Volumetric Head-Mounted Display with Locally Adaptive Focal Blocks

Dongheon Yoo*, Seungjae Lee*, Youngjin Jo, Jaebum Cho, Suyeon Choi, and Byoungho Lee, *Fellow, IEEE*

Abstract—A commercial head-mounted display (HMD) for virtual reality (VR) presents three-dimensional imagery with a fixed focal distance. The VR HMD with a fixed focus can cause visual discomfort to an observer. In this work, we propose a novel design of a compact VR HMD supporting near-correct focus cues over a wide depth of field (from 18 cm to optical infinity). The proposed HMD consists of a low-resolution binary backlight, a liquid crystal display panel, and focus-tunable lenses. In the proposed system, the backlight locally illuminates the display panel that is floated by the focus-tunable lens at a specific distance. The illumination moment and the focus-tunable lens' focal power are synchronized to generate focal blocks at the desired distances. The distance of each focal block is determined by depth information of three-dimensional imagery to provide near-correct focus cues. We evaluate the focus cue fidelity of the proposed system considering the fill factor and resolution of the backlight. Finally, we verify the display performance with experimental results.

Index Terms—Virtual reality, Head-mounted display, Three-dimensional display, Multifocal display.

1 Introduction

Virtual reality (VR) has been attracting public interests because it has a variety of applications, including entertainment, trauma treatment [1], education [2], and virtual training [3]. Over the last few years, people could experience VR applications via a head-mounted display (HMD). Although the HMD is considered the most promising platform to realize VR, customers often report visual fatigue after using HMDs for a long time. For the popularization of VR, it is important to alleviate the fatigue of HMD.

Vergence-accommodation conflict (VAC) is known as one of the main causes of the discomfort in HMD [4], [5]. Commercial HMDs stimulate the convergence of binocular eyes (vergence) to make users perceive virtual objects' depth information. The depth perception depends on the degree of rotation and alignment of both eyes [6]. On the other hand, HMDs present visual information at a single focal plane where users' eyes should focus (accommodation) to perceive sharp imagery. If a virtual object depth differs from the focal plane distance, vergence-accommodation conflict (VAC) occurs. Presenting focus cues in HMDs can mitigate the VAC, and a recent study demonstrated that supporting accurate focus cues improved user comfort [6]. Furthermore, it was shown that the absence of focus cues distorted the user's depth perception [7].

A multifocal display is one of the promising solutions for HMD with focus cues. The multifocal display floats multiple focal planes or surfaces physically via spatial [8], [9], [10] or temporal [11], [12], [13], [14], [15], [16] multiplexing approaches. These approaches aim to reconstruct continuous three-dimensional scenes by stacking two-dimensional focal planes. In multifocal displays, the density of focal planes determines depth accuracy and optical resolution limit. However, the number of planes cannot be easily increased. In general, densely stacking focal planes through spatial or temporal multiplexing requires the sacrifice of a form factor or refresh rates. Although there is a computational approach to improve the depth accuracy by dynamically moving focal planes according to target scenes [17], [18], it is still challenging to reconstruct three-dimensional scenes over a wide depth range using a few focal planes.

Recently, novel approaches were proposed to provide dense focal planes over a wide depth of field using focus-tunable lenses (FTLs) and digital micromirror devices (DMDs) [11], [13], [14]. Notably, Lee et al. [13] utilized the DMD as a fast and spatially varying binary backlight. The DMD selectively turned a liquid crystal display (LCD) panel on and off while the FTL periodically swept a specific depth range. Nevertheless, it is difficult to realize this concept in a wearable form factor since the DMD system requires additional projection or relay optics.

Here, we introduce a compact VR HMD design that supports near-correct focus cues over an extended depth of field (from 18 cm to optical infinity). Our design's distinct feature comes from the focal block, which denotes a locally adjustable focal plane. In a conventional multifocal display, visual information on a multiplexed display panel is directly projected onto each focal plane, and the focal planes remain static at specific depths. Conversely, our design divides these visual data into small blocks and allows each block to

- *Dongheon Yoo, Seungjae Lee, Youngjin Jo, Jaebum Cho, and Byoungho Lee are with the School of Electrical and Computer Engineering, Seoul National University, Seoul 08826, South Korea. E-mail: dhyou93@gmail.com, seungjae1012@gmail.com, niugnas@naver.com, chojaebum@gmail.com, byoungho@snu.ac.kr.*
- *Suyeon Choi is with Stanford University. When this research was performed, he was with the School of Electrical and Computer Engineering, Seoul National University. E-mail: suyeon@stanford.edu.*
- **The first two authors contributed equally to this research.*
- *Byoungho Lee is the corresponding author.*

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

move within the depth range. We call the block a focal block. Note that the local scene geometry in the divided display window is less complicated than the global scene geometry. As focal blocks locally fit the three-dimensional scene, the proposed system shows low depth approximation errors and high retinal image quality even with few modulations of the display panel. The provision of near-correct focus cues can also alleviate VAC related issues.

Inspired by Lee et al. [13], our design comprises a low-resolution binary backlight, a liquid crystal display panel, and binocular FTLs. The backlight divides the panel into blocks of size equal to backlight's unit pixel size. While driving the FTLs to sweep system depth range, focal blocks are floated at the moment of backlight illumination. Through numerical simulations, we show how focal blocks can improve the depth accuracy and scene reconstruction quality. Furthermore, we demonstrate implementation in a wearable form factor using off-the-shelf products. We use a light-emitting diode (LED) array and liquid lenses for the backlight and FTLs to build a proof-of-concept prototype. Several practical issues due to the low fill factor of the backlight are discussed. We attempt to address these issues using optical and computational strategies. Lastly, we conduct experiments using the prototype and present results showing the feasibility of VAC mitigation via accommodation support. The major contributions of this work are as follows:

- We introduce a new type of volumetric HMD that efficiently fits volumetric scenes with locally allocated focal blocks.
- We evaluate depth approximation of the proposed display by comparing it with dynamic multifocal displays for various volumetric scenes.
- We propose a comprehensive decomposition strategy for volumetric scenes adapted to our system, including the configuration of the focal blocks, diffusion kernel optimization, and LCD image synthesis.
- We implement an HMD prototype composed of a low-resolution LED array, a high-resolution LCD panel, and several commercialized liquid lenses. The experimental results are presented and evaluated in terms of focus cue fidelity.

2 Related Work

2.1 Near-Eye Displays with Focus Cues

Near-eye displays (NEDs) with focus cues are promising solutions for VAC mitigation. Focus cues can be supported via various display methodologies, with each method using a distinct optical strategy.

Light-field NED [19], [20], [21] aims to reconstruct a four-dimensional light field that indicates intensities and directions of light rays. In the light-field NED, reconstructed retinal image varies according to the viewing position. Suppose that the exit-pupil size of the NED is similar to that of the pupil of the human eye; then, it can also support focus cues. Although these NEDs can represent view-dependent effects, such as occlusion, they suffer from trade-off relationships between the spatial and angular resolutions.

Holographic NED [22], [23] can also restore a four-dimensional light field. However, it additionally considers the diffraction and interference of light, which cannot be modeled with ray optics. Therefore, a more accurate representation of the light field is possible in the holographic NED. However, holographic NED performances, such as an exit-pupil size and field of view, are bounded by the number of pixels of the spatial light modulator. Speckle noise due to coherent light usage should be addressed as well.

Varifocal and multifocal NEDs concentrate on the direct provision of depth information rather than light field synthesis. Varifocal NED [24], [25], as the name implies, dynamically adjusts the imaging distance of a focal plane depending on the user's gaze direction. The varifocal NED is computationally efficient since it does not require complicated rendering algorithms except for depicting artificial blur. Furthermore, it can also express occlusion optically with an additional display panel [26]. However, the transition between multiple depths might be too slow, so that users may notice focal planes being shifted. Mechanical devices for shifting can also make the system bulky. Furthermore, precise gaze tracking is required to determine the distance of the focal plane correctly.

On the other hand, multifocal NED [10], [13] reconstructs multiple focal planes by spatial or temporal multiplexing. When using the temporal multiplexing method, it is nearly impossible to float the focal planes simultaneously. However, if optical power modulation and image rendering speeds exceed the flicker fusion rate, users can observe multiple planes together without noticing the switches between focal planes. Therefore, the multifocal NED can represent natural retinal blur caused by depth differences in the focal planes; a recent study [27] reported that the retinal blur significantly helped observers identify the depth order. However, computational loads are substantial when images on the focal planes are rendered using an optimal decomposition algorithm [28]. Furthermore, a form factor and frame rates should be sacrificed in the spatial and temporal multiplexing approaches.

Recently, several studies [29], [30], [31] proposed NEDs supporting focus cues in a foveated manner. The acuity of the human visual system is high in the central region of the retina and low in the peripheral region, due to the non-uniform density of photoreceptors. Based on this characteristic, the proposed NEDs concentrated on accurately reconstructing focus cues in the central area. Akşit et al. [29] introduced to use freeform optics to create a foveated focal surface. Kim et al. [30] realized a foveated NED of a "picture in picture" architecture.

Meanwhile, using a phase-only spatial light modulator, Itoh et al. [32] demonstrated that NEDs could be utilized to modify real-world views for vision assistance, such as optical zoom and focus correction. More detailed information about NEDs can be found in a published survey [33].

