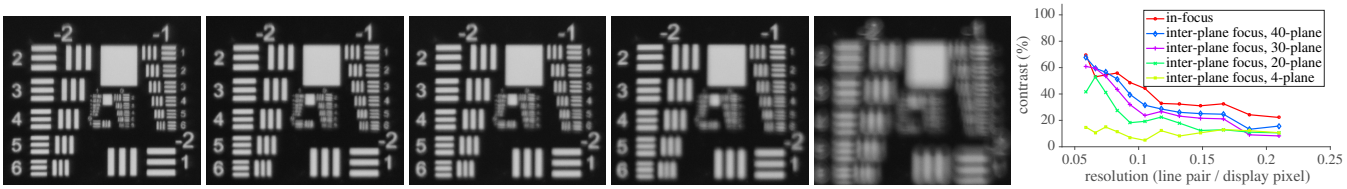
We render the scene with three methods:

- *No depth filtering*: We directly quantize the depth channel of the images to obtain the focal planes of different depths.
- *Linear depth filtering*: Following [Akeley et al. 2004], we apply a triangular filter on the focal planes based on their depths.
- *Optimization-based filtering*: We follow the formulation proposed in [Mercier et al. 2017]. We first rendered normally the desired retinal images focused at 81 depths uniformly distributed across 0 to 4 diopters in the scene with a pupil diameter of 4 mm. Then we solve the optimization problem to get the content to be displayed



(a) Camera focuses at 0.02 diopters (b) Camera focuses at 0.9 diopters (c) Modulation transfer functions of (b)

Fig. 11. Simulation results of 4-plane and 40-plane multifocal displays with direct quantization, linear depth filtering, and optimization-based filtering. The scene is at 0.02 diopters, which is an inter-plane location of the 4-plane display. (a) When the camera focuses at 0.02 diopters, the 40-plane display achieves higher spatial resolution than the 4-plane display, regardless of the depth filtering algorithm. (b) When the camera focuses at 0.9 diopters, the defocus blur on the 40-plane display closely follows that of the ground truth, whereas the 4-plane display fails to blur the low frequency contents. This can also be seen from the modulation transfer function plotted in (c).



(a) in-focus (b) 40-plane (c) 30-plane (d) 20-plane (e) 4-plane (f) modulation transfer function

Fig. 12. Captured inter-plane focused images using a 50 mm $f/1.4$ lens. The resolution chart locates on the 5th focal plane of the 40-plane display. We emulate a 4-plane and a 20-plane display by putting their focal planes on the 5, 15, 25, 35th and on the odd focal planes of the 40-plane display, respectively. (a) Camera focuses at the 5th focal plane. (b,c) Cameras focus at the estimated inter-plane locations of the 40-plane display and the 30-plane displays, respectively. (d) Camera focuses at the 6th focal plane, an inter-plane location of a 20-plane display. (e) Camera focuses at the 10th focal plane, an inter-plane location of a 4-plane display. Their modulation transfer functions are plotted in (f).

on the focal planes. We initialize the optimization process with the results of direct quantization and perform gradient descent with 500 iterations to ensure convergence.

