

An Extended Depth-of-Field Volumetric Near-Eye Augmented Reality Display

Kishore Rathinavel, *Student Member, IEEE*, Hanpeng Wang, Alex Blate, Henry Fuchs, *Life Fellow, IEEE*



Fig. 1. View through our volumetric near-eye display where virtual objects are placed among real objects at a range of distances. All virtual pixels are displayed at the proper depth simultaneously, allowing the viewer to accommodate freely anywhere in the scene. *Extreme left:* Overhead depiction of scene geometry. Icons to the left of the optical axis correspond to virtual objects (RGB bunnies), while icons to the right of the axis represent real objects (duck, cube, poster). *Remaining images, left to right:* Photos taken through the display, with camera focus set to the three distances marked on the optical axis in cm and diopters, advancing from near to far.

Abstract— We introduce an optical design and a rendering pipeline for a full-color volumetric near-eye display which simultaneously presents imagery with near-accurate per-pixel focus across an extended volume ranging from 15cm (6.7 diopters) to 4M (0.25 diopters), allowing the viewer to accommodate freely across this entire depth range. This is achieved using a focus-tunable lens that continuously sweeps a sequence of 280 synchronized binary images from a high-speed, Digital Micromirror Device (DMD) projector and a high-speed, high dynamic range (HDR) light source that illuminates the DMD images with a distinct color and brightness at each binary frame. Our rendering pipeline converts 3-D scene information into a 2-D surface of color voxels, which are decomposed into 280 binary images in a voxel-oriented manner, such that 280 distinct depth positions for full-color voxels can be displayed.

Index Terms—Near-Eye Displays, Augmented Reality, Rendering Pipeline

1 INTRODUCTION

Near-eye displays that seamlessly integrate virtual content into the real world offer exciting possibilities. Real-virtual integration could induce a paradigm shift in multiple aspects of our lives, including education, communication, entertainment, and others. Near-eye displays, as compared to spatially augmented reality and 3-D displays, allow true immersion in the sense that the near-eye display user could truly experience a virtual world around them in all directions while preserving the user's natural experience and view of the real world. However, several challenges must be addressed to realize truly immersive see-through near-eye displays. One of these is the mismatch between the vergence and accommodation cues of depth perception. Vergence is the orienting of our eyes such as to center the image of a fixated object on the fovea. Accommodation is the eye lenses' ability to change their focal length to bring the object of fixation into proper focus on the retina. These are cross-coupled physiological effects. Their absence, mismatch, or incorrect representation (may also apply to other depth cues) can disrupt the sense of presence or immersion and may cause visual discomfort, eyestrain, and nausea [7].

Some of the proposed solutions to the problem of providing such depth cues attempt to approximate focus cues. Varifocal displays,

monovision displays, and even some implementations of multifocal displays are in this category. Some other proposed solutions, such as light field displays and holographic displays, provide accurate focus cues but have limitations. Current implementations of light field displays have poor resolution or are diffraction-limited. Current implementations of holographic displays are compute-intensive and may have very small eyeboxes. Phase-only spatial light modulator (SLM) technologies also need improvement before holographic displays based on these technologies can become practical.

This paper explores a new class of displays: volumetric near-eye displays. Our approach is to sweep the virtual image plane back and forth over a wide range of diopters and use high-speed DMDs coupled with high-speed illumination to present a large number of multiple thin slices of a computer generated volume. While this sounds similar to multifocal displays that show images at various fixed depths, there is a crucial difference:

Traditionally, for multifocal displays, the computer-generated volume is decomposed into a series of image planes placed at different depths; for time-multiplexed multifocal displays, this necessitates that the focus-tunable lens or deformable mirror settle down in each focus state. Our approach is to oscillate the focus-tunable lens in a continuous state and display a stack of binary images at high-speed such that the displayed stack of images is perceived as slices of a continuous full-color volume. We decompose the computer-generated volume locally, on a per-voxel basis, and distribute the decomposition around the location of the voxel. Thus, our rendering algorithm is aware of and leverages the fact that the focus-tunable lens is in continuous motion—rather than assuming a lens that moves and settles in discrete steps. Low-level hardware access to a high-speed DMD and a high-speed HDR RGB LED light source allows control of display pattern and illumination for each binary frame. We present a rendering pipeline for volumetric

• Kishore Rathinavel is with UNC Chapel Hill E-mail: kishore@cs.unc.edu.
• Hanpeng Wang is with UNC Chapel Hill E-mail: wph17@cs.unc.edu.
• Alex Blate is with UNC Chapel Hill E-mail: blate@cs.unc.edu.
• Henry Fuchs is with UNC Chapel Hill E-mail: fuchs@cs.unc.edu.

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

near-eye displays that utilize such hardware.

One might be concerned about the computational complexity of our approach. In our implementation, we make some simplifying assumptions to reduce computational overhead. However, these simplifications might not be desirable in a human-wearable product; without these assumptions, our approach would be moderately computationally demanding. While this might be an encumbrance for today’s embedded hardware, we assert that near-eye displays (NEDs) of the future must have substantially more compute power to perform, e.g., low-latency corrections, head and eye tracking, real-world scene understanding, and so on. For example, onboard GPUs are already found in NEDs such as Microsoft HoloLens. While our current implementation is offline, we believe that future NEDs will have sufficient onboard computational resources to perform the required computations in real-time on the device.

1.1 Contributions

This paper’s main contributions are:

1. A volumetric NED exhibiting 280 perceptually simultaneous binary depth planes, each an arbitrary RGB color, situated between 15cm (6.7 diopters) and 4M (0.25 diopters) from the viewer.
2. A rendering pipeline for the new NED that decomposes 3-D graphics primitives efficiently into the set of single-color binary depth planes, such that 24 bits-per-pixel color voxels can be displayed at 280 unique depth positions.

1.2 Benefits

In addition to supporting the current volumetric display implementation, our proposed system can emulate varifocal displays and previous multifocal displays. This could allow the system to become a test-bed for future perceptual studies on accommodation. Our display allows low-level access to many stages of the graphics pipeline between GPU and the actual emission of light rays that form a retinal image. This low-level access could be used to study alternative rendering pipelines for future near-eye displays and advanced projectors. Integration of our present work and previous work with similar hardware [20, 21] could lead to a near-eye display with several desirable properties (low-latency, high dynamic range, accommodation-capable).

2 RELATED WORK

2.1 Volumetric Displays

Volumetric displays create multiple real or virtual light sources in a three-dimensional volume of space and can typically be seen from a wide range of angles around the display. These light sources are the 3-D analog of pixels and are called *voxels*. Earlier designs of volumetric displays were table-top designs and the displayed volume was confined to the *physical volume of the display* [3, 6, 33, 35, 40, 41]. One of the limitations of most of these displays is that the light sources are presented additively and view-dependent effects, such as occlusion, are absent. This limitation is overcome in Cossairt et al. [3] and in Jones et al. [14] by using anisotropic diffusers.

Our proposed display provides a methodology to create virtual light sources over an *extended volume external* to the display’s physical volume. Applied to near-eye displays, this methodology has the potential to solve the vergence-accommodation conflict and reduces the need to track accommodation state in future eye-tracking technology. To clarify, our display needs eye-tracking in the sense that the *pupil position* must be tracked, but the *accommodation state* of the pupils need not be tracked.

2.2 Accommodation supporting NEDs

2.2.1 Multifocal near-eye displays

Multifocal near-eye displays, first proposed by Akeley et al. [2], display a small number of images at different depths; the images are perceived additively [2, 10, 25–27]. In Akeley, et al. [2] and MacKenzie, et al. [27], subregions of an LCD panel were mapped to different focal planes using beam splitters. Liu and Hua [23], Love et al. [26], and

Liu et al. [25] propose a switchable lens to multiplex between the multiple focal planes. Hu and Hua [8–10] propose to use high-speed optical components, such as a DMD and a 1KHz deformable membrane mirror, to achieve a larger number of focal planes (six) than previously demonstrated.

Because a relatively small number of depth planes are used to represent objects occupying a large volume, multifocal plane displays need scene decomposition algorithms to optimally represent a 3-D scene using a few 2-D image planes. Content generated by these scene decomposition algorithms provide synthetic focus cues to represent objects that lie in between the focal planes. MacKenzie et al. [27] propose a per-pixel linear blending approach. Narain et al. [32] propose an optimized blending algorithm that can demonstrate occlusion, reflection, and non-Lambertian effects. Mercier et al. [31] and Lee et al. [18] propose a new scene decomposition techniques that are tolerant to eye movements. While scene decomposition algorithms help to depict imagery that lie between the focal planes, the spatial frequency of the fused image is inversely related to the focal plane separation [8, 11].