2.2 Multifocal Displays with Dense Focal Planes

Recently, multifocal displays generating a lot of focal planes over a wide depth of field [11], [13], [14] were proposed. In all these works, FTLs and DMDs were used to achieve the dense focal planes. While the FTLs swept the system depth range, binary images on DMDs were updated at high speeds. However, these displays represented color intensities in different ways. In the works by Chang et al. [11] and

Rathinavel et al. [14], the binary images were illuminated by high dynamic range (HDR) LED light sources. On the other hand, Lee et al. [13] used an LCD panel to express the color information of focal plane images. The DMD served as a fast and spatially adjustable backlight for the LCD panel in their work.

Our work is closely related to the work by Lee et al. in that the proposed system replaces the DMD with an LED array for compactness. Further, we propose modulating the panel a few times during a single FTL sweeping period. We demonstrate that this novel strategy can alleviate the disadvantages of using low-resolution LED arrays, such as limited depth representation for each focal block and crosstalk between adjacent focal blocks.

2.3 Multifocal Displays with Adaptive Focal Planes

With the focus-tunable optics, focal plane depths can be dynamically adjusted in multifocal displays. It was reported in several works that optimizing focal plane configurations depending on target scenes improved depth accuracy and focus cue fidelity [17], [18]. Wu et al. [17] used k-means clustering to divide the depth distribution of a volumetric scene into several sets and assigned focal planes to the center distances of the sets. Wu et al. [18] searched for the optimal configuration, changing focal plane positions exhaustively until they obtained the best retinal image quality. Although the spatial frequency and visual saliency were additionally considered compared to prior work [17], heavy computations were required for the exhaustive search.

In this work, we modify the previously proposed approaches for our system. Instead of computing the retinal image quality, we extract histograms of the visual saliency distributions so that both the depth distribution and visual saliency of a target scene can be considered. Even if our algorithm does not directly minimize the reconstruction error, it does not significantly compromise the retinal image quality. We verify this fact through quantitative evaluations of the reconstructed retinal images. Overall, our algorithm effectively decomposes a three-dimensional scene based on the visually salient regions while maintaining the overall perceptual quality.

2.4 Displays Using LED Array Backlight

Some previous works used LED array backlights for specific purposes. Huang et al. [34] employed an RGB LED array as a backlight for a monochromatic high-resolution display. In their work, a high-resolution grayscale modulator was combined with a low-resolution RGB backlight to reconstruct high-resolution full-color images. LED array backlights can be used to build high dynamic range displays as well [35]. By locally modulating the backlight brightness of a display panel, the system could have extended contrast and brightness ranges. In this study, we employ an LED array as a fast and spatially adjustable binary backlight.

2.5 Saliency Extraction

Saliency maps have considerable potential in various applications, such as image segmentation [36] and object recognition [37]. By modeling the human visual recognition procedure based on orientations, spatial frequencies, and depth distributions of scenes [38], the saliency map extraction aims to determine visually attractive regions. For the saliency map generation, various kinds of images can be utilized. A single two-dimensional RGB image [39] alone could be used to detect several features, including contrast differences and orientations. The depth map [40] and light field [38] can also be used to identify the background information accurately. In this work, we utilize the algorithm proposed by Zhang et al., where the all-in-focus RGB image, depth map, and retinal images with focal blur were used as the input dataset [40].

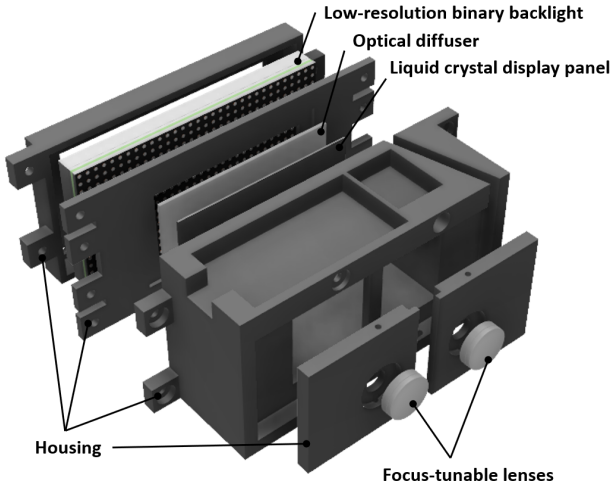

Fig. 1. Schematic of the proposed HMD consisting of a low-resolution binary backlight, an optical diffuser, a liquid crystal display panel, and focus-tunable lenses. The display panel depicts images for focal blocks, and the backlight switches the panel on and off. The diffuser spreads light from the backlight. The focus-tunable lenses serve as eyepieces of the HMD and float the focal blocks at different depths.

3 Principle

Table 1 summarizes the main abbreviations that are used to describe the proposed system in this paper. As illustrated in Figure 1, the proposed system primarily consists of a low-resolution binary backlight (LRB), a liquid crystal display panel (LDP) for color modulation, and several FTLs. We also insert an optical diffuser to alleviate screen-door effect by the LRB. The FTLs are used as eyepieces of the HMD.

To understand the working principle of our system, we examine how each optical component operates over time. We consider a simple situation in which four numbers located at different depths are optically reconstructed by our display, as described in Figure 2. In the example, the binocular FTLs are driven by analog voltage signals of $1/t_F$ kHz frequency; thus, the optical power of each FTL continuously changes according to the shape of the voltage signal. During a single sweeping period of the FTL, the frame images on the LRB and LDP are modulated $(2N_P - 1)$ and twice, respectively.

The LRB locally illuminates the LDP at a specific instant, and the color image on the LDP is floated at a distance

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3011468, IEEE Transactions on Visualization and Computer Graphics

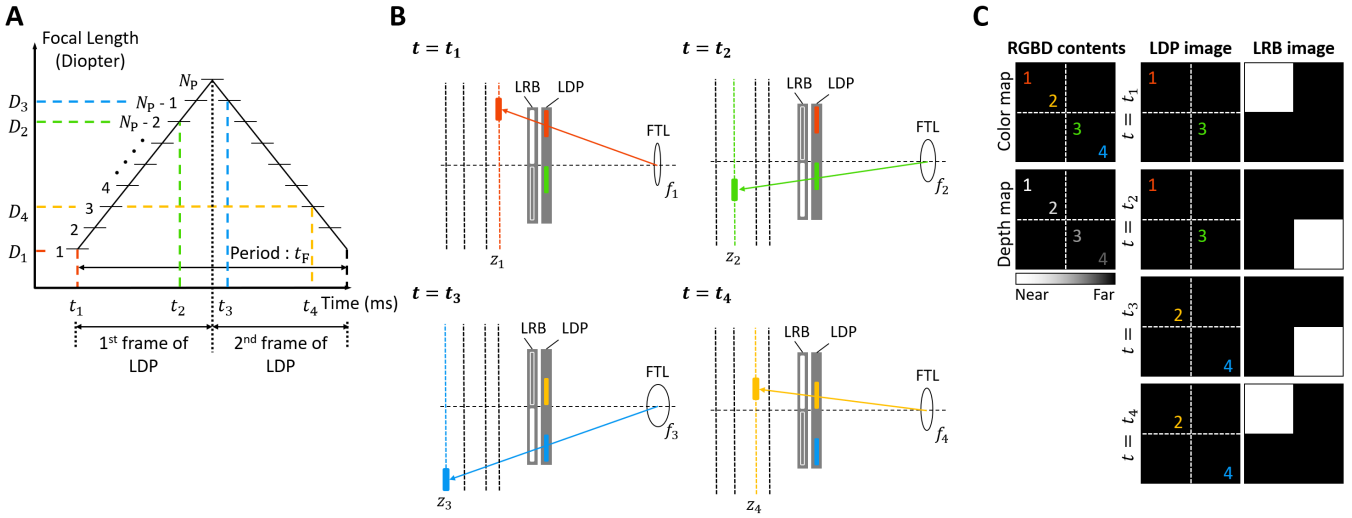

Fig. 2. Schematic of the working principle of the proposed HMD. Here, it is assumed that the LRB has 2 × 2 pixels, and two focal blocks are allocated for each partial display area. **A** In the proposed system, the focal power D_k of the FTL, which is a reciprocal value of the focal length, is modulated along the triangular waveform of the t_F period. The small horizontal lines on the focal power diagram indicate the timing when the LRB image is updated. The frame rates of the LRB are calculated to be $(1/t_F) \times (2N_P - 1)$. (**B**, **C**) While the FTL sweeps the depth range, the LRB illuminates the LDP's partial region so that the focal blocks containing partial images on the LDP can be generated at the moment of illumination. The floating distance z_k is fully determined by the focal length f_k, which is controlled according to the focal power diagram depicted in **A** and illumination timing t_k ($k = 1, 2, 3, 4$). Note that the LDP is modulated twice during a single sweeping period of the FTL, and the LDP image remains the same during each half period.

TABLE 1
Summary of abbreviations.

Abbreviation	Complete form
LRB	Low-resolution binary backlight
LDP	Liquid crystal display panel
FTL	Focus-tunable lens
DMFD	Dynamic multifocal display
PBR	Pixels-to-block ratio

determined by the FTL's optical power. The floated distance z_k is given by the thin-lens equation:

$$\frac{1}{f_k} = \frac{1}{z_T} - \frac{1}{z_k}, \tag{1}$$

where z_T and z_k denote the distances of the LDP and virtual imagery from the FTL for $k = 1, 2, 3, 4$, respectively; f_k is the focal length of the FTL. As shown in Figure 2**B**, the red number 1 and the green number 3 are floated at distances z_1 and z_2 at t_1 and t_2 when the LRB illuminates the upper left and lower right areas of the LDP image that contains both numbers. The yellow number 2 and the blue number 4 can be similarly generated at distances of z_4 and z_3 through proper illumination timing and location of the LRB.