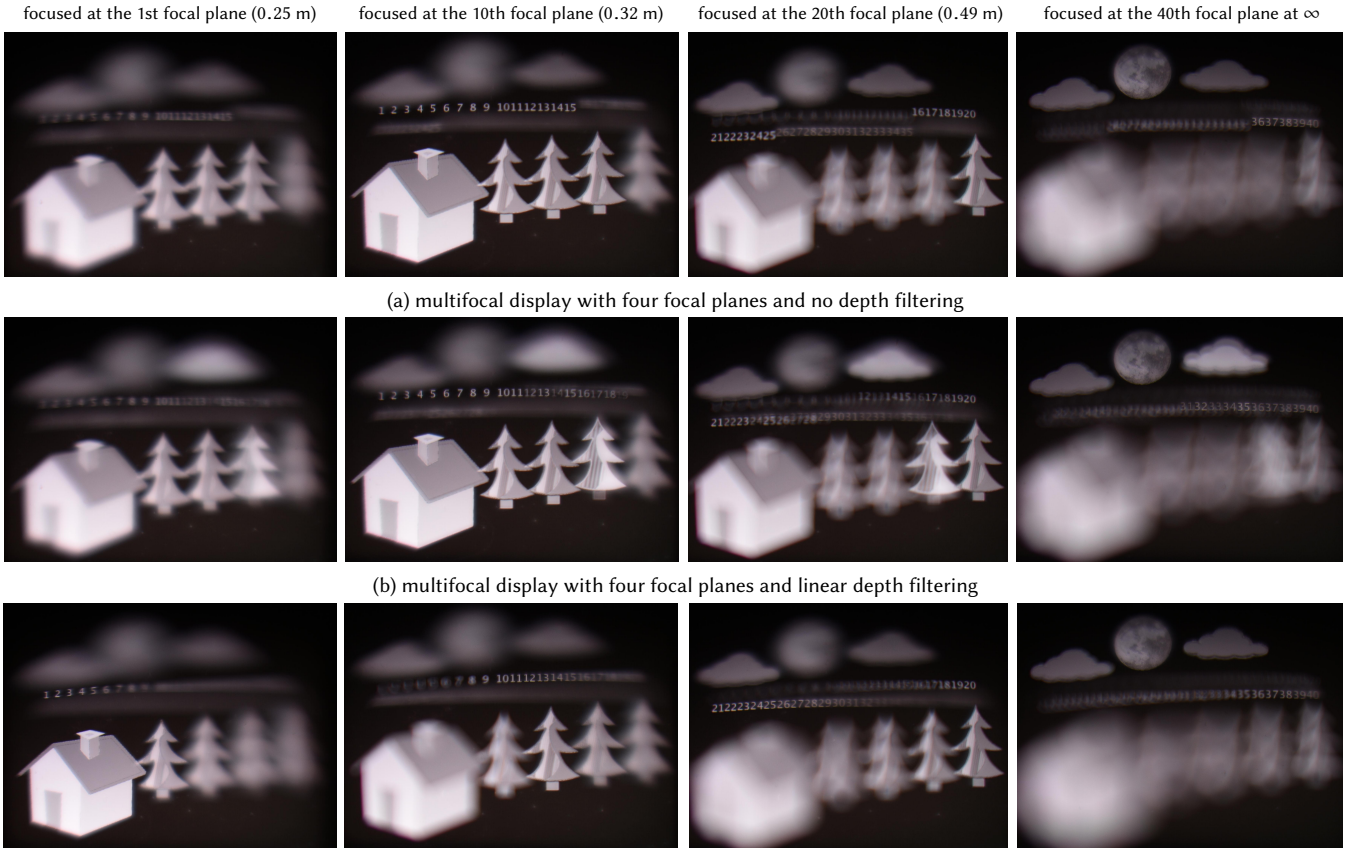
The perceived images of the resolution chart at 0.02 diopters are shown in Figure 11; a plane at 0.02 diopters is on a focal plane of the 40-plane display and is at the furthest inter-focal plane of the 4-plane display. Note that we simulate the results with pupil diameter of 4 mm, which is a typical value used to simulated retinal images of human eyes.

As can be seen from the results, the perceived images of the 40-plane display closely follow those of the ground truth — with high spatial resolution if the camera is focused on the plane (Figure 11a) and natural retinal blur when the camera is not focused (Figure 11b). In comparison, at its inter-plane location (Figure 11a), the 4-plane display has much lower spatial resolution than the other display, regardless of the depth filtering methods applied. These results verify our analysis in Section 3.

To evaluate the benefit provided by dense focal stacks in providing higher spatial resolution when the eye is focused at an inter-plane location, we implement four multifocal displays with 4, 20, 30 and 40 focal planes, respectively, on our prototype. The 4-plane display has its focal planes on the 5, 15, 25, 35th focal planes of the 40-plane display, and the 20-plane display has its focal planes on all the odd-numbered focal planes. We display a resolution chart on the fifth focal plane of the 40-plane display; this corresponds to a depth plane that all three displays can render.

To compare the worst-case scenario where an eye focuses on an inter-plane location, we focus the camera at the middle of two consecutive focal planes of each of the displays. In essence, we are reproducing the effect of VAC where the vergence cue forces the ocular lens to focus on an inter-focal plane. For the 40-plane display, this is between focal planes five and six. For the 20-plane display, this is on the sixth focal plane of the 40-plane display. And for the 4-plane display, this is on the tenth focal plane of the 40-plane display. We also focus the camera on the estimated inter-plane location of a 30-plane display. The results captured by a camera with a 50 mm $f/1.4$ lens are shown in Figure 12. As can be seen, the higher number of focal planes (smaller focal-plane separation) results in higher spatial resolution at inter-plane locations.