Similar to multifocal displays, our display can also be thought of as a view-dependent and depth-fused multifocal display. Our display has about two orders of magnitude more focal planes than previous multifocal displays which approaches a *volumetric display*’s performance. Like previous multifocal displays, our display also requires eye-tracking to provide correct occlusion and dis-occlusion effects. In this paper, we assume that the pupil position is known. Like previous multifocal displays, we also share the problem of generating synthetic focus cues through scene decomposition to represent a large 3-D scene with 2-D image planes. However, while previous methods perform the scene decomposition in an image-oriented manner, we perform the scene decomposition in a voxel-oriented manner. This is discussed in detail in Section 5.

Matsuda et al. [30] propose a multifocal display whose focal surfaces can acquire non-planar, scene-dependent surface geometry. Matsuda et al. [30] propose a rendering pipeline that converts a 3-D scene to multiple piecewise smooth 2-D surface representations that are displayed in a time-multiplex manner. In comparison with their work, our rendering pipeline generates a single 2-D surface representation of the 3-D scene, and our display does not require piecewise smooth 2-D surfaces. Our display also exhibits more uniform image quality throughout the displayed volume.

Recently, Lee et al. [19, 37] propose a multifocal plane display which uses synchronized DMD, LCD panel, and focus-tunable lens. With the exception of their LCD panel and our HDR LEDs, the hardware and operation seem similar to our display. But, because of their use of LCD panel and our use of HDR LEDs, the rendering pipelines of the two displays are different. In their display, during the focus-tunable lens’ cycle, the DMD panel is used to illuminate portions of the LCD panel resulting in color sub-images at various depths. In our display, during the focus-tunable lens’ cycle, the HDR LEDs and DMD create a series of single-color binary images which integrate together such that a color volume is perceived.

2.2.2 Light field near-eye displays

Light field displays synthesize the individual light rays that recreate the 3-D scene and can conceptually provide accurate focus cues and monocular occlusion. However, current implementations of light field displays are diffraction-limited [13, 29] or have poor resolution due to a spatial-angular resolution trade-off [12, 17]. While light field displays present a virtual pixel by displaying the light rays originating from the virtual pixel individually, our volumetric NED displays the entire set of light rays that originate from the virtual pixel simultaneously.

2.2.3 Holographic near-eye displays

Holographic displays precisely modulate the wave function of the image arriving at the pupil using a digital hologram displayed on a phase-only spatial light modulator (SLM) such as a phase-only liquid crystal on silicon (LCoS) panel. Conceptually, these displays can also provide accurate focus cues, monocular occlusion, vision correction, and non-Lambertian effects. Current implementations of holographic near-

eye displays have a very small eyebox [28], and are computationally expensive [28, 30, 38]. Our NED also has a small eyebox (4mm) and is moderately computationally intensive. Our NED's eyebox can be larger; the limiting factor for our eyebox is the focus-tunable lens's aperture (1cm). Maimone et al. [28] demonstrate a NED that can provide per-pixel focus cues for a range of 10cm to 32.5cm. In comparison, our NED provides per-pixel near-accurate focus cues for a large depth range (15cm to 4M).

2.2.4 Varifocal near-eye displays

Varifocal near-eye displays have a single image plane where the vergence and focus cues match, and this plane is moved by using focus-tunable lenses [15, 22, 34], or deformable membrane mirrors [5], or by actuating fixed-focus optical components [1]. In a varifocal display, all pixels are at the same focal plane - so virtual pixels that do not lie on the plane of focus need to be synthetically blurred in proportion to their distance from the plane of focus. Varifocal displays need to track the accommodation state of the pupil [34] or assume that the pupils are accommodated to the eye convergence distance [1, 5].

2.2.5 Accommodation-invariant displays

In Konrad et al. [16], a focus-tunable lens is set to oscillate, and the same image is displayed at several depths by using a fast switching backlight. The image is calculated to provide synthetic focus cues by emulating an average point-spread-function of the eye focused at several depths. In comparison, our NED can update the binary pattern and illumination value for each depth plane and provides near-accurate and real focus cues.

2.3 Rendering pipeline for DMD-based NEDs

Previous NEDs have used DMDs and proposed different rendering pipelines [8, 10, 20, 21]. We build upon their hardware but propose a new rendering pipeline. A detailed discussion is provided in Section 5.1.

3 SYSTEM OVERVIEW

Figure 2 shows an overview of our NED's hardware and operation. Our proposed display consists of three main active optical components namely: (1) an HDR Illuminator, (2) a DMD chip, and (3) a focus-tunable lens. These three optical components are driven at high-speed by an FPGA, a microcontroller, and custom electronics.

The focus-tunable lens is driven in a continuous mode such that its optical power follows a triangular or sinusoidal waveform. The DMD projector is synchronized with the focus-tunable lens to display a stack of binary frames in each lens cycle, and the HDR illuminator illuminates the DMD chip with a distinct selected RGB color for each binary frame. Each cycle of the focus-tunable lens is one frame of the overall display. To avoid confusion, each frame of the DMD will be referred to as *single-color binary image* whereas the 24-bit color rendering of the 3-D scene will be referred to as the *color image*.

Our DMD's refresh rate is $f_{\text{DMD}} = 16,800 \text{ Hz}$, and our target display refresh rate is $f_{\text{NED}} = 60 \text{ Hz}$. Then the number of single-color binary images displayed by the DMD in each frametime of the NED is given by

$$N_b = \frac{f_{\text{DMD}}}{f_{\text{NED}}} = 280. \quad (1)$$

These 280 single-color binary images are distributed in depth along the user's line-of-sight from 15cm to 4M. Correct modeling of the depth distribution and field-of-view (FoV) of binary images is necessary for proper rendering and color decomposition.

The optical design of our NED is discussed in Section 4 and the rendering pipeline that converts 3-D scene information into multiple single-color binary images is discussed in Section 5.

4 OPTICAL DESIGN

This section models the optical design and timing characteristics of our near-eye volumetric display to arrive at the geometry of the displayed volume, i.e., depth distribution and FoV of the binary images. The geometry of the volume is used in the rendering pipeline to decompose a 3-D scene to a volume that is displayed by the NED.

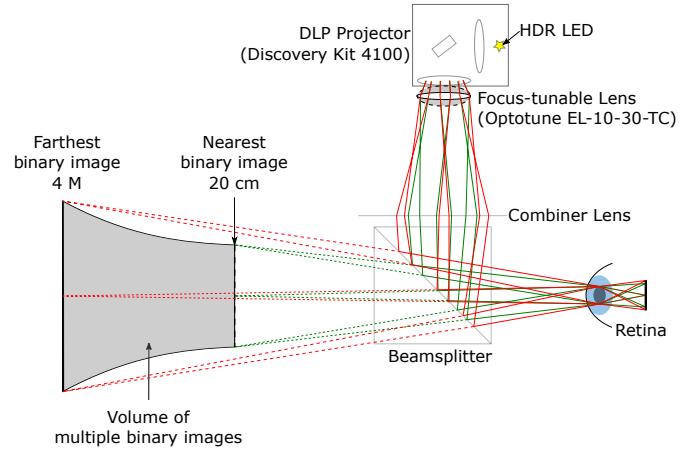


Fig. 2. Figure shows an overview of the hardware and operation of our NED. The NED is composed of a high-speed HDR LED, high-speed projector, focus-tunable lens, and other common optical components. The NED's optics, rendering pipeline, and the synchronized operation of its active components (HDR LEDs, DMD, focus-tunable lens) work together to present a color volume spanning 15cm (6.7 diopters) to 4M (0.25 diopters).

4.1 Overview of optical design

Our optical system is composed of multiple lenses (see Figure 3). The left diagram of Figure 3 shows the image formation process for any projector. Such a projector can be converted to a near-eye display by placing an eyepiece or combiner lens just after the projected image. Since the projected image for most off-the-shelf projectors would be too large, a converging lens could be placed between the projector and the combiner lens; this helps in reducing the magnification of the projected image and in reducing the form-factor of the NED. In our NED, instead of placing a static converging lens between the projector and the combiner lens, we place a focus-tunable lens (see right diagram of Figure 3) and configure the optical power of the focus-tunable lens to sweep the real image of the DMD close to the combiner lens. To see a virtual image, the combiner lens's focal length has to be less than the distance between the lens and the real image (i.e., $f_3 < o_3(t)$).

4.2 Modeling of optical design to derive volume geometry

We begin with stating the Gaussian thin-lens equation

$$\frac{1}{f} = \frac{1}{o} + \frac{1}{i}, \quad (2)$$

and associated equations

$$i = \frac{fo}{o-f}, \quad M = \frac{I}{O} = -\frac{i}{o}, \quad (3)$$

where f denotes focal lens of a thin lens, o denotes object distance, i denotes image distance, O denotes the object size, I denotes image size, and M denotes the magnification of the lens.