In our system, the focusing error can be reduced as the number of focal blocks allocated for each laterally divided area or the number of LRB pixels across the field of view increases. For the first case, the system refresh rates should be sacrificed. Therefore, we focus on how the number of LRB pixels in each direction N_B affects the focusing error of the proposed design, as shown in Figure 3. If we consider a chief ray passing through a retinal image sample and the center of the eye pupil, the focusing error E can be computed as follows:

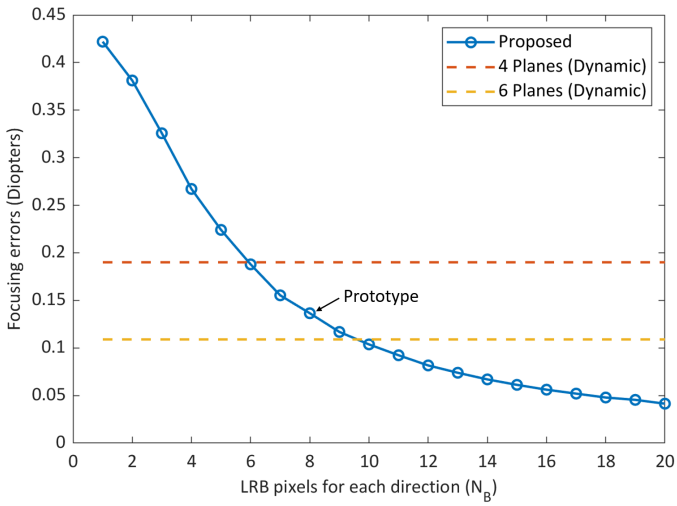

Fig. 3. Averaged focusing errors E for 130 volumetric scenes containing randomly located three-dimensional objects [41]. Our system is supposed to allocate dual focal blocks. The focusing error when the LRB has 8 pixels in each direction, which corresponds to our prototype's specification, is indicated by the black arrow in the figure. The errors of dynamic multifocal displays (DMFDs) with 4 and 6 dynamic focal planes are also plotted for comparison.

$$E(\theta_x, \theta_y) = \min_k \left| D(\theta_x, \theta_y) - \widetilde{D_k}(\theta_x, \theta_y) \right|, \tag{2}$$

where $D(\theta_x, \theta_y)$ and $\widetilde{D_k}(\theta_x, \theta_y)$ are respectively the depth (in dioptric units) of a target scene and a focal block for each viewing angle (θ_x, θ_y) indicated by the chief ray [15]. Note that the ray intersects with two focal blocks, and k is the index of the focal block closer to the target depth $D(\theta_x, \theta_y)$.

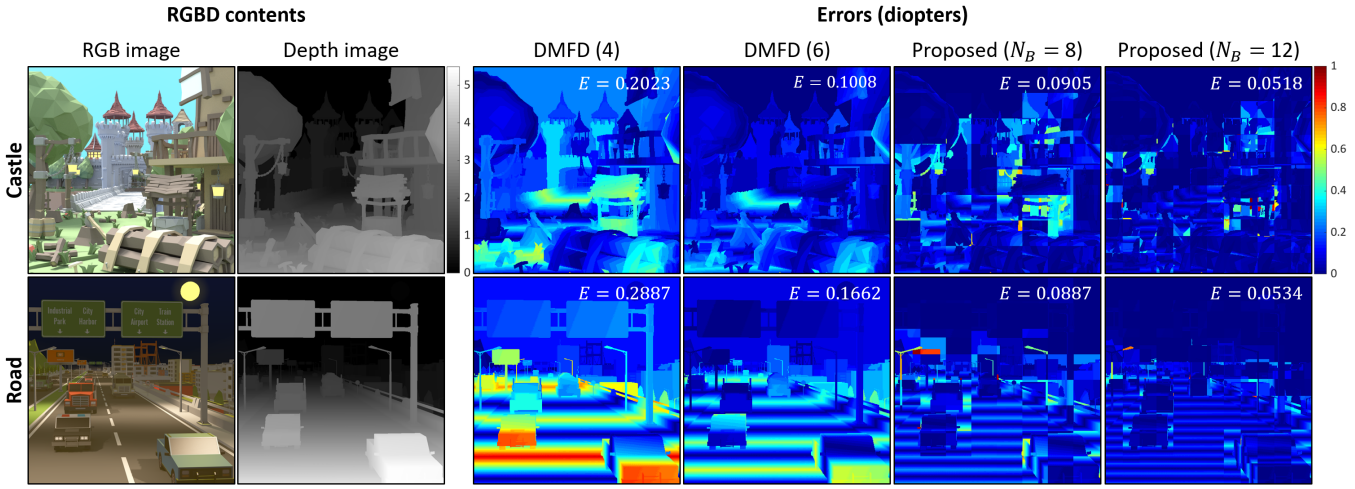

Fig. 4. Comparison of focusing errors for DMFDs of 4 and 6 planes, and proposed systems with two focal blocks, LRBs of 8 × 8 and 12 × 12 pixels. For the reconstruction target, virtual scenes (source image courtesy: www.cgtrader.com) are used that resemble real-world environments. The specifications of the scenes are summarized in Table 2. The focal planes or blocks in the displays are simply allocated using k-means clustering to concentrate on the intrinsic focusing capability, except for algorithmic influences. The averaged error value E is indicated in the upper right area of each figure.

TABLE 2
Specifications of synthesized scenes.

Scene	Field of view	Resolution	Depth range
Forest	30° × 30°	512×512	0.0 D ∼ 4.0 D
Road	30° × 30°	512×512	0.0 D ∼ 5.5 D
Castle	30° × 30°	512×512	0.0 D ∼ 4.0 D
City	10° × 10°	400×400	1.0 D ∼ 3.4 D

To evaluate our system's focus cue fidelity, we calculate the focusing errors for various volumetric scenes, as shown in Figure 3. The black arrow in the figure denotes the case that closely approximates our prototype specification. The focal blocks are supposed to be arranged according to the method proposed by Wu et al. [17]. For comparison, we also assess the fidelity for different display modes, such as dynamic multifocal displays (DMFDs) of 4 and 6 planes. The DeepFocus [41] dataset of 130 three-dimensional scenes is utilized to calculate focusing errors. In each scene, several volumetric objects are randomly placed between 0.1 and 4.0 diopter (D). According to the analysis, our prototype, which uses the LRB composed of 8×8 pixels of 3 mm size, represents the depth information of random three-dimensional scenes more accurately than the DMFD of 4 planes, but less accurately than the DMFD of 6 planes. To outperform the DMFD of 6 planes for depth approximation, the LRB should have at least 10 pixels in each direction within the field of view.

Furthermore, we investigate the focusing errors for scenes that closely resemble real-world scenarios, as shown in Figure 4. Both scenes span from 0.0 D to 5.5 D, and the mean depth errors for each display are indicated in the upper right areas of the error images. Interestingly, the proposed design of $N_B = 8$ achieves a lower mean error value E than the DMFD of 6 planes for both scenes, in contrast to the results in Figure 4. This fact is because the depth variations inside each small patch are smoother than when using random scenes. Therefore, we expect our system to be a promising solution to express natural-looking virtual scenes.

4 System Design

In this section, we introduce design methods of the proposed system for the optimal representation of retinal images. To enhance user experiences, we arrange focal blocks considering depth distributions and visual saliency of volumetric scenes. Based on the focal block arrangement, we synthesize images on the focal blocks that can minimize errors in the reconstructed retinal images.

Additionally, artifacts induced by the diffuser in the system should be considered for the synthesis. Recall that we insert an optical diffuser between the LRB and LDP to avoid the screen-door effect by the LRB. Crosstalk between the adjacent focal blocks can occur owing to the spread of light from the LRB, and such effects on image quality are not negligible. To mitigate the crosstalk, we model the diffuser behavior by designing a diffusion kernel in the algorithm. The effects of backlight diffusion on image quality are investigated by sweeping the design parameters, such as kernel width and backlight resolution. Finally, we find the kernel that can achieve optimal image quality for a backlight with a certain resolution.