Next, we compare our prototype with a 4-plane multifocal display on a real scene. Note that we implement the 4-plane multifocal



(a) multifocal display with four focal planes and no depth filtering

(b) multifocal display with four focal planes and linear depth filtering

(c) proposed display with 40 focal planes and no depth filtering

Fig. 13. Comparison of a typical multifocal display with 4 focal planes and the proposed display with 40 focal planes. The four focal planes of the multifocal display correspond to the 10th, 20th, 30th, and 40th focal plane. Images are captured with a 50 mm $f/1.4$ lens. Except for the first column, these focal planes are selected such that the 4-plane multifocal display (a) is in sharp focus. In the scene, the digits are at their indicated focal planes; the house is at the first focal plane; the trees from left to right are at 5, 10, 15, 20th focal planes; the clouds and the moon are at 30, 35, 40th, respectively.

display with our 40-plane prototype by showing contents on the 10, 20, 30, 40th focal planes. The images captured by the camera are shown in Figure 13. For the 4-plane multifocal display, when used without linear depth filtering, virtual objects at multiple depths are focus/defocus as groups; when used with linear depth filtering, same objects appearing in two focal planes reduces the visibility and thereby lowers the resolution of the display. In comparison, the proposed method produces smooth focus/defocus cues across the range of depths, and the perceived images at inter-plane locations (e.g. 0.25 m) have higher spatial resolution than the 4-plane display.

Finally, we render a more complex scene [eMirage] using Blender. From the rendered all-in-focus image and its depth map, we perform linear filtering and display the results with the prototype. Focus stack images captured using a camera are shown in Figure 14. We observe realistic focus and defocus cues in the captured images.

7 CONCLUSION

This paper provides a simple but effective technique for displaying virtual scenes that are made of a dense collection of focal planes. Despite the bulk of our current prototype, the proposed tracking technique is fairly straightforward and extremely amenable to miniaturization. We believe that the system proposed in the paper for high-speed tracking could spur innovation in not just virtual and augmented reality systems but also in traditional light field displays.

ACKNOWLEDGMENTS

The authors acknowledge support via the NSF CAREER grant CCF-1652569 and a gift from Adobe Research.

REFERENCES

Kaan Akşit, Ward Lopes, Jonghyun Kim, Peter Shirley, and David Luebke. 2017. Near-eye Varifocal Augmented Reality Display Using See-through Screens. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 189:1–189:13.

Kurt Akeley, Simon J Watt, Ahna Reza Girshick, and Martin S Banks. 2004. A Stereo Display Prototype with Multiple Focal Distances. *ACM Transactions on Graphics (TOG)* 23, 3 (2004), 804–813.

Ehsan Arbabi, Amir Arbabi, Seyedeh Mahsa Kamali, Yu Horie, MohammadSadegh Faraji-Dana, and Andrei Faraon. 2018. MEMS-tunable Dielectric Metasurface Lens. *Nature Communications* 9, 1 (2018), 812.

Stefan Bernet and Monika Ritsch-Marte. 2008. Adjustable Refractive Power From Diffractive Moiré Elements. *Applied Optics* 47, 21 (2008), 3722–3730.

