IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 26, 2017, accepted November 29, 2017, date of publication December 18, 2017,
date of current version February 14, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2782219

INVITED PAPER

Foveated Retinal Optimization for See-Through Near-Eye Multi-Layer Displays

SEUNGJAE LEE, JAEBUM CHO, BYOUNGHYO LEE, YOUNGJIN JO, CHANGWON JANG, DONGYEON KIM, AND BYOUNGHO LEE, (Fellow, IEEE)
School of Electrical and Computer Engineering, Seoul National University, Seoul 08826, South Korea

Corresponding author: Byoungho Lee (byoungho@snu.ac.kr)

This work was supported by the Institute for Information and Communications Technology Promotion Grant funded by the Korea Government (MSIT) (Development of vision assistant HMD and contents for the legally blind and low visions) under Grant 2017-0-00787.

ABSTRACT In order to implement 3-D displays with focus cues, several technologies, including multi-layer displays, have been introduced and studied. In multi-layer displays, a volumetric 3-D scene is represented by 2-D layer images via optimization process. Although this methodology has been thoroughly explored and discussed in optical aspect, the optimization method has not been fully analyzed. In this paper, we deal with pupil movement that may prevent efficient synthesis of layer images. We propose a novel optimization method called foveated retinal optimization, which considers the foveated visual acuity of human. Exploiting the characteristic of human vision, our method has tolerance for pupil movement without gaze tracking while maintaining image definition and accurate focus cues. We demonstrate and verify our method in terms of contrast, visual metric, and experimental results. In experiment, we implement a see-through near-eye display that consists of two display modules, a light guide, and a holographic lens. The holographic lens enables us to design a more compact prototype as performing the roles of an image combiner and floating lens, simultaneously. Our system achieves $38^\circ \times 19^\circ$ field of view, continuous focus cues, low aberration, small form factor, and clear see-through property.

INDEX TERMS Displays, optical signal processing, augmented reality, holographic optical elements.

I. INTRODUCTION

Since people started to have indirect experience of imaginary events and scenery through computer-generated imagery (CGI), there has been a desire to get more immersiveness and realism from the virtual contents. The desire has drawn significant efforts to realize ultimate CGI that may reconstruct a virtual scene indistinguishable from real one. In order to narrow the gap between virtual and real objects, various image processing methods have been introduced to provide more natural effects such as illumination, perspective, shading, and occlusion (psychological cues). Nevertheless, human can easily notice that CGI on ordinary 2D panels is not real, because CGI on ordinary 2D panels cannot provide full of depth information (physiological cues).

Binocular disparity, motion parallax, and accommodation are typical physiological cues that help human obtain depth information from 3D scenes. In order to provide users with physiological cues, hardware structures for displays should be modified. For instance, users may obtain binocular disparity with stereoscopic glasses in 3D theater. Head-mounted displays (HMDs) can imitate motion parallax with head-tracking sensors. Focus cues can be reconstructed by multi-layer display systems. It is feasible to integrate all of these methods into single display module (e.g. head-mounted light field stereoscopes with multi-layers [1]) so that users obtain three physiological cues simultaneously. In short, the ultimate CGI becomes more concrete term with state of the art display technologies.

However, there are still challenges especially for reconstruction of focus cues. When multi-layer displays are employed, 3D scenes are decomposed into 2D layer images that provide continuous focus cues [2]–[5]. In order to find

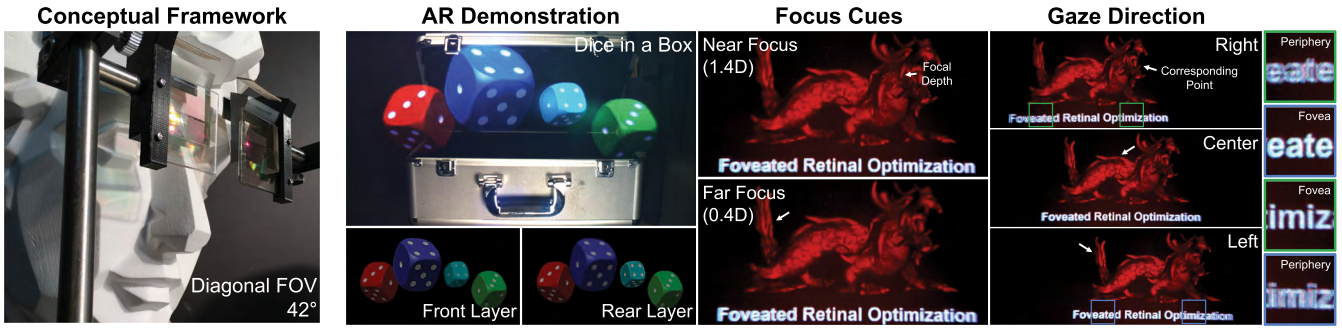

FIGURE 1. Photographs of see-through near-eye display with focus cues using a foveated retinal optimization. Conceptual framework of the proposed display is described on the left-hand side. Dice (virtual object) in a box (real object) demonstrate the see-through property of the prototype where multi-layer images are floated as shown in the bottom row. On the right-hand side, we describe photographs that show discrete focus cues at 0.7 m (1.4D) and 2.5 m (0.4D) and characteristics of the foveated retinal optimization. When the foveated retinal optimization is applied, synthesized images are clear within the fovea while the other part (periphery) is blurred. The clear region of the object varies according to the gaze direction without a tracking system.

optimal layer images, we may formulate and solve optimization problem. How we define this optimization problem is an important issue as it determines accuracy of continuous focus cues as well as fidelity of synthesized images. There have been different approaches to formulate the optimization problem such as light field synthesis [1], [6] and retinal optimization [7]. Nevertheless, it has not been thoroughly discussed and analyzed that these methods show trade-off between eye-box size, definition of synthesized images, and accuracy of reconstructed focus cues.

Here, we analyze relationship between the formulation of optimization problem and display performance. We also introduce a more practical and efficient method to find optimal layer images. This method guarantees enough eye-box, shows compelling definition of synthesized images, and reconstructs precise focus cues. We call this method foveated retinal optimization (FRO) as inspired by key ideas of previous researches related to retinal optimization [7] and foveated rendering [8], [9]. FRO considers the eye-lens and pupil movement [10], assuring accurate focus cues and enough eye-box. Moreover, exploiting the fall of the visual acuity at the periphery, FRO applies weight matrix for retinal images to enhance optimization efficiency. In summary, FRO provides similar experience of foveated rendering without a gaze tracking system.