4.1 Focal Block Arrangement

Up to this point, we have demonstrated that our system can reconstruct a volumetric scene using locally adjustable focal blocks. However, if the depth ranges of local scenes exceed 1 D, one or two focal blocks are not enough to ensure correct focus cues. Note that the distance of 1 D is recommended as the distance between adjacent focal planes for natural accommodation [42]. In this situation, we determine focal block arrangements by considering the visual saliency of the target scene. The saliency map indicates where an observer

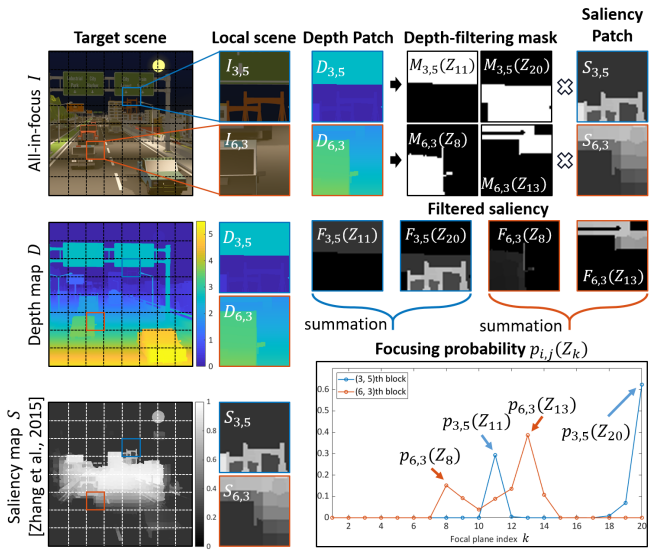

Fig. 5. Illustration of focal block arrangement. Our algorithm computes the probability $p_{i,j}(Z_k)$ that a user will focus on a specific depth Z_k in each (i,j)th block region. The depth Z_k is one of the N_P depths discretized over the entire depth range of the system. For each block region, the depth map is segmented into N_P depth bins, and each binned depth map serves as a filtering mask M for a saliency map. The saliency map S is extracted using an all-in-focus image I, depth map D, and focal stacks, as in work by Zhang et al. [40]. By summing the filtered saliency map F over the spatial dimension, the focusing probability is calculated. Among N_P depths, focal blocks are assigned to the depths with the highest focusing probability.

will be visually attracted within the scenes [38], [39], [40]. We also consider the depth distribution of the scene for correct focus cues.

Saliency map application for dynamic alignments of focal planes was introduced by Wu et al. [18]. However, as the alignments for optimal image quality are obtained using an exhaustive search, massive computational loads are required for their method. Furthermore, this technique cannot be easily adapted to our design as the focal plane images are rendered using linear depth-blending method [8]. In our system, the focal block density inside the small patch may be insufficient for the linear blending method [28]. Therefore, we introduce a new algorithm for focal block arrangements to reduce the computational loads and preserve the perceptual quality.

In Figure 5, we describe our algorithm pipeline for the arrangements with an example. The saliency map is rendered as in work by Zhang et al. [40]. Using the depth map D and the saliency map S, the probability $p_{i,j}(Z_k)$ of focusing at the depth Z_k is computed as follows:

$$p_{i,j}(Z_k) = \sum_{(x,y)\in A_{i,j}} F(x,y;z \in Z_k), \tag{3}$$

where $A_{i,j}$ indicates the (i,j)th block area; k is the index of the focal plane. The letters i and j are defined as the vertical and horizontal indexes of the LRB pixels, respectively. Assuming that the LRB is modulated N_P times during a half period of FTL sweep, the system depth range is divided into N_P bins. Among these depth bins, filtered saliency $F(Z_k)$ for the depth Z_k can be defined as follows:

$$F(Z_k) = S \circ M(Z_k), \tag{4}$$

where $\circ$ denotes the Hadamard product. The binary depth mask $M(Z_k)$ denotes the region located at kth depth bin. As our algorithm aims to find the most likely depth bins at which observers focus, focal blocks are assigned in descending order from the depth with the highest focusing probability.

TABLE 3
Quantitative retinal image quality results.

Scene	Metric	DMFD (4)	DMFD (6)	Proposed (8)	Proposed (12)
Forest	PSNR	36.42	36.79	38.60	39.86
	SSIM	0.9893	0.9899	0.9915	0.9933
Road	PSNR	36.25	38.69	38.23	40.10
	SSIM	0.9847	0.9916	0.9904	0.9934
Castle	PSNR	34.47	37.49	36.92	38.36
	SSIM	0.9794	0.9908	0.9880	0.9914
City	PSNR	31.53	32.22	35.34	36.17
	SSIM	0.9606	0.9683	0.9789	0.9825
Average	PSNR	**34.67**	**36.30**	**37.27**	**38.62**
	SSIM	**0.9785**	**0.9852**	**0.9872**	**0.9902**

4.2 Decomposition of Scene into Focal Blocks

Once the focal blocks' locations are determined, LDP images should be optimized to reconstruct accurate retinal images for various focusing states [28]. Assuming that the LRB is modulated N_P times during half a cycle of the FTL sweep, our system can form N_P focal planes composed of block-wise images. Considering the switching speed of the LED array used in our prototype, which is empirically found, we set $N_P = 20$. The retinal images are reconstructed by focusing light fields generated by the focal planes at a specific depth. Our algorithm aims to numerically model the reconstruction using matrix calculations and minimize the errors in the retinal images.

We describe retinal image reconstruction as the following form:

$$\begin{bmatrix} \tilde{\mathbf{I}}_1 \\ \tilde{\mathbf{I}}_2 \\ \vdots \\ \tilde{\mathbf{I}}_{N_R} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \cdots \\ \vdots & \ddots & \\ \mathbf{P}_{N_R 1} & & \mathbf{P}_{N_R N_P} \end{bmatrix} \begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \\ \vdots \\ \mathbf{L}_{N_P} \end{bmatrix}, \tag{5}$$

where $\tilde{\mathbf{I}}_r^{\in N_D \times 1}$ denotes a vectorized retinal image of N_D pixels given a focusing state index r. The index r varies from 1 to N_R, indicating the number of target focusing states. $\mathbf{L}_l^{\in N_D \times 1}$ is a vectorized focal plane image of N_D pixels given a layer index l. The submatrix $\mathbf{P}_{rl}^{\in N_D \times N_D}$ consists of binary values indicating whether or not a chief ray, which passes through a point on the $\mathbf{L}_l$ to N_V viewing positions, intersects with the focusing plane $\tilde{\mathbf{I}}_r$ [21]. The discrete viewing positions are sampled inside the eyebox. For simplicity, we model the human eye as a thin lens system in this simulation. However, physically accurate eye models can also be considered in our algorithm by designing the projection matrix to reflect the non-linear relationship between accommodation and position. In this case, much larger computations are required. In experiments, we choose $N_D = 470 \times 470$, $N_R = 20$, $N_V = 7 \times 7$.

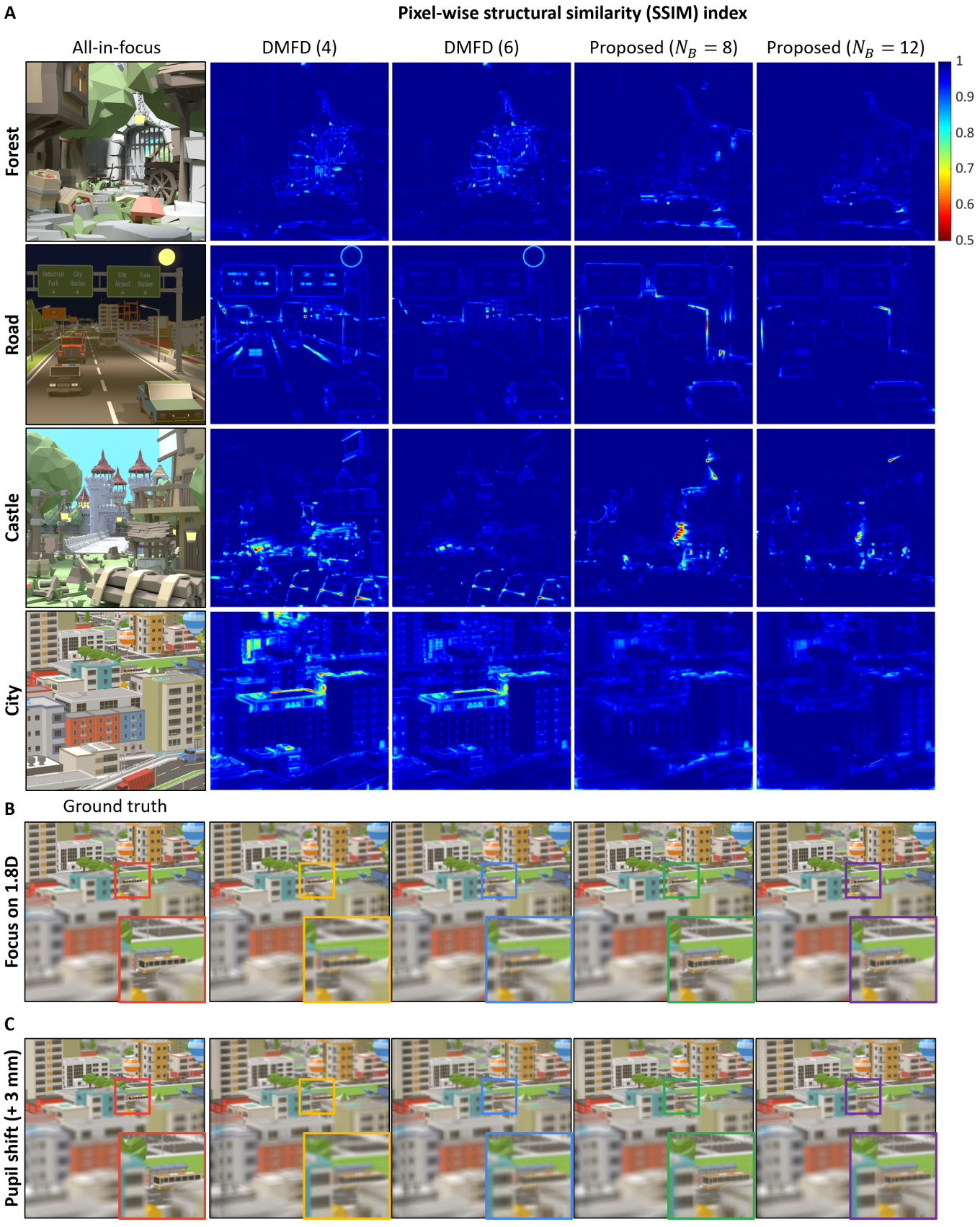

Fig. 6. Comparison of retinal image reconstruction via different display modes. We use various three-dimensional scenes for the simulations, and the scene properties are summarized in Table 2. **A** Pixel-wise structural similarity (SSIM) index for each figure is derived from a weighted sum of the SSIM images for 20 different focusing states. The weight map is estimated by the reciprocal of the circle of confusion (CoC) size. **B** Examples of the generated retinal images for the given "city" scene. The enlarged subset images are also provided for detailed comparisons. **C** Simulation results showing the effects of pupil misalignments on image quality are presented.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.3011468, IEEE Transactions on Visualization and Computer Graphics