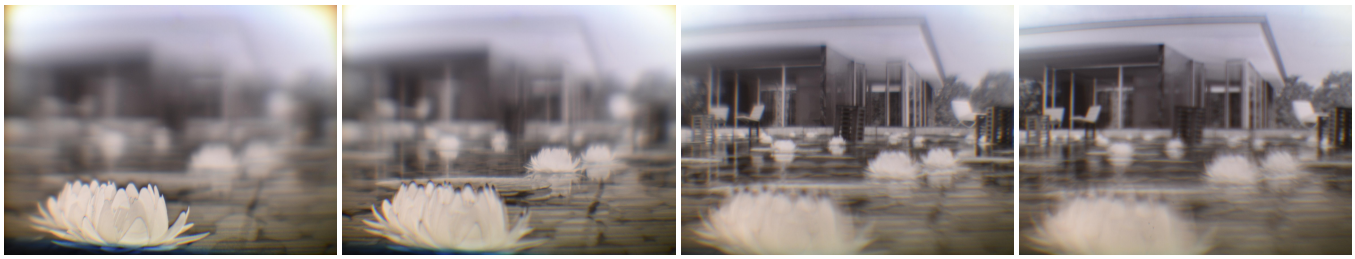


Fig. 14. Captured images focusing from near (shown at left) to far (shown at right) of a simulated scene rendered by Blender. The scene depth ranges from 50 cm (the flower at the bottom left) to infinity (the sky). The camera has a 50 mm $f/1.4$ lens. Three-dimensional scene courtesy eMirage.

Fergus W Campbell. 1957. The Depth of Field of the Human Eye. *Optica Acta: International Journal of Optics* 4, 4 (1957), 157–164.

Jen-Hao Rick Chang, BVK Vijaya Kumar, and Aswin C Sankaranarayanan. 2016. 2^{16} Shades of Gray: High Bit-depth Projection using Light Intensity Control. *Optics Express* 24, 24 (2016), 27937–27950.

Gerwin Damberg, James Gregson, and Wolfgang Heidrich. 2016. High brightness HDR projection using dynamic freeform lensing. *ACM Transactions on Graphics (TOG)* 35, 3 (2016), 24:1–24:11.

eMirage. 2017. Barcelona Pavillion. https://download.blender.org/demo/test/pabellon_barcelona_v1.scene_.zip.

David M Hoffman, Ahna 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), 13896–13903.

Hong Hua. 2017. Enabling Focus Cues in Head-mounted Displays. *Proc. IEEE* 105, 5 (2017), 805–824.

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 Transactions on Graphics (TOG)* 34, 4 (2015), 60:1–60:12.

Afsoon Jamali, Douglas Bryant, Yanli Zhang, Anders Grunnet-Jepsen, Achintya Bhowmik, and Philip J Bos. 2018a. Design of a Large Aperture Tunable Refractive Fresnel Liquid Crystal Lens. *Applied Optics* 57, 7 (2018), B10–B19.

Afsoon Jamali, Comrun Yousefzadeh, Colin McGinty, Douglas Bryant, and Philip Bos. 2018b. A Continuous Variable Lens System to Address the Accommodation Problem in VR and 3D Displays. In *Imaging and Applied Optics*. 3Tu2G.5.

Alexei A. Goon Jannick P. Rolland, Myron W. Krueger. 1999. Dynamic Focusing in Head-mounted Displays. *Proceeding of SPIE* 3639 (1999), 3639–3639–8.

Paul V Johnson, Jared AQ Parnell, Joohwan Kim, Christopher D Saunter, Gordon D Love, and Martin S Banks. 2016. Dynamic Lens and Monovision 3D Displays to Improve Viewer Comfort. *Optics Express* 24, 11 (2016), 11808–11827.

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. In *Conference on Human Factors in Computing Systems (CHI)*. 1211–1220.

Robert Konrad, Nitish Padmanaban, Keenan Molner, Emily A Cooper, and Gordon Wetzstein. 2017. Accommodation-invariant Computational Near-eye Displays. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 88:1–88:12.

Gregory Kramida. 2016. Resolving the Vergence-accommodation Conflict in Head-mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 22, 7 (2016), 1912–1931.

Douglas Lanman and David Luebke. 2013. Near-eye Light Field Displays. *ACM Transactions on Graphics (TOG)* 32, 6 (2013), 220:1–220:10.

Seungjae Lee, Youngjin Jo, Dongheon Yoo, Jaebum Cho, Dukho Lee, and Byoungho Lee. 2018. TomoReal: Tomographic Displays. *arXiv:1804.04619* (2018).

Peter Lincoln, Alex Blate, Montek Singh, Andrei State, Mary C. Whitton, Turner Whitted, and Henry Fuchs. 2017. Scene-adaptive High Dynamic Range Display for Low Latency Augmented Reality. In *Proceedings of the 21st ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*.

Peter Lincoln, Alex Blate, Montek Singh, Turner Whitted, Andrei State, Anselmo Lastra, and Henry Fuchs. 2016. From Motion to Photons in 80 Microseconds: Towards Minimal Latency for Virtual and Augmented Reality. *Transactions on Visualization and Computer Graphics* 22, 4 (2016), 1367–1376.