Due to the presence of multiple lenses in the optical stack (see Figure 3), we analyze the image formation of each lens separately and consider the image formed by each lens as the object for the next lens. This gives the following geometric relations:

$$o_2 = i_1 - d_1, \quad o_3(t) = d_2 - i_2(t), \quad i_e(t) = i_3(t) + d_e. \quad (4)$$

The relationship between the distance from the DMD to the projection lens (o_1), focal length of the projection lens f_1 , and the projected image distance (i_1) is given by

$$i_1 = \frac{f_1 o_1}{o_1 - f_1}. \quad (5)$$

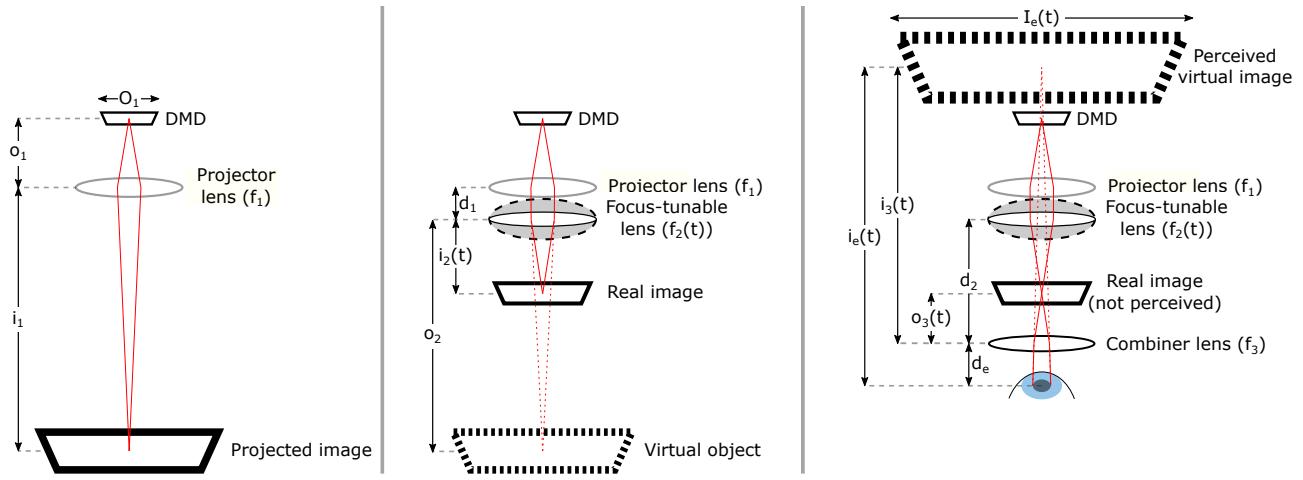


Fig. 3. Our NED's optics can be analyzed in three stages. Figure shows the unfolded optics and ray diagram for each stage. *Left:* Image formation for the DMD projector using manufacturer-provided projection optics. *Middle:* Adding a focus-tunable lens at the exit pupil of the DMD projector causes the real image of the DMD to be formed closer; Configuring the focus-tunable lens power to continuously oscillate causes the real image of the DMD to also oscillate. *Right:* A combiner lens finally creates a virtual image of the DMD that can be seen by the eye.

The relationship between the object distance $o_2 = i_1 - d_1$, focal length ($f_2(t)$), and image distance ($i_2(t)$) for the focus-tunable lens is given by

$$i_2(t) = \frac{f_2(t)o_2}{o_2 - f_2(t)} = \frac{f_2(t)(i_1 - d_1)}{(i_1 - d_1) - f_2(t)}. \quad (6)$$

The optical power of the focus-tunable lens ($f_2(t)$) can be configured to maintain a constant value or follow a time-varying square, triangular, or sinusoidal waveform. Other waveforms may be possible with custom electronics, but for this paper, we analyze only the triangular and sinusoidal waveforms of lens power.

To define optical power of the focus-tunable lens as a function of time, we define some standard parameters for time-varying signals: *DC bias*, *amplitude*, and *half-time period*. Let the DC bias, which is the average value of the signal over one full-time period be denoted by D . Let the amplitude, which is half of the peak-to-peak value, be denoted by A . Note that each cycle of the lens is the frametime of the NED. Hence, the frequency of the lens is equal to the refresh rate of the NED (f_{NED}). Let a denote half a time period ($a = \frac{1}{2f_{\text{NED}}}$). The optical power of the focus-tunable lens, when following a triangular waveform, can be modeled as

$$f_2(t) = D - A \left(\frac{1}{2} - \left| \frac{t-a}{a} \right| \right), \quad (7)$$

and when following a sinusoidal waveform, it can be modeled as

$$f_2(t) = D + A \sin(2\pi f_{\text{NED}} t). \quad (8)$$

And finally, the relationship between the object distance ($o_3(t)$), focal length (f_3), and image distance ($i_3(t)$) for the combiner lens is given by

$$i_3(t) = \frac{f_3 o_3(t)}{o_3(t) - f_3} = \frac{f_3(i_2(t) + d_2)}{(i_2(t) + d_2) - f_3}. \quad (9)$$

The above equations (4) to (9) are sufficient to calculate the depth of each binary image plane. The FoV of the virtual binary images is found by repeated application of the magnification formula from Equation (2):

$$M = M_1 M_2 M_3, \quad (10)$$

$$I_e = M O_1, \quad (11)$$

$$\theta_{\text{FoV}} = 2 \tan^{-1} \left(\frac{I_e}{2i_e} \right). \quad (12)$$

In our system, these are the values for the known quantities: $f_{\text{DMD}} = 16,800$ Hz, $f_{\text{NED}} = 60$ Hz, $f_1 = 2.96$ cm, $o_1 = 3$ cm, $O_1 = 1.778$ cm (diagonal size of the DMD module), $d_1 = 3$ cm, $f_3 = 6$ cm, $d_2 = 12$ cm, $d_e = 3$ cm, $D = 14$, $A = 4$, and $a = 8.33$ ms. Equations (1) to (12) are evaluated with the above values to calculate the geometry of the displayed volume. The geometry of the volume is graphed in Figure 4.

The above formulation and graphs in Figure 4 shows only 140 unique depth planes over the time period because depth values in the first half of the time period are repeated in the second half of the time period. Our implementation is slightly different from this - we apply a small phase difference to equations (7) to (8) to get 280 unique depth values.

4.3 Sinusoidal vs. triangular waveforms

In optical imaging systems, including the human eye, the blur of an object that is defocused is directly proportional to the difference between the actual distance of focus and the distance of the object in units of diopter. In our NED, the depth distribution of images should ideally be dioptrically equidistant from each other. From Figure 4, it can be seen that the lens power following a triangular waveform results in a near-linear and equidistant distribution of virtual image planes in dioptric space. Hence, we implemented the rendering pipeline and electronic synchronization assuming that the lens sweeps a triangular waveform. However, when we used the sinusoidal waveform in place of a triangular waveform, keeping everything else such as the color decomposition, and electronic synchronization the same, we didn't notice a significant difference in the displayed volume geometry and image quality. This may be either because the difference between the triangular and sinusoidal waveforms is negligible compared to the minimum dioptric difference required to make a perceptual difference or because the lens' triangular and sinusoidal waveforms are similar, which can often happen with physical systems due to inertia/friction, etc., especially at higher frequencies.

5 RENDERING PIPELINE

In this section, we first discuss the rendering pipelines of previous DMD-based NEDs, then describe our full rendering pipeline from graphics primitives to single-color binary images, and finally discuss the benefits and limitations of our rendering pipeline.