We expect distinct improvements and benefits with the proposed FRO as follows. First, FRO has enlarged eye-box while maintaining optimization performance within fovea. Second, FRO can be operated with less memory capacity and applied to ordinary multi-layer displays with either additive or attenuation layers. Third, FRO can be adopted for any display system that employs layered panels for focus cues and has a fixed viewing position. Thus, it has a wide range of applications including computational eyeglasses [11], computational displays [12]–[15], and light field stereoscope for VR [1].

In this study, FRO is applied to multi-layer displays with additive layers. We demonstrate the competence of FRO using contrast ratio curves and visual difference (HDR-VDP-2 [16]), which correspond to accuracy of focus cues and fidelity of retinal images. We also implement a prototype for see-through HMDs with additive layers, which is one of core applications of FRO. As our prototype replaces an image combiner and a floating lens with a single holographic lens [17]–[20], the prototype has a competitive form factor as well as field of view (FOV), transmittance, and resolution. Display performance of our prototypes may surpass related works for AR using partially reflective surface [21]–[23], or some other emerging technologies [24]–[27]. Figure 1 shows photographs of our prototype and how FRO works in a practical application.

II. RELATED WORK

A. NEAR-EYE DISPLAYS WITH FOCUS CUES

Along with binocular disparity, focus cues are essential for more immersive and realistic experience. In stereoscopic 3D displays, reconstruction of focus cues becomes more important for mitigation of the visual fatigue caused by vergence-accommodation conflict (VAC). Conventional stereoscopic 3D displays have a single plane where accommodation of users should be fixed to perceive the sharpest images. On the other hand, vergence of eyes could be driven within wide range of depth by corresponding binocular disparity. The VAC is caused by mismatch between accommodation and vergence which are coupled with each other [28]. This conflict not only leads users to get tired but also brings visual difference from real environment, which prevents immersion in the virtual environment. In order to overcome this problem, there have been several efforts to provide focus cues.

Holographic Displays aim to reconstruct original wavefront of real object via phase and amplitude modulation of waves. It is considered as ultimate methodology to provide accurate focus cues for mitigation of VAC. There have been numerous efforts to utilize holographic displays for practical applications. For instance, see-through holographic HMDs using holographic optical elements were reported by several

groups [19], [24], [29]. However, holographic displays suffer from significant trade-off between FOV, resolution, and eye-box size. It is hard to achieve large FOV, high resolution, and enough eye box with state-of-the-art spatial light modulators.

Focus-tunable Displays are another approach to provide focus cues, which may scan a specific depth range. The focus-tunable system could be implemented by using a liquid lens [30], [31] or a digital micromirror device [22]. According to the focus-depth of users measured by additional equipments (e.g. gaze tracker), the depth of imaging plane is modulated to provide precise focus cues. However, the necessity of the gaze tracker makes it hard to achieve enough frame rate and small form factor. If the gaze tracking is not preferred or feasible, the focus-tunable system could be converted to multi-layer displays via the temporal multiplexing method.

Multi-Layer Displays can provide continuous focus cues within a working range (mostly between layers) [1], [7], [23], [32]. Computationally optimized 2D layer images may reconstruct a volumetric object between the layers. In this study, we categorize these methods in terms of optimization approach: light field synthesis [1] and retinal optimization [7]. Categorization standard is optimization targets: 4D light fields or perceived retinal images. Note that display mechanisms (e.g. addition or attenuation) are not considered in this classification because most methods can be adjusted for both mechanisms [1], [11].

B. OPTICAL SEE-THROUGH NEAR-EYE DISPLAYS

There are several methods to implement see-through near-eye displays: partially reflective surfaces [21], [23], [30], diffractive or holographic optical elements [17], [20], [33]–[36], and some emerging concepts [11], [27], [37]. All of these methods have the same goal to achieve a wide field of view, compact form factor, large eye-box, high transparency, and clear formation of virtual images. These optical properties determine economical and technical values of see-through near-eye displays.

In this study, we design a prototype for AR HMD with focus cues using holographic lenses, which extends AR tabletop multi-layer displays [25] to AR HMD. Our system needs a single holographic layer regardless of the number of focal planes, mitigating several limitation of the tabletop prototype such as form factor, ghost effect, and aberration. Also, the prototype has compelling performance compared to representative systems for AR HMDs with focus cues.

C. FOVEATED RENDERING

Foveated rendering utilizes the fall of visual acuity in periphery of retina [38] for enhancement of rendering speed. This approach has clear advantage in HMDs for VR since 3D contents could be updated in faster frame rate [8], [9], [39]. The core idea of the foveated rendering is to render contents of the periphery at the lower sampling rate. It enables rendering speed to be faster while the difference is hard to detect due to the decline of visual receptors in periphery. FRO adopts this idea to improve display performance of multi-layer displays rather than rendering speed of 3D contents.

III. FOVEATED RETINAL OPTIMIZATION

A. FOCUS CUE RECONSTRUCTION VIA MULTI-LAYER DISPLAYS

In this section, we will review the optical principle of multi-layer displays with focus cues. First, it will be described how we define focus cue. Second, we will analyze optical validity of multi-layer displays to reconstruct focus cues. Third, retinal optimization, a computational method to decompose 3D scenes into layer images, will be illustrated [7].

A primary factor to determine the accommodation is contrast gradient of perceived images [40]. The contrast should be smoothly increased to a peak in order to stimulate observers to focus on the desired depth [5]. In the smoothly varying condition, observer likely focuses on a depth where the contrast of perceived images is maximized. Thus, we may guess the depth where the accommodation is stimulated via analysis of the contrast curves, and demonstrate the ability of focus cue reconstruction.

In multi-layer displays, the contrast curve of perceived images is computationally optimized in order to provide appropriate focus cues. For instance, retinal optimization finds layer images by comparing perceived retinal images with target retinal images according to the focal depth of the eyes as shown in Fig. 2(a). In additive layered system, retinal optimization shows superior performance compared to the depth-weighted blending: more natural focus cues and accurate occlusion effect. However, this method supposes that gaze is fixed on a specific direction, usually center, which means that pupil position should be on a static point. This assumption causes undesired noise when gaze direction changes (i.e. pupil swim effect), which will be described in the following section.