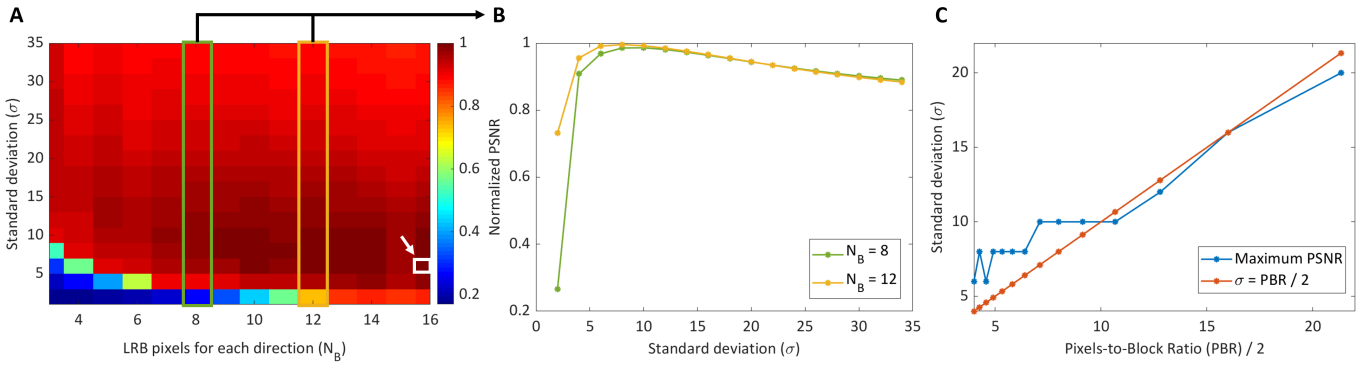

Fig. 7. Effects of diffusion kernel width and pixels-to-block ratio (PBR) on retinal image quality. **A** We compute the mean PSNR values of 20 retinal images for different values of LRB pixels (N_B) and standard deviations (σ) of the diffusion kernel. These PSNR values are also averaged over "road", "forest", and "castle" scenes, and the scene properties are summarized in Table 2. For each scene, the image resolution is set to 128×128 pixels. All values are then normalized by the maximum PSNR value denoted by the white arrow. **B** We plot normalized PSNR values against standard deviations of the diffusion kernel. **C** Standard deviations for optimal image qualities are plotted against half of the PBR values.

Since focal planes are formed due to LRB illumination through the LDP, focal plane images can be represented as shown below:

$$\begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \\ \vdots \\ \mathbf{L}_{N_P} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \cdots \\ \vdots & \ddots & \\ \mathbf{B}_{N_P 1} & & \mathbf{B}_{N_P N_L} \end{bmatrix} \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \vdots \\ \mathbf{D}_{N_L} \end{bmatrix}, \quad (6)$$

where $\mathbf{D}_d{}^{\in N_D \times 1}$ denotes a vectorized LDP image of N_D pixels for the frame index d. N_L means the number of LDP modulations during the single period of the entire system operation. The N_L also denotes the number of focal blocks for each partial region of LDP. We choose $N_L = 2$ in our prototype.

Submatrix $\mathbf{B}_{ld}{}^{\in N_D \times N_D}$ encodes the LRB behavior when $\mathbf{D}_d$ is displayed on the LDP and FTL generates the lth focal plane. Since we insert a diffuser between the LRB and LDP, the diffuser behavior, which is modeled with a Gaussian kernel [34], is also considered in the submatrix $\mathbf{B}_{ld}$ as follows:

$$B_{ld}(x, y) = G(x - x_l, y - y_l), \quad (7)$$

where $B_{ld}(x, y)$ denotes two-dimensional LRB image associated with $\mathbf{B}_{ld}$ when the LRB pixel of physical location (x_l, y_l) is turned on. Gaussian kernel $G(x, y)$ is defined as follows:

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left[-\left(\frac{x^2 + y^2}{2\sigma^2}\right)\right], \quad (8)$$

where σ means standard deviation of the kernel.

Overall, we find optimal LDP images by solving the following problem:

$$\min_{\mathbf{D}} \|\mathbf{WI} - \mathbf{WPBD}\|^2, \quad (9)$$

where $\mathbf{I}$ represents a set of ground truth retinal images for N_R focusing states. $\mathbf{W}^{\in N_D \times N_D}$ is a weight matrix. The weight matrix can be designed to redistribute errors between retinal images for different focusing states. We set the weight matrix as an identity matrix for simplicity, but finding optimal values for the weights depending on the specific application can be interesting future work. Since illumination intensity cannot be less than zero, there exists a constraint whereby all elements of $\mathbf{D}$ should be non-negative. Therefore, we solve Equation (9) via Simultaneous algebraic reconstruction technique [43] (SART), which is a gradient-based optimization method.

After allocating focal blocks and rendering LDP images, we assess the retinal blur fidelity of our design by comparing reconstructed retinal images for different types of multifocal displays. Four computer-generated scenes of different properties are used for evaluation, and scene properties are summarized in Table 2. We use the scene called "city" of a much narrower field of view to investigate the system performance when a high-resolution environment is assumed. Although the actual distance ranges where virtual objects are located differ from each other, all scenes are supposed to be extended from 0.0 D to 5.5 D. Therefore, focal blocks are not allocated between the actual range of the utilized scene, but between the system depth range.

The retinal image quality is quantitatively measured in terms of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index [44], as shown in Table 3. The averaged PSNR values over entire focal stacks for each scene are derived from the weighted sum of mean square errors. The weight image is estimated by the reciprocal of the pixel-wise circle of confusion (CoC) map c, which is defined as follows [41], [45]:

$$c = P\frac{d_r}{d_a}\left|1 - \frac{d_a}{d_s}\right|, \quad (10)$$

where P and d_r denote pupil size and distance between the retinal plane and the crystalline lens of the human eye, respectively. They are assumed as 6 mm and 25 mm. d_a and d_s mean accommodation and stimulus distances. Similarly, SSIM values in Table 3 are computed by averaging pixel-wise SSIM images as demonstrated in Figure 6. Overall, averaged metric values for all scenes are calculated for each display method.

Meanwhile, pupil misalignment may occur inside the eyebox and degrade retinal image quality. Since our decomposition algorithm samples different viewing positions and considers light rays from them, the proposed system has some tolerance for pupil movement compared to other

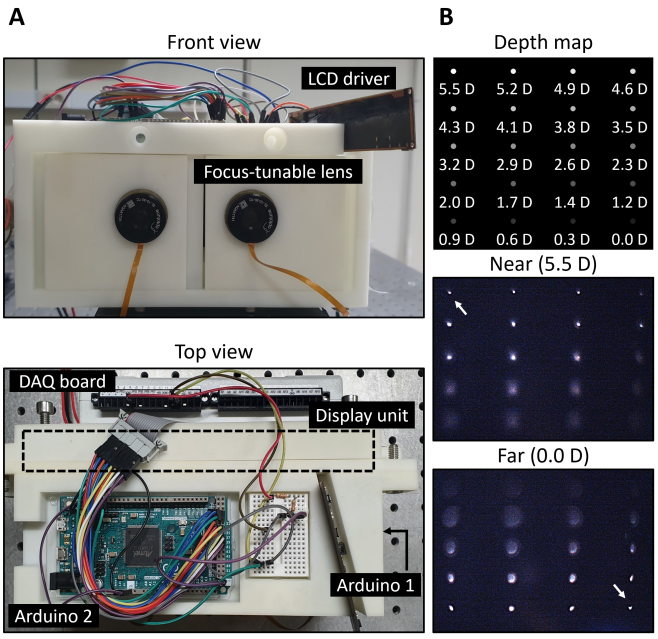

Fig. 8. Implementation of hardware prototype. **A** Photograph of the prototype. The display unit, which consists of an RGB LED array, an optical diffuser, and an LCD panel, is placed behind focus-tunable lenses. All components are packaged inside the frame. **B** Captured point-spread function (PSF) maps for different focal settings of the camera. White arrows in the photographs indicate focused points. It is demonstrated through the PSF maps that focal blocks can be freely located between 20 different distances. The distances are sampled uniformly from 0.0 D and 5.5 D (∼18 cm).

display modes, as shown in Figure 6**C**. Note that the pupil size is set as 6 mm in the simulation.