5.1 Rendering pipeline for previous DMD-based NEDs

5.1.1 Low latency and HDR NEDs

Most display technologies that employ a DMD also use a constant intensity or bivalent illumination source and use pulse train modulation to create grayscale or color imagery [21]. Recently, Lincoln et al. [20]

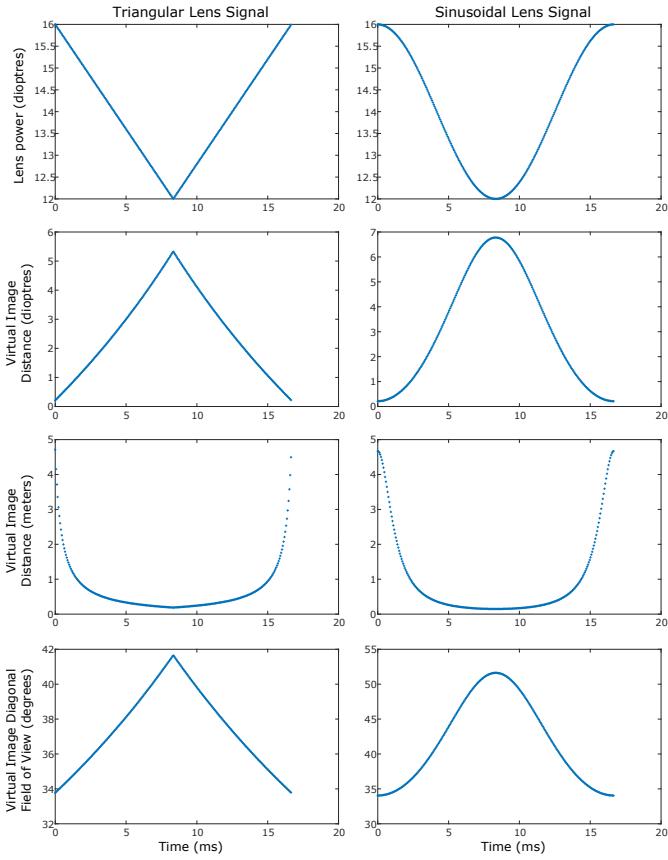


Fig. 4. Graphs modeling the depth distribution and FoV of the displayed single-color binary images that compose the volume formed by synchronizing the DMD projector and a continuously oscillating focus-tunable lens. The oscillating lens's optical power can follow a triangular waveform (Left column) or a sinusoidal waveform (Right column). Data presented in these graphs are used in the rendering pipeline to convert 3-D scene information to multiple single-color binary images that are displayed by the NED. Equations used to generate these graphs are described in Section 4.

demonstrated a DMD-based display system which used a controllable high-speed HDR illuminator. They demonstrated that the intensity and color of the illumination could be changed over a wide range on a per-binary frame basis. They also proposed a new color to binary decomposition method, which they call *Direct Digital Synthesis (DDS)*. Let d be the desired color intensity value, g be the generated color intensity value, and s be the step index of the binary representation of the value of d . Then, DDS decomposition from color to binary values can be represented per color channel as shown below:

$$g = \sum_{s=0}^{n-1} (2^s \times \text{bit}(d, s)). \quad (13)$$

5.1.2 Multifocal plane NEDs

Previous DMD-based multi-focal plane displays [8, 10] decomposed a 3-D scene to a stack of *color images* fixed at the various depths. In these approaches, the focus-tunable lens or deformable membrane mirrors would step through a set of focal lengths, and at each focal length, *after the lens stabilizes*, a series of binary images was displayed by one of the classical pulse train modulation schemes to generate color imagery. For such color image plane based approaches, we provide equations below for the relationship between the DMD's framerate (f_{DMD}), number of focal planes (N_{planes}), framerate of the NED (f_{NED}), and the color depth per color channel (N_{gray}). Up to some extent, a multifocal NED with a larger number of N_{planes} can present better imagery because

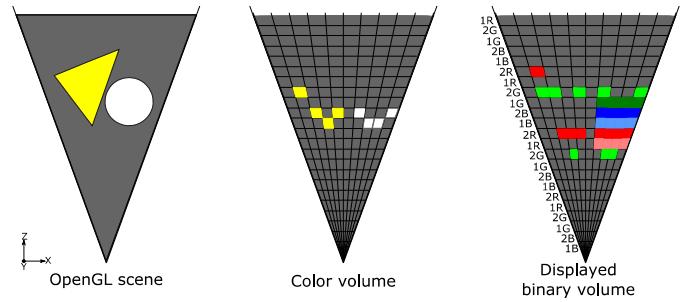


Fig. 5. Diagram shows the stages of our rendering pipeline: voxelization (see Section 5.3) and binary decomposition (see Section 5.4). For ease of representation, the figure depicts the rendering pipeline for a simple 2-D graphics and 6 bits-per-pixel imagery. Actual implementation uses 3-D graphics and 24 bits-per-pixel imagery. The numbers along the displayed binary volume's frustum indicate the intensity level and color of the RGB LED that illuminates the current binary image.

the scene decomposition algorithms of depth fused multifocal displays trade-off the spatial frequency of the fused image and the focal plane separation [8, 11].

In case of classical pulse train modulation schemes, the number of focal planes is given by

$$N_{\text{planes}} = \frac{f_{\text{DMD}}}{3f_{\text{NED}}(2^{N_{\text{gray}}} - 1)}. \quad (14)$$

Using DDS decomposition, N_{planes} can be increased significantly as shown by the following equation

$$N_{\text{planes}} = \frac{f_{\text{DMD}}}{3f_{\text{NED}}N_{\text{gray}}}. \quad (15)$$

A DMD-based multifocal display which decomposes 3-D scene information to color image planes which are in turn decomposed from color images to binary images based on DDS decomposition has not been demonstrated. If it were demonstrated with our hardware ($f_{\text{DMD}} = 16,800$ Hz, $N_{\text{gray}} = 8$, $f_{\text{NED}} = 60$ Hz), we would achieve $N_{\text{planes}} = 11$ color image planes. However, we propose a further improvement below based on voxel-oriented decomposition rather than image-oriented decomposition.

5.2 An overview of our rendering pipeline

The pipeline currently handles only opaque polygons; transparency and other primitives are left to future work. Our rendering pipeline is composed of two steps: (1) *voxelization*, i.e., the process of converting 3-D polygonal data to a 2-D surface composed of color voxels (3-D equivalent of pixels) that best approximates 3-D polygonal data; and (2) *decomposition* of the color voxels into a series of binary images and corresponding illumination values; these data are used by the display to present a series of single-color binary images to the viewer.

5.3 Voxelization: Graphics primitives to 2-D surface

Using 3-D models and scene data, an OpenGL renderer generates an RGB image and a linearized 16-bit depth map of the current scene at the resolution of the DMD display (1024×768). The 16-bit values of the depth map are remapped to the 280 depth values of the focal planes supported by our optical design. This results in a 2-D surface, composed of color voxels, in a $1024 \times 768 \times 280$ volume.

5.4 Binary Decomposition: Color voxels to binary images

Our key observation is that the binary representation of a color voxel need not start or end at one of the modulo $3 \times (2^{N_{\text{gray}}} - 1)$ planes as proposed by earlier binary multifocal displays. It need not start or end at one of the modulo $3 \times N_{\text{gray}}$ also, as would be the case for a multifocal NED which displays color image planes using DDS decomposition.

Instead, the decomposition of a color voxel to binary voxel can begin and end at arbitrary depths.

When converting from color volume data to binary images, the intensity and color of each color voxel tell us the binary pattern that represents it, and the depth of the color voxels tells us the center around which the binary pattern should be distributed. The binary voxels that encode the color voxel are distributed along the perspective projection lines that pass through the color voxel’s location and the distribution is centered around the color voxel’s location. Figure 5 provides a visualization of our rendering pipeline. For ease of representation, Figure 5 depicts the rendering pipeline for equidistant focal planes and for 2-D graphics generating 6 bits-per-pixel imagery. Our implementation handles 24-bits-per-pixel color imagery.

If this decomposition was implemented in an acyclic manner, the number of unique color voxel depths would be $N_b - (3 \times N_{\text{gray}}) + 1$, which is 257 planes in our case. However, we could implement this decomposition in a cyclic manner, and in this case, the number of unique color voxel depths would be equal to N_b , which is 280 in our case.

Even though we depict in Figure 5 that the decomposition happens in a perspectively shaped volume, it can be implemented as a decomposition on a rectangularly shaped volume. This is indeed the case in our implementation. This is not an issue because when the NED displays the single-color binary images, it does a near-inverse perspective transformation.

5.5 Display: Binary images to Retinal image

The binary images generated are displayed on a DMD in sync with a focus-tunable lens sweeping a sinusoidal or triangular waveform for the optical power of the lens. The single-color binary images displayed by the NED are integrated by the eye to see a color volume. Displaying the binary images in our prototype display is a near inverse-perspective transformation. It is not a perfect inverse-perspective transformation due to the slight change in FoV of images seen over the cycle of the lens (see Figure 4). This near inverse-perspective transformation allows us to perform the transformation of RGB and depth images to color voxels, and the transformation of color voxels to binary voxels in an orthographic space.

5.6 Limitations

5.6.1 Depth and Spatial resolution

Conceptually, the minimum non-zero separation in depth between two voxels in our display is 1 depth plane which averages to $\frac{6.7 \text{ diopters}}{280 \text{ focal planes}} = 0.024 \text{ diopters}$. However, because the binary voxels are spread across multiple binary image planes, we should expect to see a blur for the color voxel along the optical axis which could lead to a loss in the depth resolution of the NED. Since each color voxel is represented by multiple binary voxels and the brighter binary voxels are going to be perceived more strongly, we calculate the depth blur as the *weighted* standard deviation of sorted depth values for a moving window of length $3 \times N_{\text{gray}} = 24$; this is graphed in the top row of Figure 6.