B. PUPIL SWIM EFFECT

The pupil swim effect in multi-layer displays occurs because the rotation axis of eyes is separated from the center of eye-lens as shown in Fig. 3 [3]. As the optical axis rotates according to the gaze of the eye, angular shift appears from the multi-layer displays. The angular shift causes dislocation of layer images so that the fidelity of synthesized images is declined. It is determined by the layer gap and the rotation angle (θ_x, θ_y) of the eyes as follows.

$$\begin{aligned}\tan\Delta\phi_x &= \frac{r(d_2-d_1)\tan\theta_x}{\sec^2\theta_x + r(d_1+d_2)\tan^2\theta_x},\\ \tan\Delta\phi_y &= \frac{r(d_2-d_1)\cos\theta_x\tan\theta_y}{\cos^2\theta_x + (1+rd_1+rd_2)\tan^2\theta_y},\end{aligned} \tag{1}$$

where r is the distance between the rotation and optical axes of eyes. d_1 and d_2 are the depths of layer 1 and 2 in the dioptric unit. $\Delta\phi_x$ and $\Delta\phi_y$ denote the angular shift on the yz and zx planes, respectively. The pupil swim effect becomes more significant when the layer gap of multi-layer displays gets wider or the range of the pupil swim expands. According

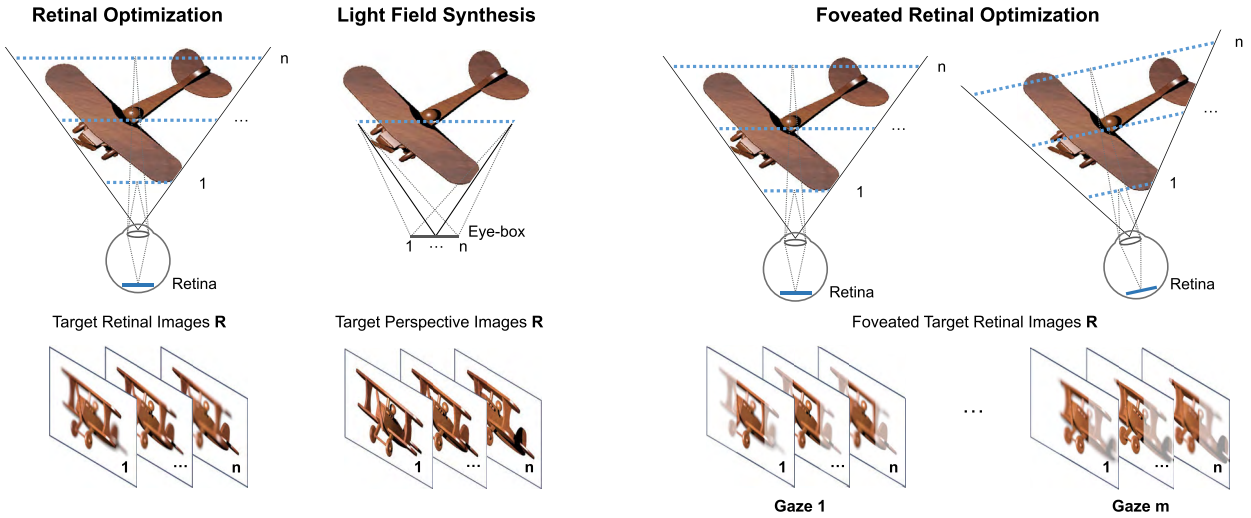

FIGURE 2. Schematic diagram to describe the retinal optimization, light field synthesis, and foveated retinal optimization. (a) Retinal optimization aims to reconstruct target retinal images with an assumption that the gaze is fixed at a reference point. (b) Light field synthesis optimizes light fields within the eye-box. (c) Foveated retinal optimization considers the gaze movement of users so that the system has enough tolerance for the pupil swim. In this method, weight matrix is applied for the retinal images in order to enhance optimization performance in terms of necessary memory capacity and image fidelity within the foveated region.

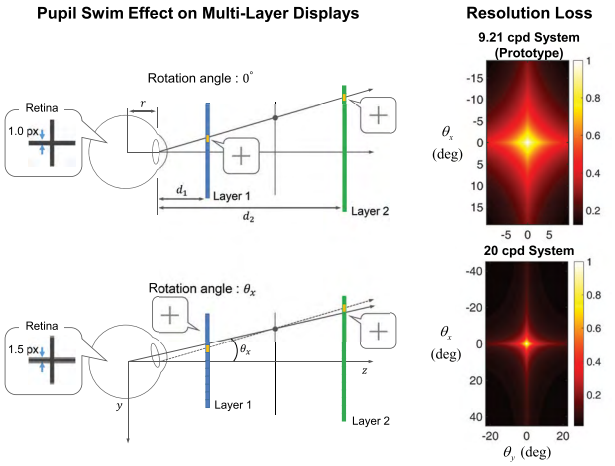

FIGURE 3. Illustration of the pupil swim effect on the *yz* plane. In multi-layer displays, information on an intermediate plane between layers is reconstructed by both layers. If the gaze of eyes moves, the angular shift of coupled pixels occurs between the layers. As the size of synthesized pixel gets larger, the spatial frequency of reconstructed images is degraded if the pupil swim effect is neglected. On the right-hand side, we describe the resolution loss caused by the pupil swim effect. The resolution loss is defined as multiplication of the vertical and horizontal losses. Two different results are shown according to specifications of multi-layer systems: (top) FOV of 38°×19° and resolution of 9.21 cpd; (bottom) FOV of 90°×45° and resolution of 20 cpd.

to Eq. 1, we can describe the maximum frequencies f_x and f_y (cpd) of perceived retinal images synthesized by additive layers when retinal optimization is applied.

Result shown in Fig. 3 implies that the pupil swim effect can limit the maximum frequency of images in practical conditions: near-eye displays with FOV 38°×19°, resolution 700×350 (9.21 cpd), and two layers of 0.6D gap is limited to reconstruct light fields lower than 3.91 cpd (43% of original cpd) when the rotation angle is 10°. This effect becomes much more severe if we suppose further advanced near-eye displays with higher FOV 90° and dense resolution of 20 cpd. In this condition, the maximum frequency is degraded to 2.83 cpd (14% of original cpd) when the rotation angle 22.5°. Moreover, this loss of resolution can be more serious when more layers are adopted to enhance the depth of field.

In order to deal with the pupil swim effect, we may apply alternative approach called light field synthesis. As shown in Fig. 2(b), this approach is inspired by super multi-view displays where focus cues are provided via reconstruction of appropriate motion parallax within eye-box. This approach guarantees large eye-box and the tolerance for movement of pupil (pupil swim). However, layer decomposition using the motion parallax is performed in metric unit while human visual system responses in dioptric unit due to the eye-lens. When metric distance between adjacent layers gets larger, accuracy of focus cues could be diminished. Also, consideration of the motion parallax within eye-box derives degradation in definition of synthesized images. In short, light field synthesis sacrifices accuracy of focus cues and resolution of synthesized images to enlarge eye-box.