According to the results in Table 3, the proposed design with two focal blocks and LRB of 8×8 pixels shows better retinal image quality than DMFD of 4 planes for all scenes. However, DMFD of 6 planes fits "road" and "castle" scenes better than the proposed design with LRB of 8×8 pixels. Although the design can closely follow the depth information of both scenes, as shown in Figure 4, the crosstalk between focal blocks deteriorates resultant image qualities. Dividing display area more finely can further increase the retinal blur fidelity, and we numerically verify that the design using LRB of 12×12 pixels can adapt to all scenes with sufficient image qualities. Note that we use the LRB of 3 mm pixel size for our prototype so that 8×8 LRB pixels are included in the field of view. In order for 12×12 LRB pixels to be included, LRB of 2 mm pixel size should be used.

4.3 Optimizing Backlight Diffusion

Our system utilizes an optical diffuser to smooth out the screen-door effect by LRB. However, the backlight diffusion induces crosstalk between adjacent focal blocks. Since it can significantly distort the depth information of reconstructed scenes, diffusion kernel width, which is modeled by a standard deviation of Gaussian function, should be well adjusted for optimal retinal image quality.

Previously, Huang et al. [34] created a system that also exploited low-resolution LED backlight with color information and a high-resolution grayscale modulator to generate a high-resolution display. In the prior work, Huang et al. investigated the effects of two parameters on the rendered image quality: the ratios between unit pixel sizes of the backlight and the grayscale modulator, and the diffusion kernel width. The effects are evaluated through errors in reconstructed images, and the errors are computed by sweeping both parameters. However, the results cannot be identically applied to our design as the goals of both systems are different. More specifically, our design does not aim to reconstruct two-dimensional scenes but volumetric scenes.

Therefore, we analyze the effects of pixels-to-block ratio (PBR) and diffusion kernel width on retinal blur fidelity, as shown in Figure 7. The PBR is defined as the ratio of the pixel size of LRB to that of LDP. The N_B values of 8 and 12 in Figure 7**B** correspond to the half of PBR values of 8 and about 5.33, respectively. Obviously, PSNR values are observed to be significantly low when standard deviations are much less than half of the PBR values. This fact is due to the perceivable artifacts of pixelated LRB structure.

A key observation from these results is that both parameters should be similar for optimal image quality. According to the prior work [34], a standard deviation value close to the given PBR value results in moderate image quality. However, the kernel width has a significant impact on our design as the crosstalk between the focal blocks affects the point-spread functions (PSFs) of different focusing states. Considering these results, we manually set the standard deviation of the diffusing kernel to about 33 LDP pixels. This number of pixels is close to half the PBR value (∼29 pixels) in our prototype.

5 Implementation and Results

We built a compact proof-of-concept HMD prototype to demonstrate the feasibility of our design. All the components of our prototype are commercially available and packaged within a 3D-printed frame, as shown in Figure 8. In this section, we elaborate on the specifications of our prototype in terms of hardware and software.

5.1 Hardware

For binocular eyepieces, liquid lenses from Optotune (EL-10-30-TC-VIS-12D) that can change foci from 8.3 D to 20 D are used. Each lens has a 10 mm aperture and can be driven by an external analog voltage signal at a speed of 60 Hz [13], [46]. The driving circuit board adapted for lens operation (Optotune USB Lens Driver 4) is also used. When the lens is driven using an external voltage signal, the lens changes the focal power according to the shape of the signal. We generate a triangular voltage signal using the data acquisition (DAQ) board from National Instruments for the lens operation and apply it to the lens driver. Note that it was reported in [14] that there was negligible difference in system operation when using triangular and sinusoidal voltage signals. Overall, the FTLs sweep between 0.0 D and 5.5 D in our prototype.

We use a Topfoison TF60010A LCD panel with a resolution of 2560×1440 pixels. The maximum frame rate of the panel is restricted to 60 Hz. Owing to the limited LCD modulation speed, the FTLs and LCD are driven at 30

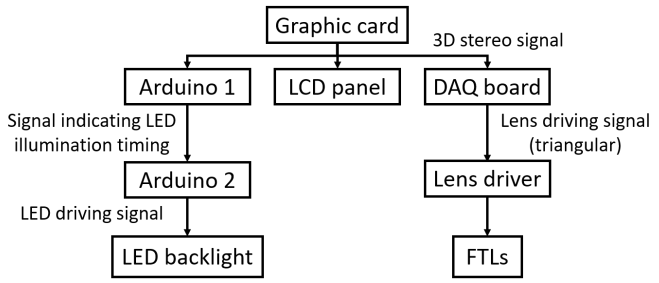

Fig. 9. Schematic showing synchronization of the active components of our prototype. The starting points of all components are equally matched based on the 3-D stereo voltage signal from the graphic card. For LED backlight illumination, two Arduino boards are utilized. The first Arduino outputs 8×40 square pulses during a single period of the stereo signal, and the pulses indicate the moment of turning on the backlight. The second Arduino drives the LED array with several voltage signals designed according to the manufacturer guidelines. According to the stereo signal, the DAQ board generates a lens driving signal so that the FTLs operate in phase with other components.

Hz and 60 Hz, respectively. The LCD panel is modulated twice during a single operation period of the FTLs. We use an NVIDIA Quadro P4000, which can output a 3-D stereo voltage signal to precisely synchronize the display components. Although the brightness may differ from the desirable case of 120 Hz modulation, there is a negligible difference in performance in synthesizing the volumetric scenes. Furthermore, because the display panels with frame rates higher than 120 Hz are already commercialized, we expect it to be sufficiently plausible to realize our design for 60 Hz operation.

For the backlight unit, a 32×64 RGB LED matrix of 3 mm pixel pitch from Adafruit is adopted (Adafruit Product ID 2279). Although the LED matrix is meant to be driven with pulse width modulation (PWM) for bitwise color representation, it is possible to use the matrix as a fast switching binary backlight without PWM for the prototype. Our prototype sequentially floats 20 focal planes, so the LED matrix modulates the frame image at a speed of 20 times the refresh rate of the LCD. We use two Arduino DUE microcontroller boards for running the matrix. The active display components operate in synchronized states as the starting points of the external driving signals are precisely matched to the 3-D stereo signal from the graphic card; Figure 9 briefly explains the synchronization method of all the components in the prototype.

Holographic diffuser with a 60° diffusing angle (Edmund Optics) is inserted between the LED matrix and LCD panel to smooth the edges between the LED pixels. The diffuser and LED matrix are about 3.5 mm apart to achieve a Gaussian kernel with a standard deviation of 1.7 mm (33 LCD pixels). The final display region consists of 8×8 LED pixels and 470×470 LCD pixels, which constitutes an area of about 1.36 inches diagonally. Each FTL is located 50 mm from the display unit. The monocular field of view of our prototype is about 35° diagonally, assuming an eye relief of 14 mm.

5.2 Software

We used computer-generated volumetric scenes rendered using Blender 2.76 in this work. The source of each scene is marked herein, and perspective color intensities and depth maps are extracted using Blender 2.76. We assume a pupil diameter of 6 mm supposing the situation of viewing a bright monitor. The 7×7 perspective images are sampled from uniformly distributed viewpoints inside the pupil. Twenty retinal images for different focusing states are depicted by refocusing the multiview images [47]. The optimized intensity profiles on the LCD are rendered using a PC with two Intel Xeon Bronze 3104 1.70GHz CPUs with 384 GB RAM. Two LCD frames are synthesized using CPU-only implementations and are optimized to reconstruct 20 retinal images correctly; we empirically determined that the optimization required about 150 iterations to converge for each volumetric scene. The optimizations of the two LCD images with RGB color required an average of 62 min for the 150 iterations.

5.3 Calibration

To ensure that our system supports 20 different focal states correctly, we capture each focal plane separated by about 0.37 D using an 8.9 megapixel camera from FLIR (GS3-U3-89S6C-C) and an FTL of 16 mm aperture, which can change the focal power from -10 D to 10 D (EL-16-40-TC-VIS-20D, Optotune). As the lens adjusts the focal length depending on the applied current, we first measured several current values for different imaging distances. Then, we fitted these samples into the mapping functions that convert imaging distance to the current value. We refined driving signals of FTLs, considering the phase delay between lens operation and the signal by manually imaging focal planes with the calibrated lens. We found that the FTL operation and driving signal were out of phase by 25°, which corresponds to $33.33 \times \frac{25}{360} = 2.31$ ms. The computed phase delay was similar to the value in work by Rathinavel et al. [14] (2.38 ms). After calibration, we verified the focusing capability of the proposed system by capturing PSF maps, with each point denoting a different distance, as shown in Figure 8.

5.4 Experiment Results

In Figure 10, we present experimental results compared to the simulation results. The retinal images are numerically synthesized based on the prototype specifications. Before the experiment, the target volumetric scenes were decomposed into two LCD images and 2×20 LRB images. The LCD and LRB images were combined to generate $2 \times (8 \times 8)$ focal blocks. As seen in the figure, our system reconstructs three-dimensional scenes over a broad depth of field. Photographs of the results were captured using an 8.9 megapixel CMOS camera from FLIR (GS3-U3-89S6C-C) and f/1.4 C-mount lens of 16 mm focal length (TUSS LYM1614, entrance pupil size of 11.4 mm).