Sheng Liu, Dewen Cheng, and Hong Hua. 2008. An Optical See-through Head Mounted Display with Addressable Focal Planes. In *IEEE/ACM International Symposium on Mixed and Augmented Reality*. 33–42.

Sheng Liu and Hong Hua. 2009. Time-multiplexed Dual-focal Plane Head-mounted Display with a Liquid Lens. *Optics Letters* 34, 11 (2009), 1642–1644.

Patrick Llull, Noah Bedard, Wanmin Wu, Ivana Tosic, Kathrin Berkner, and Nikhil Balram. 2015. Design and Optimization of a Near-eye Multifocal Display System for Augmented Reality. In *Imaging and Applied Optics*. JTH3A.5.

Gordon D Love, David M Hoffman, Philip JW 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), 15716–15725.

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 (2012), 21–21–9.

Kevin J MacKenzie, David M Hoffman, and Simon J Watt. 2010. Accommodation to Multiple-focal-plane Displays: Implications for Improving Stereoscopic Displays and for Accommodation Control. *Journal of Vision* 10, 8 (2010), 22.

Andrew Maimone, Andreas Georgiou, and Joel S Kollin. 2017. Holographic Near-eye Displays for Virtual and Augmented Reality. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 85:1–85:16.

Nathan Matsuda, Alexander Fix, and Douglas Lanman. 2017. Focal Surface Displays. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 86:1–86:14.

Olivier Mercier, Yusufu Sulai, Kevin Mackenzie, Marina Zannoli, James Hillis, Derek Nowrouzezahrai, and Douglas Lanman. 2017. Fast Gaze-contingent Optimal Decompositions for Multifocal Displays. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 237:1–237:15.

Daniel Miau, Oliver Cossairt, and Shree K Nayar. 2013. Focal Sweep Videography with Deformable Optics. In *IEEE Conference on Computational Photography (ICCP)*.

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 Transactions on Graphics (TOG)* 34, 4 (2015), 59:1–59:12.

Optotune. 2017. Optotune Electrically Tunable Lens EL-10-30. http://www.optotune.com/images/products/Optotune.

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. *Proceedings of the National Academy of Sciences* 114, 9 (2017), 2183–2188.

Sowmya Ravikumar, Kurt Akeley, and Martin S Banks. 2011. Creating Effective Focus Cues in Multi-plane 3D Displays. *Optics Express* 19, 21 (2011), 20940–20952.

Shinichi Shiwa, Katsuyuki Omura, and Fumio Kishino. 1996. Proposal for a 3-D Display with Accommodative Compensation: 3DDAC. *Journal of the Society for Information Display* 4, 4 (1996), 255–261.

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. In *ITE Technical Report 22.1*. 33–36.

Qi Sun, Fu-Chung Huang, Joohwan Kim, Li-Yi Wei, David Luebke, and Arie Kaufman. 2017. Perceptually-guided Foveation for Light Field Displays. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 192:1–192:13.

Nelson V Tabiryan, Svetlana V Serak, David E Roberts, Diane M Steeves, and Brian R Kimball. 2015. Thin Waveplate Lenses of Switchable Focal Length–New Generation in Optics. *Optics express* 23, 20 (2015), 25783–25794.

Varioptic. 2017. Varioptic Variable Focus Liquid Lens ARCTIC 25H. http://varioptic.com/media/cms_page_media/45/MADS_-_160429_-_Arctic_25H_family.pdf.

Dhanraj Vishwanath and Erik Blaser. 2010. Retinal Blur and the Perception of Egocentric Distance. *Journal of Vision* 10, 10 (2010), 26.

Simon J Watt, Kurt Akeley, Marc O Ernst, and Martin S Banks. 2005. Focus Cues Affect Perceived Depth. *Journal of Vision* 5, 10 (2005), 7.

Gordon Wetzstein, Douglas Lanman, Wolfgang Heidrich, and Ramesh Raskar. 2011. Layered 3D: Tomographic Image Synthesis for Attenuation-based Light Field and High Dynamic Range Displays. *ACM Transactions on Graphics (TOG)* 30, 4 (2011), 95:1–95:12.

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.