C. FORMULATION OF FOVEATED RETINAL OPTIMIZATION

Here, our approach for this problem maintains high definition of synthesized images within the fovea while enlarging the eye-box. Similar with the retinal optimization, FRO also finds optimal layer images by comparing perceived images R with target retinal images R_t. On the other hand, the gaze direction is considered in FRO to enlarge the eye-box. Accordingly, R_t consists of $m \times n$ images that are rendered at corresponding eye state defined by gaze direction (g_n directions) and focus depth of eye (f_m depths).

We demonstrate the significance of FRO in Fig. 4 ('Jadeplant' from the Middlebury 2014 dataset [41]).

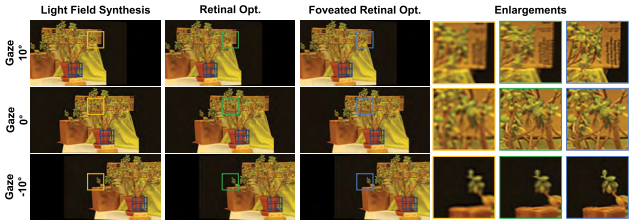

FIGURE 4. Trade-off between eye-box size and image definition. Light field synthesis suffers from limited image definition while retinal optimization is vulnerable to the pupil movement. On the other hand, our approach provides well synthesized images of high definition within the fovea regardless of the gaze direction. For clear illustration, we simulate additive 2-layer display where the two layers are located at 0.4D and 1.6D from observers. Target 3D object is a 2D image that is floated on the depth of 1.0D. Each reconstructed retinal image is perspective view of corresponding gaze direction (or rotation of eyes).

Compared to the light field synthesis [1] and retinal optimization [7], FRO is relatively free from the trade-off among eye-box, definition, and image fidelity. We formulate a weight matrix that enables us to minimize the degradation of contrast within the fovea while enlarging the eye-box. The weight matrix reflects the fall of visual acuity in peripheral region of the retina. As a result, FRO provides similar experience of foveated rendering [8], [9]. However, FRO does not need a gaze tracking system, which is essential for foveated rendering.

In the following sections, FRO is described and analyzed in details including formulation of optimization problem. We first review procedure of retinal optimization [7] and introduce modified retinal optimization, which could be extended to FRO.

1) MODELING HUMAN VISUAL SYSTEM

Retinal optimization supposes an approximated human visual system that consists of a planar image sensor and a thin lens of tunable focal length, each of which corresponds to retina and eye-lens, respectively. Based on the human visual system, perceived retinal images are predicted according to the focal length of eye-lens. The perceived retinal images are acquired by adding up retinal projection of each layer. The retinal projection of each layer can be determined by corresponding incoherent point spread function (PSF) of the eye-lens. PSF for each layer image is defined by depth of the layer z_l and accommodation depth of observers z_f as $h(z_f, z_l)$.

If we suppose a display system with n_L additive layers, perceived retinal images $r(z_f)$ are given by

$$R(z_f) = \sum_{j=1}^{n_L} H(z_f, z_l^{(j)}) L^{(j)}, \tag{2}$$

where capital letter denotes Fourier transform of corresponding variable. $z_l^{(j)}$ is depth of jth-layer, and $l^{(j)}$ denotes displayed images on the jth-layer.

Here, we apply ray approximated PSF rather than conventional definition of PSF based on waveoptics. This ray approximation is valid in human visual system where aberration is significant due to the focusing error [42]. If the ray approximated PSF is applied, retinal optimization could be utilized for multi-layer displays with attenuation layers as well as additive layers. Also, this approximation significantly reduces memory capacity required to solve optimization problem. While conventional PSF corresponds to a dense projection matrix, ray approximated PSF composes a sparse projection matrix P. Using the ray approximated PSF, we can formulate least squares problem to find optimal layer images as follows.

$$\min \left\| \sum_{a=1}^{n_F} \left(r_T(z_f^{(a)}) - P^{(a)} l \right) \right\|, \tag{3}$$

where r_T is target retinal image and elements of the vector l are bounded between 0 and 1.

In order to extend the range of tolerance for the pupil swim while minimizing decrement of optimization efficiency, FRO considers the rotation of eyes that corresponds to the pupil swim. FRO applies an approximated optical model of human eyes that has a fixed rotation axis. FRO considers the rotation of eyes which corresponds to the pupil swim. In order to consider pupil swim, we duplicate the projection matrix along the row as modulating the gaze direction of the eye ($g^{(1)}$, ..., $g^{(n_G)}$) as shown in Fig. 2(c). The gaze direction denotes the vector between the center of retina and the optical axis of the eye. Each duplicated projection matrix links the retina and display panels in the corresponding eye's gaze direction. Each retinal image according to the gaze direction $g^{(k)}$ and focal depth $z_f^{(a)}$ is given by

$$r(z_f^{(a)}, g^{(k)}) = \sum_{j=1}^{n_L} P^{(a,k,j)} l^{(j)} = P^{(a,k)} l. \tag{4}$$

2) WEIGHT MATRIX AND LEAST SQUARES PROBLEM

Finally, we may formulate FRO by applying a weight matrix W based on the assumption that optic nerves are concentrated on the fovea. Then, problem for FRO is given by

$$\min \left\| \sum_{a=1}^{n_F} \sum_{k=1}^{n_G} \left(W r_T(z_f^{(a)}, g^{(k)}) - W P^{(a,k)} l \right) \right\|. \tag{5}$$

Note that $W r_T$ indicates the foveated target retinal images. Weight matrix could be designed according to uniform, step, or human visual acuity distribution model [8] as shown in Fig. 5.

As shown in the simulation results, we may observe the trade-off between central contrast and peripheral fidelity. If uniform weight model is employed, the separation of images is mitigated while contrast within the fovea is traded. We note that weight distribution is not linearly reflected to the contrast distribution of reconstructed retinal images. Because of the correlation of directional view images, we could use drastic weight model (i.e. step model) that ignores periphery region.