6 Discussion

6.1 Limitations

The retinal blur fidelity of our system depends on the scene properties since the system tries to fit local scenes with few focal blocks. Although our design reconstructs the depth information better than the DMFD of 6 planes when target volumetric scenes resemble the natural environment,

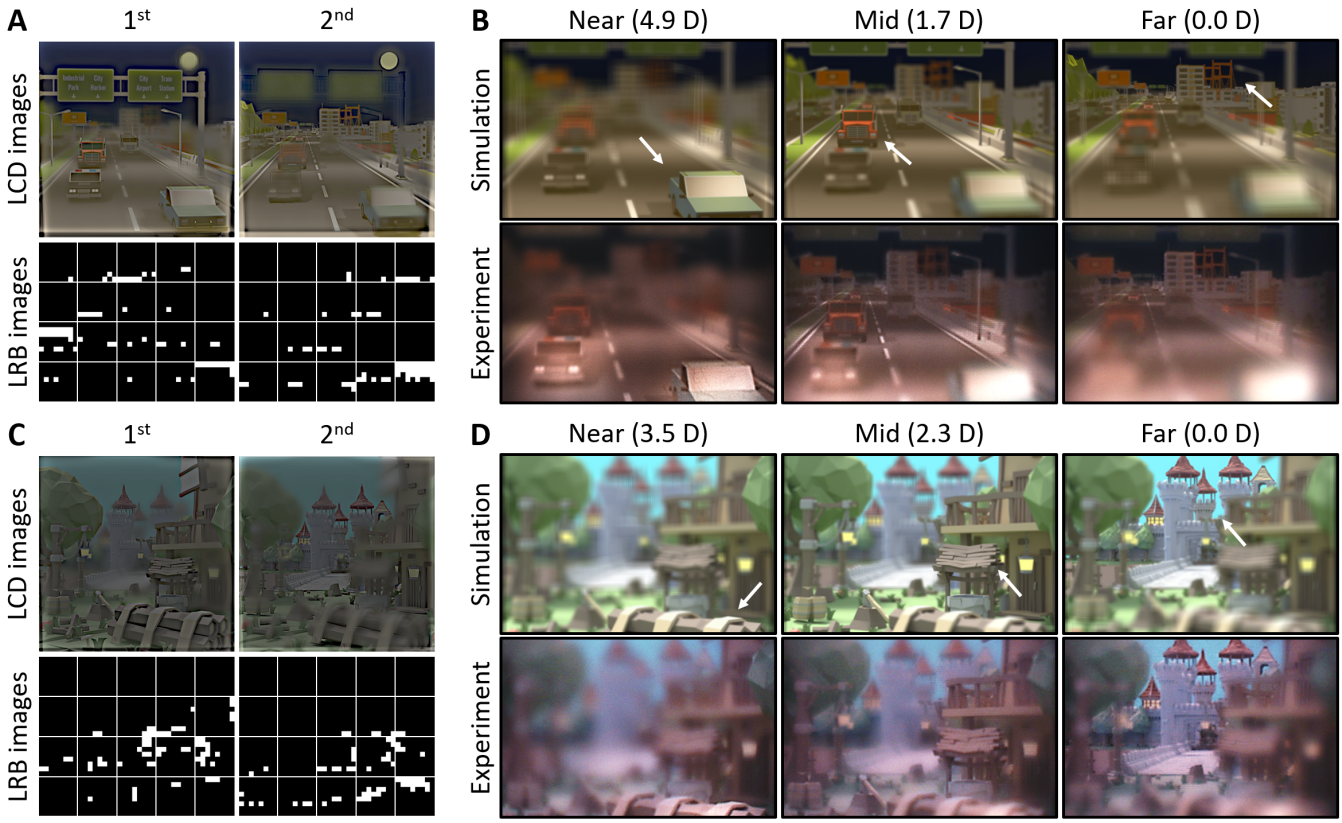

Fig. 10. Experimental results for our prototype. (**A**, **C**) Frame images of the LCD panel and LRB for computer-generated scenes of "road" and "castle". In the experiment, the images labeled as "1st" are displayed during the first half period of FTL sweep, and the images labeled "2nd" are displayed during the second half period. Among the 20 LRB images for each set, the upper left and lower right images represent the LED illuminations when the nearest and farthest focal planes are floated. (**B**, **D**) Comparison between numerical simulation results and captured results: the white arrows denote the focused object for the provided camera setting.

the relationship is reversed when reconstructing scenes of various objects arranged in a stochastic manner. As shown in Figure 3, this scene dependency can be reduced if an LRB with smaller pixels is used.

The diffuser to alleviate screen-door effect could increase the depth errors because of the crosstalk between adjacent focal blocks. We show through numerical simulations that the crosstalk can be alleviated by optimizing the diffusion kernel width. Meanwhile, Akşit [48] demonstrated an approach to improve the pixel fill factor by spatiotemporal means using an optical scanning mirror. As their structure is similar to our system, except for the scanning mirror, the concept could be considered as an alternative solution to improve the backlight fill factor of our system. Furthermore, Sitter et al. [49] proposed a carefully crafted diffractive film instead of a diffuser to increase the fill factor; this scheme could be easily applied to our system by replacing the diffuser with the diffractive film.

The vertical scanning method also sets a practical limit for the depth errors within each focal plane. However, if the refresh rate of the backlight exceeds a particular value and the maximum depth error within the focal plane is less than 0.3 D, the perceptual difference from the ideal case is negligible. Note that the depth of field of the human eye is known to be around 0.3 D under normal viewing conditions [50]. Alternatively, we can remove the errors using a backlight array of active matrix [48].

The real-time operation of our system is difficult due to the computational load for optimization. The timeline for optimization mainly involves extracting the visual saliency and rendering the LDP images. For the saliency map calculation, the required time could be further reduced, as reported in a prior work [51]. For the LDP image rendering, a convolutional neural network (CNN) could be applied to our system. The real-time rendering techniques using CNN for various three-dimensional displays were discussed by Xiao et al. [41].

The limited aperture size of the FTL serves as a barrier to the field of view. However, the field of view could be further expanded by utilizing a liquid lens with a 16 mm aperture (EL-16-40-TC-VIS-5D, Optotune). The operating frequency should be adjusted to about 50 Hz for the wide lens based on the trade-off relationships between aperture size, focal power range, and settling time.

6.2 Future Work

Using CNNs not only reduces the rendering time of LDP images but also can improve our system's perceptual quality. Recently, it was proposed to use neural networks for measuring the perceptual distortions of images since the typical image metrics, such as PSNR and SSIM, often fail to reflect the characteristics of human perception [52]. The

decomposition algorithm considering the movement of the eye pupil [10], [53] could also be applied to our LDP rendering algorithm to expand the effective eyebox. Lastly, the algorithm for focal block arrangements can be improved based on user studies. It would also be useful to determine optimal arrangements for different applications given a large dataset of visual saliencies [29].

7 Conclusion

A commercialized VR HMD floats a single focal plane at a fixed distance. However, vergence-accommodation conflict can occur due to the single focal plane, which causes an observer's visual fatigue. In this paper, we proposed a novel design of VR HMD supporting near-correct focus cues over a wide depth of field (from 18 cm to optical infinity). The proposed design could be realized in a wearable form factor using a low-resolution binary backlight, an LCD panel, and several focus-tunable lenses. We demonstrated through numerical simulations and experiments that a few locally adaptive focal blocks could accurately reconstruct various volumetric scenes. We also built a proof-of-concept prototype using off-the-shelf products to verify the feasibility of the proposed design. We believe that the proposed system could inspire further developments for compact and accommodation-supporting VR HMDs.

Acknowledgments

This research was supported by the Projects for Research and Development of Police Science and Technology under the Center for Research and Development of Police Science and Technology and the Korean National Police Agency (PA-H000001).

References

[1] J. Garrick and M. B. Williams, *Trauma treatment techniques: Innovative trends*. Routledge, 2014, vol. 12, no. 1-2.

[2] C. Moro, Z. Štromberga, A. Raikos, and A. Stirling, "The effectiveness of virtual and augmented reality in health sciences and medical anatomy," *Anatomical Sciences Education*, vol. 10, no. 6, pp. 549–559, 2017.

[3] A. O. Dourado and C. Martin, "New concept of dynamic flight simulator, part I," *Aerospace Science and Technology*, vol. 30, no. 1, pp. 79–82, 2013.

[4] M. Lambooij, M. Fortuin, I. Heynderickx, and W. IJsselsteijn, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *Journal of Imaging Science and Technology*, vol. 53, no. 3, pp. 30 201–1, 2009.

[5] M. Urvoy, M. Barkowsky, and P. Le Callet, "How visual fatigue and discomfort impact 3d-tv quality of experience: A comprehensive review of technological, psychophysical, and psychological factors," *Annals of Telecommunications-annales des télécommunications*, vol. 68, no. 11-12, pp. 641–655, 2013.

[6] G.-A. Koulieris, B. Bui, M. S. Banks, and G. Drettakis, "Accommodation and comfort in head-mounted displays," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, p. 87, 2017.

[7] R. T. Held, E. A. Cooper, and M. S. Banks, "Blur and disparity are complementary cues to depth," *Current Biology*, vol. 22, no. 5, pp. 426–431, 2012.

[8] K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks, "A stereo display prototype with multiple focal distances," *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3, pp. 804–813, 2004.

[9] M. A. Reichow and D. M. Joseph, "Three dimensional display with multiplane image display elements," Feb. 11 2014, uS Patent 8,646,917.

[10] O. Mercier, Y. Sulai, K. Mackenzie, M. Zannoli, J. Hillis, D. Nowrouzezahrai, and D. Lanman, "Fast gaze-contingent optimal decompositions for multifocal displays," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 6, p. 237, 2017.

[11] J.-H. R. Chang, B. V. Kumar, and A. C. Sankaranarayanan, "Towards multifocal displays with dense focal stacks," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 6, pp. 1–13, 2018.