Similarly, due to the slightly changing FoV of binary images across the lens cycle (shown in Figure 4), we should expect to see a blur perpendicular to the optical axis which could lead to a loss in the spatial resolution of the NED. This blur is minimum for pixels close to the optical axis and maximum for pixels at the periphery. The maximum angular blur perpendicular to the optical axis is calculated as the standard deviation of FoV values for a moving window of length $3 \times N_{\text{gray}} = 24$; this is graphed in the bottom row of Figure 6.

The blur perpendicular to the optical axis can be reduced by performing a calibration to determine the actual FoV of each binary image plane and modifying the color to binary decomposition algorithm to take into account the deformed volume geometry. We did not perform such a calibration process for this paper. The blur along the optical axis, however, is more fundamental to the display technology. It could be reduced by advanced color volume to binary volume decomposition schemes.

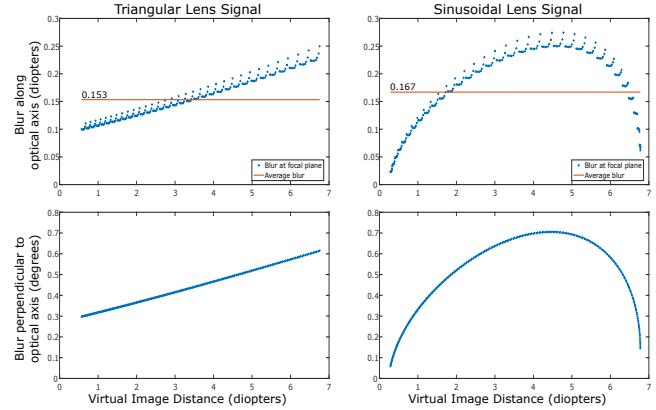


Fig. 6. *Top row:* Graphs indicate the depth blur for a color voxel at each depth plane and the average depth blur for color voxels of all depth planes. The depth blur arises because the rendering pipeline decomposes each color voxel to multiple single-color binary voxels which are spread along the perspective projection lines. *Bottom row:* The optics of our NED cause the FoV of the virtual image to slightly change over the lens cycle; this changing FoV is graphed. This creates a blur perpendicular to the optical axis leading to a loss in spatial resolution.

Previous works have suggested slightly different values for the focal plane separation required for a good multifocal display. Rolland et al. [36] suggest 0.143 diopters, Akeley et al. [2] design their prototype with image spacings of 0.67 diopters, Liu et al. [24] and Watt et al. [39] suggest 0.6 diopters, and MacKenzie et al. [27] suggest 1 diopter. As shown in Figure 6, our display has a maximum depth-blur of 0.3 diopters and an average depth-blur of 0.167 diopters.

5.6.2 Voxel-fighting in a dynamic display implementation

Here we discuss a minor limitation in extending our proposed offline rendering pipeline to a dynamic display. Observe that to decompose a single color voxel for a 24 bits-per-pixel image, we require 24 binary voxels. In the case of a static display and a cyclic implementation of our decomposition algorithm, this means that a color voxel at, say, the 280th focal plane would be decomposed into binary voxels that range from binary image indices 268 to 12. However, in a dynamic display case, we run into the issue that a new frame is received for each display cycle.

If the incoming frame information completely replaces the previous frame information, there could be a loss of brightness and bit-depth for the color voxels in the last few focal planes. Alternatively, if we design the NED to start displaying the new frame information only after it finishes displaying the previous frame information, the DMD display’s cycle would quickly fall out-of-sync with the lens cycle. With a modified rendering pipeline, for which the framerate of the NED is slightly lower than the frequency of the focus-tunable lens, and very good synchronization of the lens and the DMD, this would not be an issue. Alternatively, we could carry over the information of the last few focal planes of the previous frame to the new frame while giving priority/preference to the new frame’s information.

6 IMPLEMENTATION

6.1 Overview and Software

To test our ideas, we developed a hardware prototype of a monocular near-eye display and implemented an offline version of our proposed rendering pipeline. The offline rendering pipeline begins with the rendering of a virtual scene using OpenGL/GLSL to generate an RGB image and a linearized 16-bit depth map of the virtual scene. The RGB image and depth map are processed in MATLAB to generate a series of binary images and RGB LED brightness values. The binary images are uploaded to and displayed by the DMD controller, and the RGB LED brightness values are used by a custom RGB LED driver for precise high-speed control over each LED’s brightness. An ARM-based

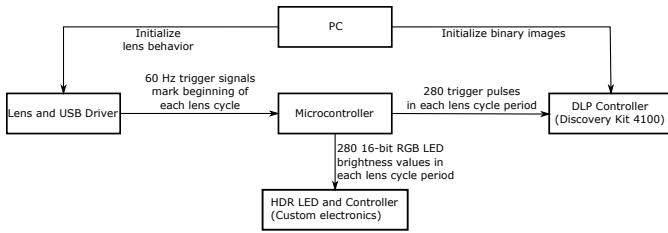


Fig. 7. Diagram shows the various hardware components and their timing relations to each other in the display's operational state.

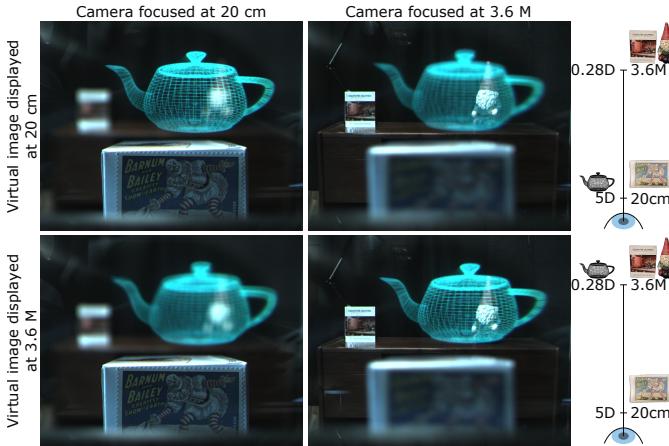


Fig. 8. View through our near-eye display when only one out of the 280 binary image planes is encoded with a binary image. This figure gives an idea of how each binary image is perceived by an eye or a camera. When all binary images are encoded with appropriate content, a time-integrated color volume occupying a large depth range can be seen (see Figure 10).

microcontroller provides synchronization between the lens, the DMD controller, and the custom RGB LED driver. Below we discuss each hardware component in detail.

6.2 Hardware

Focus-tunable Lens The focus-tunable lens used is the Optotune EL-10-30-TC-VIS. The optical power of the lens is controlled via a manufacturer-provided software and a USB-connected lens driver. The optical power can be set to a static value, or it can be set to follow a rectangular, sinusoidal, or triangular signal for a wide range of frequencies (0.25 Hz to 2000 Hz). For this paper, all experiments were conducted with the optical power of the lens configured to follow a triangular signal of 60 Hz frequency, and the maximum and minimum lens powers of the triangular signal were approximately $50m^{-1}$ and $120m^{-1}$.

Optics Other than the focus-tunable lens, we use the manufacturer-provided optical engine of the TI Discovery 4100 Kit (STAR-07 optical module), a Fresnel lens (60mm focal length), and a beamsplitter that allows the display to optically integrate the real world view and the imagery of the virtual scene.

DMD controller The DMD controller we use is the Texas Instruments (TI) Digital Light Processing (DLP) Discovery 4100 Kit which drives an XGA (1024 x 768) DMD module. The display system is capable of displaying binary images at up to 17241 Hz which would allow 287 binary images to be displayed in each lens cycle. This would need precise synchronization between the lens signal and the DMD controller, which is not afforded by the current implementation. For a more robust system, we display 280 images in each lens cycle and design the system such that the 280 images are guaranteed to be displayed slightly before the beginning of the next lens cycle.