In this study, we employ the step model for the weight matrix that has uniform weight only within the central region (8°) and zero weight for the other region. The main

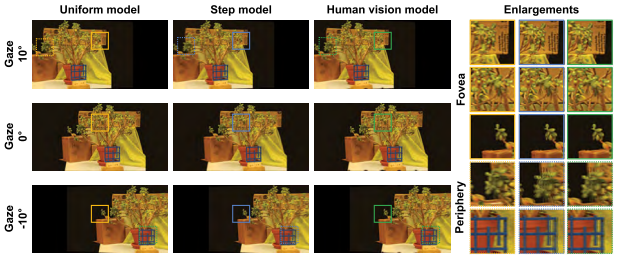

FIGURE 5. Trade-off between contrast in foveated region and fidelity in periphery region according to the weight model. Uniform model: all retinal pixels have the same weight. Step model: only central pixels have the uniform weight while the other pixels are ignored. Human vision model: weight distribution of pixels follow the acuity curve of human vision system.

reason to use the step model is advantage in computation load. The step model is efficient for weight optimization since we can squeeze the projection matrix. The weight matrix could be considered by canceling the corresponding row of projection matrix rather than multiplying with error column vector. Also, we can employ squeezed projection matrix with less number of rows. As a result, optimization for 100 iteration takes 150 seconds when the step model is applied, which is 3 times faster than the human vision model (450 seconds).

IV. IMPLEMENTATION

A. SPECIFICATIONS OF RENDERING AND EXPERIMENT

In this study, we suppose that the pupil is circular with the radius of 4 mm and apart from the retina by 25 mm. We assume that the rotation axis of the eye is fixed on the center of the eye which is separated with the pupil by 12.5 mm. Gaze direction is horizontally varied from -15° to 15°.

We employ the 3D contents with depth information to render the set of target retinal images using POV-Ray and ray tracing method on MATLAB. The spatial resolution of target retinal images is 1200×600 while the resolution of layer images is 350×700. Note that the resolution is different as human visual system generally has higher resolution than display panels. Modulating the focal depth of the eye-lens and the gaze direction, we render 5×7 target retinal images. Focal blur effect is applied to the retinal images based on the ray tracing between 11×11 pixels on the pupil and retina planes.

Simultaneous algebraic reconstruction technique (SART) [43] is employed for FRO. 2.8 GHz 64-bit Intel Xeon E5-2680 workstation with 256GB of RAM is used to find an optimal solution of the least square problems. Foveated region for FRO is assumed as central part (about 9°) of retinal images whose resolution is 300×600. The optimization takes about 150 seconds for the system 1 and 99 seconds for the system 2 using 100 iterations without parallelization. It could be operated at 10 Hz in the following condition: less resolution of retinal image 700×350, 5 iterations, and parallelization with CUDA.

We align the 2-layer images by modulating the displayed position and scaling the size of each layer image. The alignment is performed separately according to the color of the

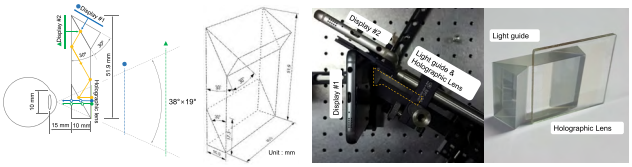

FIGURE 6. Schematic diagram of see-through HMDs with focus cues. The system consists of three parts: display modules, light guide, and image combiner. Two display panels coupled with the light guide are floated at different depths by the image combiner. Detail specifications of the light guide is described on the right-hand side.

layer images in order to alleviate chromatic aberration that occurs by stacking holographic lenses. Layer images cover FOV of 38°×19°, which is demonstrated in Supplementary Material.

Photographs are captured by using a CCD camera mounted with a lens (f/2.8) of Tamron (23FM12-L). The pupil swim of human eyes is approximated to the rotation of the CCD camera around a fixed axis. The rotation axis of the camera is separated by 35 mm from optical axis. In this condition, the pupil swim effect of 15° is approximated to 5° rotation of the camera. This approximation enables us to capture whole synthesized scenes while observing pupil swim effect.

B. SEE-THROUGH NEAR-EYE DISPLAYS WITH FOCUS CUES USING HOLOGRAPHIC LENSES

Fig. 6 illustrates a draft design of see-through displays with focus cues. This system consists of three parts: display module, light guide, and image combiner. Virtual information is generated from the display module and transferred to the image combiner by the light guide. The light guide enables us to enhance the form factor of the display system. After passing through the light guide, the virtual information is integrated into real-world scene by the image combiner. Observer can see the virtual information on a distant plane floated by the image combiner as well as clear real-world scene without distortion. We design this concept considering focus cues, FOV, form factor, image formation, eye-box, and eye relief which are important factors for near-eye displays. These specifications are investigated with experiment and optical simulation tool called CODE V as presented in Supplementary Material.

In the display module, two display panels are employed to establish multi-layer displays and provide users with focus cues. For the light sources of these display panels, laser diodes or organic light emitting diodes (OLED) are good candidates due to the narrow spectral bandwidth suitable for holographic lenses. Light fields from the display panels are merged by a beam splitter and coupled into the light guide. The light fields propagate along the light guide via the total internal reflection. These light fields are coupled out of the light guide by a half mirror which is placed in front of user's eye. As the light fields are introduced to a holographic lens, virtual images of the display panels are formed on distant layers.

C. EXPERIMENTS

A light guide with the specifications is customized as shown in Fig. 6. Detailed specification of the light guide is available in Supplementary Material. OLED panels of Samsung Galaxy S6 panels are used for display panels, which support 570 dpi resolution. We use a portion of display panels whose resolution and size are 700×350 and 31 mm×15.5 mm, respectively. The display gap between the displays and the light guide determines the depth where virtual panels are floated. We verify our prototype in two different conditions: In a system 1, the virtual planes are floated at the depth of 0.7 m (1.4D), and 2.5 m (0.4D). In this specification, we can see more distinct and clear optical blur and pupil swim effect. In a system 2, the depth of virtual panels are set at 1 m (1.0D), and 2.5 m (0.4D). The 0.6D gap between virtual panels is narrow enough to generate continuous focus cues between the layers.

Fig. 7 demonstrates display results of the see-through HMD prototype. Four 3D contents are used to show validity of our prototype. The first results (Chinese dragon) are photographs of the system 1 whose layers are separated by the larger diopter (1.0D). Since the gap between layers is large, the optical blur and pupil swim effect are clearly observed as demonstrated in the results. The other three results (dice, tropical fish, toy plane) are display photographs of the system 2 which can provide continuous focus cues. In these results, we confirm that the prototype can express full-color images and correct focus cues. As we can see in enlarged photographs of 4 different 3D contents, FRO enables the system to show clear images and correct optical blur effects within the interest region where observer gazes. On the other hand, blurred synthesized images are observed in unnoticed area, periphery of the retina.