[12] X. Hu and H. Hua, "High-resolution optical see-through multi-focal-plane head-mounted display using freeform optics," *Optics Express*, vol. 22, no. 11, pp. 13 896–13 903, 2014.

[13] S. Lee, Y. Jo, D. Yoo, J. Cho, D. Lee, and B. Lee, "Tomographic near-eye displays," *Nature Communications*, vol. 10, no. 1, p. 2497, 2019.

[14] K. Rathinavel, H. Wang, A. Blate, and H. Fuchs, "An extended depth-at-field volumetric near-eye augmented reality display," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 11, pp. 2857–2866, 2018.

[15] N. Matsuda, A. Fix, and D. Lanman, "Focal surface displays," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, p. 86, 2017.

[16] D. Yoo, S. Lee, Y. Jo, J. Cho, S. Choi, and B. Lee, "15 focal planes head-mounted display using led array backlight," in *Optical Design Challenge 2019, International Society for Optics and Photonics*, vol. 11040, 2019, p. 110400D.

[17] W. Wu, K. Berkner, I. Tošić, and N. Balram, "Personal near-to-eye light-field displays," *Information Display*, vol. 30, no. 6, pp. 16–22, 2014.

[18] W. Wu, P. Llull, I. Tosic, N. Bedard, K. Berkner, and N. Balram, "Content-adaptive focus configuration for near-eye multi-focal displays," in *2016 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2016, pp. 1–6.

[19] D. Lanman and D. Luebke, "Near-eye light field displays," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, p. 220, 2013.

[20] 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 Transactions on Graphics (TOG)*, vol. 34, no. 4, p. 60, 2015.

[21] S. Lee, C. Jang, S. Moon, J. Cho, and B. Lee, "Additive light field displays: Realization of augmented reality with holographic optical elements," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, p. 60, 2016.

[22] A. Maimone, A. Georgiou, and J. S. Kollin, "Holographic near-eye displays for virtual and augmented reality," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, p. 85, 2017.

[23] C. Jang, K. Bang, G. Li, and B. Lee, "Holographic near-eye display with expanded eye-box," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 6, pp. 1–14, 2018.

[24] K. Akşit, W. Lopes, J. Kim, P. Shirley, and D. Luebke, "Near-eye varifocal augmented reality display using see-through screens," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 6, pp. 1–13, 2017.

[25] D. Dunn, C. Tippets, K. Torell, P. Kellnhofer, K. Akşit, P. Didyk, K. Myszkowski, D. Luebke, and H. Fuchs, "Wide field of view varifocal near-eye display using see-through deformable membrane mirrors," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 4, pp. 1322–1331, 2017.

[26] T. Hamasaki and Y. Itoh, "Varifocal occlusion for optical see-through head-mounted displays using a slide occlusion mask," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 5, pp. 1961–1969, 2019.

[27] M. Zannoli, G. D. Love, R. Narain, and M. S. Banks, "Blur and the perception of depth at occlusions," *Journal of Vision*, vol. 16, no. 6, pp. 17–17, 2016.

[28] 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 Transactions on Graphics (TOG)*, vol. 34, no. 4, p. 59, 2015.

[29] K. Akşit, P. Chakravarthula, K. Rathinavel, Y. Jeong, R. Albert, H. Fuchs, and D. Luebke, "Manufacturing application-driven foveated near-eye displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 5, pp. 1928–1939, 2019.

[30] J. Kim, Y. Jeong, M. Stengel, K. Akşit, R. Albert, B. Boudaoud, T. Greer, J. Kim, W. Lopes, Z. Majercik *et al.*, "Foveated AR: Dynamically-foveated augmented reality display," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, pp. 1–15, 2019.

[31] J. Spjut, B. Boudaoud, J. Kim, T. Greer, R. Albert, M. Stengel, K. Akşit, and D. Luebke, "Toward standardized classification of foveated displays," *IEEE Transactions on Visualization and Computer Graphics*, 2020.

[32] Y. Itoh, T. Langlotz, S. Zollmann, D. Iwai, K. Kiyokawa, and T. Amano, "Computational phase-modulated eyeglasses," *IEEE Transactions on Visualization and Computer Graphics*, 2019.

[33] G. A. Koulieris, K. Akşit, M. Stengel, R. K. Mantiuk, K. Mania, and C. Richardt, "Near-eye display and tracking technologies for virtual and augmented reality," in *Computer Graphics Forum*, vol. 38, no. 2. Wiley Online Library, 2019, pp. 493–519.

[34] F.-C. Huang, D. Pajak, J. Kim, J. Kautz, and D. Luebke, "Mixed-primary factorization for dual-frame computational displays." *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, p. 149, 2017.

[35] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, "High dynamic range display systems," *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3, pp. 760–768, 2004.

[36] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915–1926, 2012.

[37] D. Gao and N. Vasconcelos, "Discriminant saliency for visual recognition from cluttered scenes," in *Advances in Neural Information Processing Systems*, 2005, pp. 481–488.

[38] S. Wang, W. Liao, P. Surman, Z. Tu, Y. Zheng, and J. Yuan, "Salience guided depth calibration for perceptually optimized compressive light field 3d display," in *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*. IEEE, 2018, pp. 2031–2040.

[39] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[40] J. Zhang, M. Wang, J. Gao, Y. Wang, X. Zhang, and X. Wu, "Saliency detection with a deeper investigation of light field." in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2015, pp. 2212–2218.

[41] L. Xiao, A. Kaplanyan, A. Fix, M. Chapman, and D. Lanman, "Deepfocus: Learned image synthesis for computational display," in *ACM SIGGRAPH 2018 Talks*, 2018, pp. 1–2.

[42] K. J. MacKenzie, D. M. Hoffman, and S. J. Watt, "Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control," *Journal of Vision*, vol. 10, no. 8, pp. 22–22, 2010.

[43] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (sart): A superior implementation of the art algorithm," *Ultrasonic Imaging*, vol. 6, no. 1, pp. 81–94, 1984.

[44] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli *et al.*, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[45] P. Chakravarthula, D. Dunn, K. Akşit, and H. Fuchs, "FocusAR: Auto-focus augmented reality eyeglasses for both real world and virtual imagery," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 11, pp. 2906–2916, 2018.

[46] S. Choi, S. Lee, Y. Jo, D. Yoo, D. Kim, and B. Lee, "Optimal binary representation via non-convex optimization on tomographic displays," *Optics Express*, vol. 27, no. 17, pp. 24 362–24 381, 2019.

[47] K. Takahashi, Y. Kobayashi, and T. Fujii, "From focal stack to tensor light-field display," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4571–4584, 2018.

[48] K. Akşit, "Patch scanning displays: Spatiotemporal enhancement for displays," *Optics Express*, vol. 28, no. 2, pp. 2107–2121, 2020.

[49] B. Sitter, J. Yang, J. Thielen, N. Naismith, and J. Lonergan, "78-3: Screen door effect reduction with diffractive film for virtual reality and augmented reality displays," in *SID Symposium Digest of Technical Papers*, vol. 48, no. 1. Wiley Online Library, 2017, pp. 1150–1153.

[50] F. W. Campbell, "The depth of field of the human eye," *Optica Acta: International Journal of Optics*, vol. 4, no. 4, pp. 157–164, 1957.

[51] P. Dabkowski and Y. Gal, "Real time image saliency for black box classifiers," in *Advances in Neural Information Processing Systems*, 2017, pp. 6967–6976.

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

[53] 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*, vol. 6, pp. 2170–2180, 2018.

Dongheon Yoo received his B.S. in electrical engineering from Seoul National University in 2017, where he is currently pursuing a Ph.D. in the School of Electrical Engineering. His research interests include three-dimensional displays, virtual reality, near-eye displays, and deep learning.

Seungjae Lee received his B.S. in electrical engineering from Seoul National University in 2015, where he is currently pursuing a Ph.D. in the School of Electrical Engineering. His research interests include 3-D displays, holographic optical elements, augmented reality, near-eye displays, and digital holographic microscopy.

Youngjin Jo received his B.S. in semiconductor engineering from Hoseo University in 2015. He is currently pursuing a Ph.D. from the School of Electrical Engineering at Seoul National University. His research interests include 3-D displays, human perception, augmented reality, and near-eye displays.

Jaebum Cho received his B.S. in electrical engineering from Korea University in 2013. He is currently pursuing a Ph.D. degree from the OEQE Laboratory at Seoul National University. His research interests cover holographic displays, computer generated holograms, deep learning, and optical simulations.

Suyeon Choi received his B.S. in electrical and computer engineering from Seoul National University in 2019, and he is currently pursuing an MSc in electrical engineering at Stanford University. His research interests include 3-D displays, computational imaging and optimization. He was with Seoul National University when this research was conducted.

Byoungho Lee (M94SM00F14) is a Professor of the School of Electrical and Computer Engineering at Seoul National University, Korea. He received his Ph.D. from the University of California at Berkeley (EECS) in 1993. He is a fellow of the IEEE, SPIE, Optical Society of America (OSA), and Society for Information Display (SID). He was the President of the Optical Society of Korea in 2019. He is a member of the Korean Academy of Science and Technology and a senior member of the National Academy of Engineering of Korea. He has served on the Board of Directors of the OSA. He has received many awards, including the National Jin-Bo-Jang Badge of Science of Korea in 2016. His research fields include augmented reality displays, three-dimensional displays, and metamaterial applications.