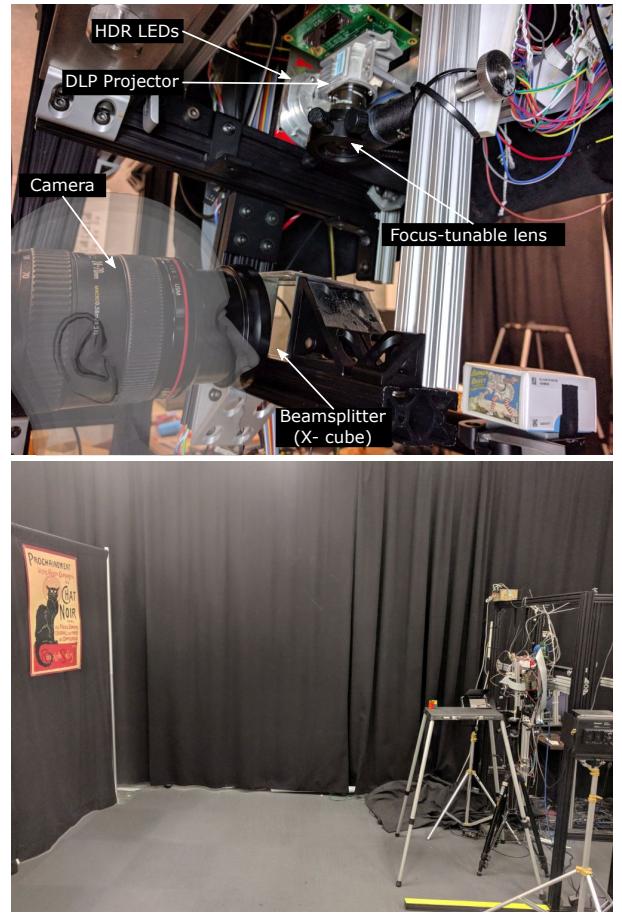


Fig. 9. *Top:* Our prototype display. *Bottom:* The staged real-world scene used to collect all see-through images and videos. Multiple objects (a tiny rubber ducky (2cm height), a wristwatch, a Rubik's cube, and a wall poster) are arranged progressively from near to far. Virtual objects are rendered in this staged real-world scene such that each virtual object is located at the same depth as one of the real world objects (See Figure 10).

Custom RGB LED Illuminator A PCB mounted RGB LED is controlled using electronics consisting of Digital-Analog-Converters (DACs), Op-Amps, etc. The board listens for three 16-bit binary codes over Serial Peripheral Interface (SPI) protocol over three parallel buses and sets each color LED to the brightness level corresponding to the received code. The board is capable of illuminating the DMD with a wide range of brightnesses and color combinations. 2^{16} levels of intensity are possible for each color LED and all color LEDs can be driven in parallel. The full-scale rise and fall times of each channel are approximately 500ns; every binary frame can be illuminated at a distinct intensity and color mix. This RGB LED illuminator is the same as the one used in Lincoln et al. [20]; please refer to that paper for more details.

PC A PC using Intel Xeon E5-2630 2.4 GHz processor with an NVIDIA GeForce GTX 980 running Windows 7 is used to implement an offline version of the proposed rendering pipeline.

6.3 Operational detail

Figure 7 gives an overview of the operation of the NED. The binary images are uploaded to the DMD controller using the ALP 4.1 Controller Suite. The DMD controller is configured to advance frames each time it receives a trigger signal from the microcontroller. At the end of the sequence of images, the DMD cycles back to the first image. The frametime of the DMD was set to the minimum possible frametime of $58 \mu s$.

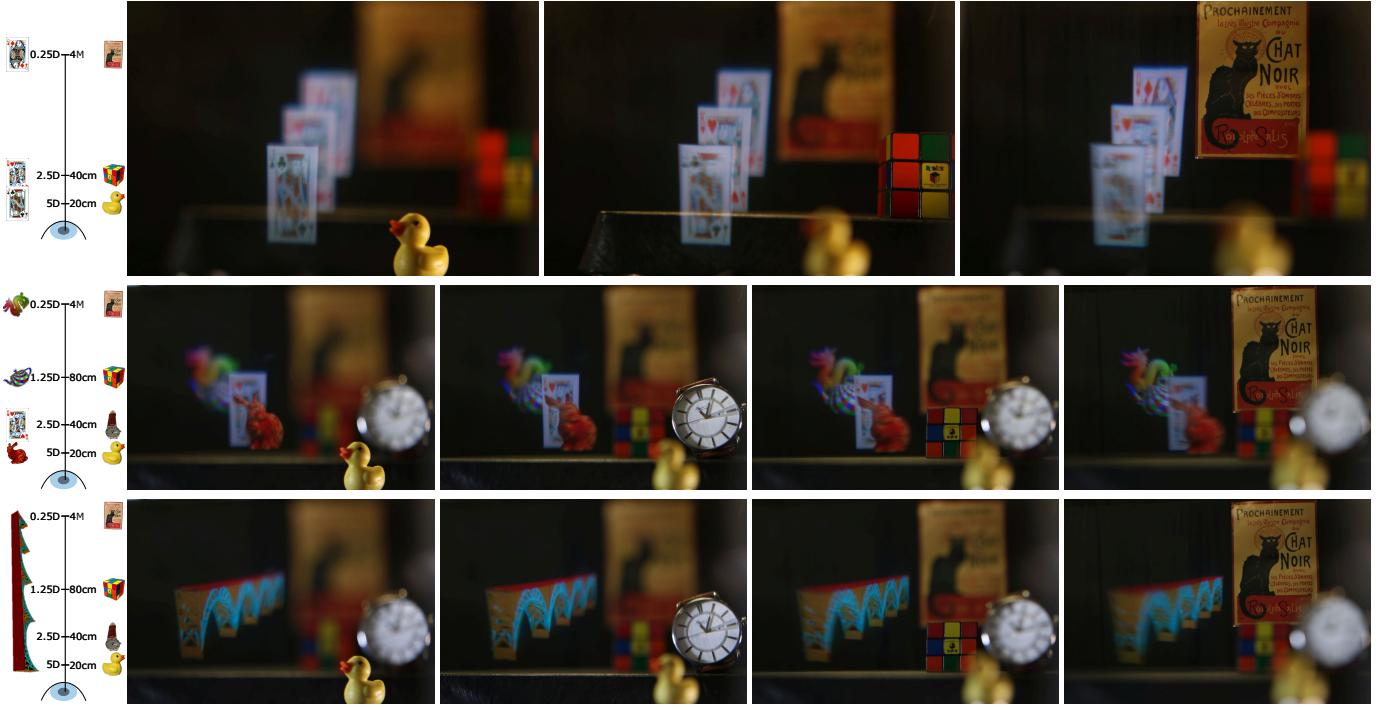


Fig. 10. View through our volumetric near-eye display where virtual objects are placed among real objects at a range of distances. *Extreme left:* Overhead depiction of scene geometry. Icons to the left of the optical axis correspond to virtual objects, while icons to the right of the optical axis correspond to real objects. *Other images, left to right, in each row:* Photos taken through the display where the focus of the camera is adjusted progressively from near to far. In each row, the only difference between the see-through views is the camera’s focus settings - this demonstrates the ability of the display to provide proper focus cues for all virtual pixels simultaneously, allowing the viewer to freely accommodate in the scene without any feedback to the display.

The lens controller outputs a trigger signal whose rising edges correspond to the beginning of each lens’s cycle. The lens operates at a frequency of 60 Hz. The lens’ trigger signal is detected by the microcontroller, which then performs 280 instances of these operations before the next lens’ trigger signal: (1) microcontroller outputs a trigger signal to the DMD controller, and (2) microcontroller sends three 16-bit words to the LED controller. Each 16-bit word specifies the brightness of a color LED. The microcontroller ensures that the DMD updates and illumination values are phase-locked to the lens cycle.

Figure 9 shows an image of our display’s hardware and the experimental setup.

6.3.1 Calibrating phase delay

In our experience, we’ve found that there is a phase delay between the lens signal and the displayed image plane depth estimated in Figure 4. This phase delay was calibrated visually by generating a synthetic stack of images in which each image has a single feature (like a cross-hair), but the feature is placed at a different location in each image. By setting the camera lens to nearest focus, it was visually determined that the 180th image out of the 280 image stack is in focus, which meant that the lens trigger signal and our system’s display of the virtual images are out of phase by $\frac{180-140}{280} \times 16.67ms = 2.38ms$. To correct for this, the binary images uploaded to the DMD controller were cyclically rearranged such that the 180th binary image is moved to the 140th index.

6.4 Future implementation improvements

The results can be visually improved by performing white-balance correction, gamma curve calibration, and calibrating for the non-uniform frustum as shown in the last row of Figure 4. In our experience, we didn’t find the change in FoV to reduce the image quality significantly, but it does make long straight objects slightly curved, especially when the straight objects are places towards the periphery of the display.

7 RESULTS

Cameras To record images and videos of the see-through view of the display, cameras that approximate the human eye were placed behind the display (see Figure 9) at a distance that approximated the eye relief of a human viewer (2cm away from the beamsplitter). See-through image results presented in this paper were recorded using a Canon T6i Rebel camera with a Canon 24-70mm f/2.8 lens. See-through video results (in accompanying video submission) were recorded using a Point Grey Chameleon3 camera with a Fujinon 2.8-8mm f/1.2 lens. A 4mm aperture was used in both cameras to emulate the human pupil diameter while collecting results. When using PointGrey cameras, the nearest distance was chosen to be 15cm (6.7 diopters), and when using the DSLR camera, the minimum distance was 20cm (5 diopters) because the lens could not focus closer.