V. EVALUATION AND ANALYSIS

1) DEMONSTRATION OF ENHANCEMENT

Fig. 8 shows a part of simulation results according to light field synthesis [1], retinal optimization [7], and FRO. We employ HDR-VDP-2 [16] to estimate the visual metric that indicates probability to recognize difference between synthesized images and target images. We perform this simulation for 4 different target objects (Chinese dragon, dice, tropical fish, and toy plane). Using simulation results, we compare each method in terms of point signal to noise (PSNR), foveated point signal to noise (F_PSNR), image quality (Q) [16], and foveated image quality (F_Q). We estimate F_PSNR and F_Q by computing weighted sum of errors where the weight is determined by the slope of MAR line [8]. Obtained factors are described in Table 1. According to the results, we can observe trade-off between definition of synthesized images and eye-box and significance of FRO.

2) ACCURACY OF FOCUS CUES

In order to demonstrate performance of FRO to reconstruct focus cues, we estimate contrast ratio curves shown in Fig. 9.

TABLE 1. Performances of decomposition algorithms in terms of PSNR, F_PSNR, Q, and F_Q.

	PSNR	F_PSNR	Q	F_Q
Light field synthesis	**38.1194**	37.0806	71.0172	71.4405
Modified retinal opt.	37.7128	37.0496	**71.7425**	72.4615
Foveated retinal opt.	37.5877	**38.5902**	69.3947	**74.7313**

In this simulation, we suppose a 2D plane image between layers as a target 3D object. We derive contrast ratio of each frequency by using Fourier transformed retinal images. According to the results, FRO provides accurate focus cues within fovea regardless of the gaze direction. Contrast is also maintained when gaze direction is changed so that high contrast retinal images are achieved.

VI. DISCUSSION

A. FOVEATED RETINAL OPTIMIZATION

We introduce FRO which is a novel layer decomposition method for multi-layer displays with focus cues. FRO considers various characteristics of human visual systems such as optical effects of eye-lens, pupil swim, and foveated visual acuity. Consideration of these characteristics enables us to find optimal layer images in a more convincing and efficient way compared to previous works. For instance, multi-layer displays using FRO could reconstruct more accurate focus cues and provide higher definition image within fovea. Also, FRO has robust reliability for pupil swim effect that causes loss of image fidelity when retinal optimization is employed.

In summary, FRO aims to utilize the decline of attention in the visual periphery for enhancement in eye-box, accuracy of focus cues, and image fidelity. Compared to retinal optimization [7], FRO provides enlarged eye-box and foveated experience [9] without a pupil tracking system. Utilizing ray approximated PSF for retinal optimization [42], FRO may reduce computation load and operate for either additive or attenuation layers. This approach can be applied for a wide range of applications including computational eyeglasses [11], computational displays [12]–[15], and light field stereoscope for VR [1].

There could be another approach to mitigate the pupil swim effect, which was reported in Akeley's thesis [3]. His approach derived from linear blending [2] could be considered as approximation of FRO with step weight model. Our method may exceed this approximation as it is based on the optimal blending [7] rather than linear blending [2]. The main advantages of optimal blending are reproduction of focus cues, occlusion, and blur effect. Furthermore, the separation of images could be mitigated via FRO

FIGURE 7. Photographs of the prototype for see-through HMD with focus cues. The first images (Chinese dragon) are display results of the system 1 with the 1.0D gap where we can observe clear pupil swim effect and focal blur. The other three images (dice, tropical fish, and toy plane) are display results of the system 2 with the 0.6D gap that can provide continuous focus cues. The see-through property is demonstrated with example applications of AR (real objects: a ribbon, box, and toys for fishbowl). Target 3D objects, layer images, focus cues, and significance of FRO are illustrated with the enlargements of the results. Gains of some enlarged images are slightly adjusted to mitigate difference in brightness for fair comparison. Additional experiment, simulation results, and sources of the 3D contents are presented in Supplementary Material.

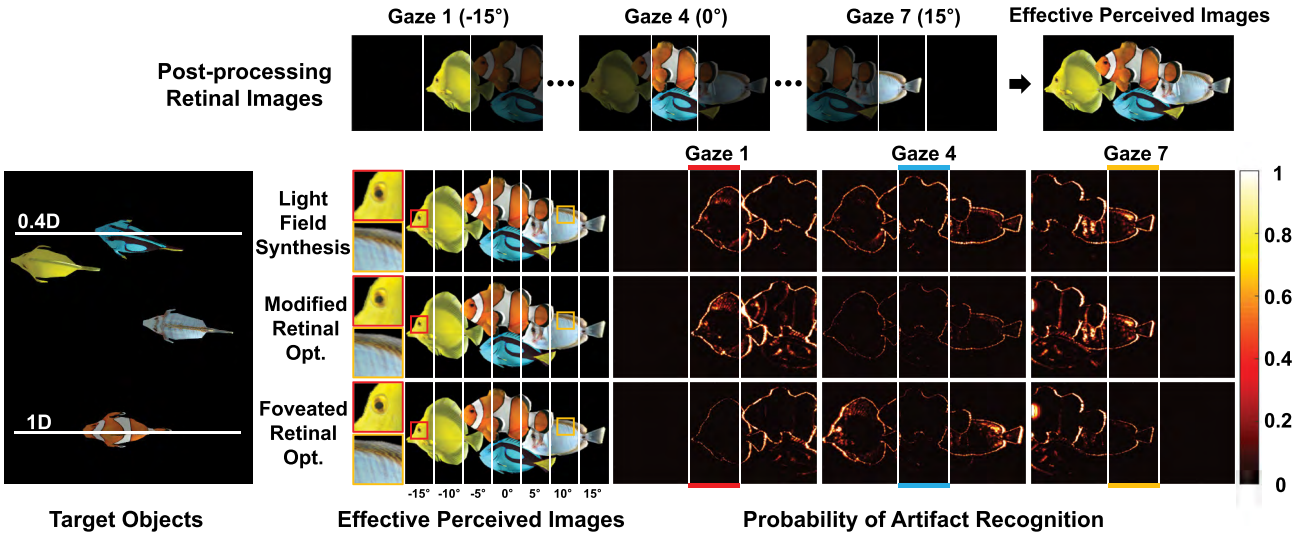

FIGURE 8. Simulation results of light field synthesis, retinal optimization, and FRO for multi-layer displays. On the left-hand side, target 3D objects (tropical fish) are described, which are decomposed into 2 additive layer images by each method. Using decomposed layer images, we predict how user perceives layer images according to gaze direction. In order to give more intuitive comparison, we define effective perceived images obtained by cropping retinal images within central part (8°) and synthesizing cropped images. This procedure is described in the top row and the results are described in the first column. On the right hand side, we also compare quality of retinal images according to gaze direction via HDR-VDP2 analysis. Each image shows 2D probability map [0-1] of noticing visual difference between multi-layer displays and target 3D objects. These maps represent maximum probability over all focal depth of observer. More results can be consulted in Supplementary Material.

as shown in Fig. 5, which is not feasible via Akeley's approximation.

B. SEE-THROUGH NEAR-EYE DISPLAYS WITH FOCUS CUES

We introduce and demonstrate see-through HMD for AR using a holographic lens. The holographic lens can function as a floating lens and image combiner simultaneously, which enables us to design more compact configuration for near-eye displays. We could observe pixel structures of display panels with 570 dpi, which validates the performance of the holographic lens as a floating lens. The prototype achieves a $38° \times 19°$ FOV with the eye-box of 10 mm and the eye relief of 15 mm, which is comparable to commercial products and prototypes of recent research papers. The thickness of the light guide is about 15 mm. The prototype can provide continuous focus cues between 1 m (1D) and 2.5 m (0.4D), which can be adjusted to any other combination that covers diopter range of 0.6D. Since our prototype is relatively free from the diffraction limit caused by stacking LC panels, it is also possible to provide discrete focus cues separated by larger diopter without loss of the resolution.

C. LIMITATIONS

1) CALIBRATION PROCESS FOR FOVEATED RETINAL OPTIMIZATION

Although FRO could be performed for multi-layer displays without a gaze tracking system, calibration process is necessary to find the rotation axis of each user. In practical applications, we should apply a calibration process that consists of 3 steps, which is also applied in this study to determine the rotation axis of the CCD camera. First, system displays

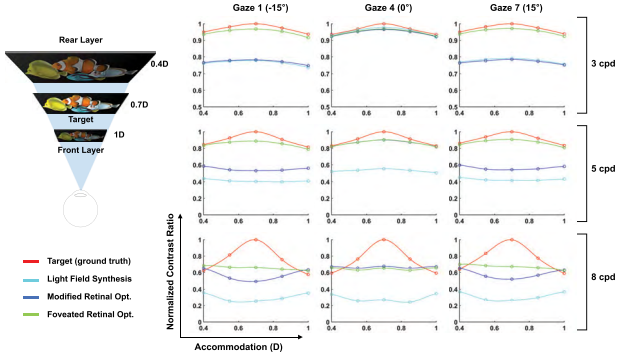

FIGURE 9. Contrast ratio curves of multi-layer displays when corresponding optimization methods and display mechanisms are applied. We derive contrast ratio curves according to the gaze direction by using Fourier transform of retinal images within the central region (8°). As we can see in these results, foveated retinal optimization shows the most convincing and reliable performance for reconstruction of focus cues: we may compare the depth where maximum contrast is achieved, the peak value of the contrast, and variation of contrast gradient.

a cross mark on each layer at the center. Then, users move manually the cross on the front layer to make the two crosses overlapped. Second, system moves the cross marks along the same direction. Note that the two crosses are no longer overlapped since the gaze of users moves and the pupil swim occurs. Again, users make the two crosses overlapped by moving the front layer. The system saves the pixel movement conducted by users. Third, system finds the rotation axis by using relationship between the layer gap and the pixel movement.

2) LUMINANCE AND CONTRAST

The display results of our prototype show limited luminance and contrast since most of light from display panel is attenuated by passing through the light guide and the holographic

lens. At least 87.5% of the light is lost by the two slanted (30°) beam splitters of the light guide and another loss is caused by the holographic lens. The amount of loss by the holographic lens is determined by the diffraction efficiency. In addition, only small angular portion of the light from display panels can arrive to observer's eye. We can enhance the luminance by employing pico projector modules and diffusers instead of display panels. There are some other methods to enhance luminance and contrast simultaneously. Using a backlight source with smaller diffusing angle could be an alternative solution. Smaller diffusing angle is also helpful to diminish the ghost images.

3) MORE THAN 2 LAYERS AND DEPTH OF FIELD

The prototype of the system 1 employs 2 layers to deliver continuous focus cues between the layers. As the prototype reconstructs light fields via addition, the depth of field of the prototype is limited between two layers since each layer is represented by a straight line in the light field spectrum [7], [25]. This limited depth of field should be improved as the accommodation range of human visual system is much larger than 0.6D. In order to enhance the depth of field, we can increase the number of layers. A possible approach is to design hybrid multi-layer displays, which combines two attenuation-based displays [1], [44], [45] by the beam splitter. With this approach, the system can have 4 layers that cover wider depth of field (1.8D) than the current system (0.6D).

VII. CONCLUSION

We formulated a novel optimization method to decompose a 3D scene into 2D layer images, which is called FRO. Inspired by the basis of the foveated rendering, FRO utilizes the fall of visual acuity in periphery of retina for enhancement of optimization performance. Employing FRO, multi-layer displays could reconstruct accurate focus cues, high definition retinal images, and enlarged eye-box. We emphasize that FRO enables multi-layer displays to provide similar experience of foveated rendering without a gaze tracking system. Validity of FRO was thoroughly analyzed and demonstrated via comparison with previous approaches in terms of contrast ratio and visual difference. FRO could be applied to a wider range of applications including multi-layer displays with either additive or attenuation layers. We also introduced a see-through HMD with focus cues, which is one of the most promising applications of FRO. Combining multi-layer displays and holographic lenses, our prototype has advantages in form factor and see-through property. The prototype allowed observers to see real-world superimposed by virtual objects with full-color, high resolution, and focus cues. We expect that our work can be applied to several applications including AR and VR.

ACKNOWLEDGMENT

This work was supported by Covestro AG (formerly, Bayer Material Science AG) for providing the photopolymer Bayfol HX film used for recording the holographic lenses.

REFERENCES

[1] 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.*, vol. 34, no. 4, 2015, Art. no. 60.

[2] K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks, "A stereo display prototype with multiple focal distances," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 804–813, 2004.