Our display is capable of presenting virtual imagery closer than 15cm and farther than 4M, but we were constrained by the recording camera (for the minimum distance) and by the lab space (for the farthest distance). Although virtual images closer or farther than what is demonstrated here may not be required for near-eye displays, this may be of use in some other application.

Setup Figure 9 shows the monocular display prototype, the positioning of the camera in place of a viewer’s eye and the staged real-world scene consisting of a large poster, and smaller objects such as a Rubik’s cube, a wristwatch and a tiny rubber ducky (2cm height). The real world objects are arranged from small to large progressively away from the display approximately along the line of sight of the see-through view. Virtual objects are scaled progressively from small to large away from the display for the virtual objects to subtend approximately the same angle at the camera.

See-through images Figure 10 shows the see-through views of our NED when displaying a virtual scene registered to the staged real-world scene. Even though our RGB LED illuminator can produce high-dynamic range and consequently very bright virtual imagery [20],



Fig. 11. Images show the see-through view through the display when the optical power of the focus-tunable lens follows a triangular waveform and a sinusoidal waveform. For both these images, the voxelization and color volume to binary volume decomposition was performed assuming a triangular waveform. We don't observe a significant difference in the see-through images.

we're currently displaying moderately bright imagery at 24 bits-per-pixel. A black background screen is used to improve visibility and contrast of the virtual objects.

As discussed earlier, the field of view of the virtual image changes during the lens cycle. Ideally, this needs to be calibrated for, but we didn't. Other first order and second order optical aberrations present in all optical systems may also need to be calibrated. Each of these optical aberrations likely varies with depth across the volume. Optical aberrations are observable in our system, e.g., in the second row of Figure 10, even when focused at the correct depth, the bottom portion of the Jack card is blurred relative to the top.

To demonstrate that our display shows high-resolution imagery when each binary image is considered individually, Figure 8 shows the see-through view when only one of the 280 binary images was encoded with the image of a wire model of a teapot. We plan to develop a methodology for the calibration of the volume as future work - our expectation is this would lead to high-resolution imagery across the entire volume.

Figure 11 shows a comparison between the see-through views when the lens signal follows a triangular and a sinusoidal waveform. This is discussed in Section 4.3.

Video results An accompanying file with this paper shows videos of the see-through view. The video results show a larger range of depth (15cm - 4M) compared to the image results (20cm - 4M) because the camera used to record the video was able to focus closer. In these videos, a flicker is seen propagating back and forth through the displayed volume. This flicker is an artifact of the video capture and is not human-visible. The flicker arises because of the slight discrepancy between the display frame time (16.67 ms) and the minimum shutter speed possible on the camera (16.74 ms). The flicker moves back and forth in the volume because the camera samples the whole volume once and a small portion of the volume twice - and because of this, it starts to sample the volume in the subsequent frame from a slightly different starting position of the volume.

8 DISCUSSION

8.1 Limitations

Here we discuss the limitations of our proposed technique and its implementation. We review methods for overcoming these limitations.

8.1.1 High computation

Our proposed rendering pipeline, under our current assumptions of (1) opaque 3-D models (2) uniform FoV throughout the lens cycle, allow us to implement the entire algorithm (from graphics primitives to binary images) very efficiently. If the above assumptions were removed, the decomposition would become complex. However, we started off with the assumption that future near-eye displays will most likely have an onboard GPU and consequently, the techniques presented here are closely related to some of the graphics algorithms already implemented very efficiently in GPUs.

8.1.2 Non-zero pixel blur along optical axis

We presented two sources for blur that could reduce the depth resolution and the spatial resolution of the displayed imagery. Of these, the spatial resolution can be improved by a calibration process that takes into account the changing FoV over the lens cycle. The source of the depth blur is more fundamental to the approach. It arises from the effort to represent a color voxel in a depth-fused manner with multiple binary voxels. This could be reduced in the future with advanced color to binary volume decomposition algorithms.

8.1.3 Implementation Limitations

Static display Our current implementation relies on an offline rendering pipeline implementation. We believe that our implementation can be made into a dynamic display in a system similar to that presented in recent low-latency and HDR display work [20, 21].

Bulky Optics The bulk of the optics is due to the large optical engine of the DLP Discovery 4100 kit, and the tiny aperture of the focus-tunable lens. Other DMD development boards have much smaller optics, and we also note that there is a commercially available AR display that uses a DMD chip [4]. The small aperture of the focus-tunable lens constrains the optical design and limits the etendue of the system. There are focus-tunable lenses with a wider aperture which could be used. Our NED, if implemented with alternative components, could approach moderate form factor.

Bulky electronic components All of the driving electronics (DLP Discovery 4100 kit, custom RGB LED controller, microcontroller) could be reimplemented in a compact ASICs device.

8.2 Future work

Our near-eye display can emulate some other display technologies, such as multifocal and varifocal displays, and is thus suitable as a versatile platform for user studies. The current work could benefit greatly from a compact, wearable, wide-FoV, binocular, and real-time implementation. Since the hardware platform and application are similar, this work could be integrated with recent low-latency [21], and HDR AR [20] displays work. This would require a holistic approach to a near-eye display rendering pipeline. Another opportunity for research is to investigate if this display can be made entirely independent of eye-tracking requirements.

9 CONCLUSION

We have introduced a near-eye volumetric display capable of presenting a large volume over an extended depth-of-field created external to the display's physical volume. We view our system as a hybrid between traditional volumetric displays that create the volume within the confines of the display's physical volume, and view-dependent multifocal near-eye displays. We presented the optical design of our implementation and the rendering pipeline that synthesizes the volume for our display. Our main contribution is a new color-volume-to-binary-volume decomposition algorithm that does not require image planes, but rather decomposes each color voxel into a series of binary voxels around the color voxel's position. With our optical setup and rendering pipeline, we demonstrate a full-color volumetric display refreshed at 60 Hz and comprising 280 focal planes, each at a unique depth, ranging from 15cm (6.7 diopters) to 4M (0.25 diopters). We hope that this system will inspire future research work in near-eye displays to rethink the rendering pipeline.

ACKNOWLEDGMENTS

The authors wish to thank Peter Lincoln and Andrei State for helpful discussions, and Jim Mahaney for engineering support.

This research was supported in part by NSF grant 1405847, by a research gift from Intel Corporation, and by the BeingTogether Center, a collaboration between Nanyang Technological University (NTU) Singapore and University of North Carolina (UNC) at Chapel Hill. The BeingTogether Center is supported by the National Research Foundation, Prime Minister's Office, Singapore, under its International Research Centers in Singapore Funding Initiative.

REFERENCES

- [1] K. Aksit, W. Lopes, J. Kim, P. Shirley, and D. Luebke. Near-eye Varifocal Augmented Reality Display Using See-through Screens. *ACM Trans. Graph.*, 36(6):189:1–189:13, Nov. 2017. doi: 10.1145/3130800.3130892
- [2] K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks. A Stereo Display Prototype with Multiple Focal Distances. In *ACM SIGGRAPH 2004 Papers*, SIGGRAPH ’04, pp. 804–813. ACM, New York, NY, USA, 2004. doi: 10.1145/1186562.1015804
- [3] O. S. Cossairt, J. Napoli, S. L. Hill, R. K. Dorval, and G. E. Favalora. Occlusion-capable multiview volumetric three-dimensional display. *Appl. Opt.*, 46(8):1244–1250, Mar 2007. doi: 10.1364/AO.46.001244
- [4] D. S. Dewald, A. T. Evans, N. Welch, A. Gross, and G. Hill. The Avegant Glyph: Optical Design Considerations and Approach to Near-eye Display. *SID Symposium Digest of Technical Papers*, 47(1):69–71, 2016. doi: 10.1002/sdtp.10609
- [5] D. Dunn, C. Tippets, K. Torell, P. Kellnhofer, K. Akit, P. Didyk, K. Myszkowski, D. Luebke, and H. Fuchs. Wide Field Of View Varifocal Near-Eye Display Using See-Through Deformable Membrane Mirrors. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1322–1331, April 2017. doi: 10.1109/TVCG.2017.2657058
- [6] G. E. Favalora, J. Napoli, D. M. Hall, R. K. Dorval, M. Giovinco, M. J. Richmond, and W. S. Chun. 100-million-voxel volumetric display. In *Cockpit Displays IX: Displays for Defense Applications*, vol. 4712, pp. 300–313. International Society for Optics and Photonics, 2002. doi: 10.1117/12.480930
- [7] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks. Vergence accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3):33, 2008. doi: 10.1167/8.3.33
- [8] X. Hu and H. Hua. Design and Assessment of a Depth-Fused Multi-Focal-Plane Display Prototype. *Journal of Display Technology*, 10(4):308–316, April 2014. doi: 10.1109/JDT.2014.2300752
- [9] X. Hu and H. Hua. High-resolution optical see-through multi-focal-plane head-mounted display using freeform optics. *Opt. Express*, 22(11):13896–13903, Jun 2014. doi: 10.1364/OE.22.013896
- [10] X. Hu and H. Hua. Design and tolerance of a free-form optical system for an optical see-through multi-focal-plane display. *Appl. Opt.*, 54(33):9990–9999, Nov 2015. doi: 10.1364/AO.54.009990
- [11] H. Hua. Enabling Focus Cues in Head-Mounted Displays. *Proceedings of the IEEE*, 105(5):805–824, May 2017. doi: 10.1109/IPROC.2017.2648796
- [12] H. Hua and B. Javidi. A 3D integral imaging optical see-through head-mounted display. *Opt. Express*, 22(11):13484–13491, Jun 2014. doi: 10.1364/OE.22.013484
- [13] F.-C. Huang, K. Chen, and G. Wetzstein. The Light Field Stereoscope: Immersive Computer Graphics via Factored Near-eye Light Field Displays with Focus Cues. *ACM Trans. Graph.*, 34(4):60:1–60:12, July 2015. doi: 10.1145/2766922
- [14] A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec. Rendering for an Interactive 360° Light Field Display. In *ACM SIGGRAPH 2007 Papers*, SIGGRAPH ’07. ACM, New York, NY, USA, 2007. doi: 10.1145/1275808.1276427
- [15] R. Konrad, E. A. Cooper, and G. Wetzstein. Novel optical configurations for virtual reality: Evaluating user preference and performance with focus-tunable and monovision near-eye displays. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI ’16, pp. 1211–1220. ACM, New York, NY, USA, 2016. doi: 10.1145/2858036.2858140
- [16] R. Konrad, N. Padmanaban, K. Molner, E. A. Cooper, and G. Wetzstein. Accommodation-invariant computational near-eye displays. *ACM Trans. Graph.*, 36(4):88:1–88:12, July 2017. doi: 10.1145/3072959.3073594
- [17] D. Lanman and D. Luebke. Near-eye Light Field Displays. *ACM Trans. Graph.*, 32(6):220:1–220:10, Nov. 2013. doi: 10.1145/2508363.2508366
- [18] S. Lee, J. Cho, B. Lee, Y. Jo, C. Jang, D. Kim, and B. Lee. Foveated Retinal Optimization for See-Through Near-Eye Multi-Layer Displays. *IEEE Access*, 6:2170–2180, 2018. doi: 10.1109/ACCESS.2017.2782219
- [19] S. Lee, Y. Jo, D. Yoo, J. Cho, D. Lee, and B. Lee. Tomoreal: Tomographic displays. *CoRR*, abs/1804.04619, 2018.
- [20] P. Lincoln, A. Blate, M. Singh, A. State, M. C. Whitton, T. Whitted, and H. Fuchs. 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*, I3D ’17, pp. 15:1–15:7. ACM, New York, NY, USA, 2017. doi: 10.1145/3023368.3023379
- [21] P. Lincoln, A. Blate, M. Singh, T. Whitted, A. State, A. Lastra, and H. Fuchs. From Motion to Photons in 80 Microseconds: Towards Minimal Latency for Virtual and Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics*, 22(4):1367–1376, April 2016. doi: 10.1109/TVCG.2016.2518038
- [22] S. Liu, D. Cheng, and H. Hua. An optical see-through head mounted display with addressable focal planes. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 33–42, Sept 2008. doi: 10.1109/ISMAR.2008.4637321
- [23] S. Liu and H. Hua. Time-multiplexed dual-focal plane head-mounted display with a liquid lens. *Optics letters*, 34(11):1642–1644, 2009.
- [24] S. Liu and H. Hua. A systematic method for designing depth-fused multi-focal plane three-dimensional displays. *Opt. Express*, 18(11):11562–11573, May 2010. doi: 10.1364/OE.18.011562
- [25] S. Liu, H. Hua, and D. Cheng. A Novel Prototype for an Optical See-Through Head-Mounted Display with Addressable Focus Cues. *IEEE Transactions on Visualization and Computer Graphics*, 16(3):381–393, May 2010. doi: 10.1109/TVCG.2009.95
- [26] G. D. Love, D. M. Hoffman, P. J. Hands, J. Gao, A. K. Kirby, and M. S. Banks. High-speed switchable lens enables the development of a volumetric stereoscopic display. *Opt. Express*, 17(18):15716–15725, Aug 2009. doi: 10.1364/OE.17.015716
- [27] K. J. MacKenzie, D. M. Hoffman, and S. J. Watt. Accommodation to multiplefocalplane displays: Implications for improving stereoscopic displays and for accommodation control. *Journal of Vision*, 10(8):22, 2010. doi: 10.1167/10.8.22
- [28] A. Maimone, A. Georgiou, and J. S. Kollin. Holographic Near-eye Displays for Virtual and Augmented Reality. *ACM Trans. Graph.*, 36(4):85:1–85:16, July 2017. doi: 10.1145/3072959.3073624
- [29] A. Maimone, D. Lanman, K. Rathinavel, K. Keller, D. Luebke, and H. Fuchs. Pinlight Displays: Wide Field of View Augmented Reality Eyeglasses Using Defocused Point Light Sources. In *ACM SIGGRAPH 2014 Emerging Technologies*, SIGGRAPH ’14, pp. 20:1–20:1. ACM, New York, NY, USA, 2014. doi: 10.1145/2614066.2614080
- [30] N. Matsuda, A. Fix, and D. Lanman. Focal Surface Displays. *ACM Trans. Graph.*, 36(4):86:1–86:14, July 2017. doi: 10.1145/3072959.3073590
- [31] O. Mercier, Y. Sulai, K. Mackenzie, M. Zannoli, J. Hillis, D. Nowrouzezahrai, and D. Lanman. Fast Gaze-contingent Optimal Decompositions for Multifocal Displays. *ACM Trans. Graph.*, 36(6):237:1–237:15, Nov. 2017. doi: 10.1145/3130800.3130846
- [32] R. Narain, R. A. Albert, A. Bulbul, G. J. Ward, M. S. Banks, and J. F. O’Brien. Optimal Presentation of Imagery with Focus Cues on Multi-plane Displays. *ACM Trans. Graph.*, 34(4):59:1–59:12, July 2015. doi: 10.1145/2766909
- [33] Y. Ochiai, K. Kumagai, T. Hoshi, J. Rekimoto, S. Hasegawa, and Y. Hayasaki. Fairy lights in femtoseconds: Aerial and volumetric graphics rendered by focused femtosecond laser combined with computational holographic fields. *ACM Trans. Graph.*, 35(2):17:1–17:14, Feb. 2016. doi: 10.1145/2850414
- [34] N. Padmanaban, R. Konrad, T. Stramer, E. A. Cooper, and G. Wetzstein. Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *Proceedings of the National Academy of Sciences*, 2017. doi: 10.1073/pnas.1617251114
- [35] H. Refai. Static Volumetric Three-Dimensional Display. *J. Display Technol.*, 5(10):391–397, Oct 2009.
- [36] J. P. Rolland, M. W. Krueger, and A. A. Goon. Dynamic focusing in head-mounted displays. In *Stereoscopic Displays and Virtual Reality Systems VI*, vol. 3639, pp. 463–471. International Society for Optics and Photonics, 1999. doi: 10.1117/12.349412
- [37] D. Y. J. C. D. L. B. L. Seungjae Lee, Youngjin Jo. Shape scanning displays: tomographic decomposition of 3d scenes, 2018. doi: 10.1117/12.2315655
- [38] L. Shi, F.-C. Huang, W. Lopes, W. Matusik, and D. Luebke. Near-eye light field holographic rendering with spherical waves for wide field of view interactive 3d computer graphics. *ACM Trans. Graph.*, 36(6):236:1–236:17, Nov. 2017. doi: 10.1145/3130800.3130832
- [39] L. R. Simon J. Watt, Kevin J. MacKenzie. Real-world stereoscopic performance in multiple-focal-plane displays: How far apart should the image planes be?, 2012. doi: 10.1117/12.908883
- [40] D. Smalley, E. Nygaard, K. Squire, J. Van Wagoner, J. Rasmussen, S. Gneiting, K. Qaderi, J. Goodsell, W. Rogers, M. Lindsey, et al. A photophoretic-trap volumetric display. *Nature*, 553(7689):486, 2018.
- [41] A. Sullivan. DepthCube solid-state 3D volumetric display, 2004. doi: 10.1117/12.527543