[3] K. Akeley, *Achieving Near-Correct Focus Cues Using Multiple Image Planes*. Stanford, CA, USA: Stanford Univ., 2004.

[4] S. Liu and H. Hua, "A systematic method for designing depth-fused multi-focal plane three-dimensional displays," *Opt. Exp.*, vol. 18, no. 11, pp. 11562–11573, 2010.

[5] S. Ravikumar, K. Akeley, and M. S. Banks, "Creating effective focus cues in multi-plane 3D displays," *Opt. Exp.*, vol. 19, no. 21, pp. 20940–20952, 2011.

[6] M. Liu, C. Lu, H. Li, and X. Liu, "Near eye light field display based on human visual features," *Opt. Exp.*, vol. 25, no. 9, pp. 9886–9900, 2017.

[7] 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.*, vol. 34, no. 4, 2015, Art. no. 59.

[8] B. Guenter, M. Finch, S. Drucker, D. Tan, and J. Snyder, "Foveated 3D graphics," *ACM Trans. Graph.*, vol. 31, no. 6, 2012, Art. no. 164.

[9] A. Patney *et al.*, "Towards foveated rendering for gaze-tracked virtual reality," *ACM Trans. Graph.*, vol. 35, no. 6, 2016, Art. no. 179.

[10] M. R. Clark and L. Stark, "Control of human eye movements: III. Dynamic characteristics of the eye tracking mechanism," *Math. Biosci.*, vol. 20, pp. 239–265, Aug. 1974.

[11] A. Maimone and H. Fuchs, "Computational augmented reality eyeglasses," in *Proc. ISMAR*, 2013, pp. 29–38.

[12] A. Maimone, G. Wetzstein, M. Hirsch, D. Lanman, R. Raskar, and H. Fuchs, "Focus 3D: Compressive accommodation display," *ACM Trans. Graph.*, vol. 32, no. 5, Sep. 2013, Art. no. 153.

[13] S. Moon, S.-G. Park, C.-K. Lee, J. Cho, S. Lee, and B. Lee, "Computational multi-projection display," *Opt. Exp.*, vol. 24, no. 8, pp. 9025–9037, 2016.

[14] N. Matsuda, A. Fix, and D. Lanman, "Focal surface displays," *ACM Trans. Graph.*, vol. 36, no. 4, 2017, Art. no. 86.

[15] R. Konrad, N. Padmanaban, K. Molner, E. A. Cooper, and G. Wetzstein, "Accommodation-invariant computational near-eye displays," *ACM Trans. Graph.*, vol. 36, no. 4, 2017, Art. no. 88.

[16] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, 2011, Art. no. 40.

[17] I. Kasai, Y. Tanijiri, E. Takeshi, and U. Hiroaki, "A practical see-through head mounted display using a holographic optical element," *Opt. Rev.*, vol. 8, no. 4, pp. 241–244, 2001.

[18] H. Peng *et al.*, "Design and fabrication of a holographic head-up display with asymmetric field of view," *Appl. Opt.*, vol. 53, no. 29, pp. H177–H185, 2014.

[19] G. Li, D. Lee, Y. Jeong, J. Cho, and B. Lee, "Holographic display for see-through augmented reality using mirror-lens holographic optical element," *Opt. Lett.*, vol. 41, no. 11, pp. 2486–2489, 2016.

[20] S. Lee, B. Lee, J. Cho, C. Jang, J. Kim, and B. Lee, "Analysis and implementation of hologram lenses for see-through head-mounted display," *IEEE Photon. Technol. Lett.*, vol. 29, no. 1, pp. 82–85, Jan. 1, 2017.

[21] J. Hong, S.-W. Min, and B. Lee, "Integral floating display systems for augmented reality," *Appl. Opt.*, vol. 51, no. 18, pp. 4201–4209, 2012.

[22] X. Hu and H. Hua, "High-resolution optical see-through multi-focal-plane head-mounted display using freeform optics," *Opt. Exp.*, vol. 22, no. 11, pp. 13896–13903, 2014.

[23] C.-K. Lee, S. Moon, S. Lee, D. Yoo, J.-Y. Hong, and B. Lee, "Compact three-dimensional head-mounted display system with Savart plate," *Opt. Exp.*, vol. 24, no. 17, pp. 19531–19544, 2016.

[24] H.-J. Yeom *et al.*, "3D holographic head mounted display using holographic optical elements with astigmatism aberration compensation," *Opt. Exp.*, vol. 23, no. 25, pp. 32025–32034, 2015.

[25] S. Lee, C. Jang, S. Moon, J. Cho, and B. Lee, "Additive light field displays: Realization of augmented reality with holographic optical elements," *ACM Trans. Graph.*, vol. 35, no. 4, 2016, Art. no. 60.

[26] K. Wakunami *et al.*, "Projection-type see-through holographic three-dimensional display," *Nature Commun.*, vol. 7, Oct. 2016, Art. no. 12954.

[27] J.-Y. Hong *et al.*, "See-through optical combiner for augmented reality head-mounted display: Index-matched anisotropic crystal lens," *Sci. Rep.*, vol. 7, Jun. 2017, Art. no. 2753.
[28] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, "The zone of comfort: Predicting visual discomfort with stereo displays," *J. Vis.*, vol. 11, no. 8, Aug. 2011, Art. no. 11.
[29] A. Maimone, A. Georgiou, and J. S. Kollin, "Holographic near-eye displays for virtual and augmented reality," *ACM Trans. Graph.*, vol. 36, no. 4, 2017, Art. no. 85.
[30] S. Liu, D. Cheng, and H. Hua, "An optical see-through head mounted display with addressable focal planes," in *Proc. ISMAR*, 2008, pp. 33–42.
[31] 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 *Proc. CHI*, 2016, pp. 1211–1220.
[32] G. D. Love, D. M. Hoffman, P. J. W. Hands, J. Gao, A. K. Kirby, and M. S. Banks, "High-speed switchable lens enables the development of a volumetric stereoscopic display," *Opt. Exp.*, vol. 17, no. 18, pp. 15716–15725, 2009.
[33] H. Takahashi and S. Hirooka, "Stereoscopic see-through retinal projection head-mounted display," *Proc. SPIE*, vol. 6803, 2008, p. 68031N, Feb. 2008.
[34] H. Mukawa *et al.*, "A full-color eyewear display using planar waveguides with reflection volume holograms," *J. Soc. Inf. Display*, vol. 17, no. 3, pp. 185–193, 2009.
[35] Y. Wu, C. P. Chen, L. Zhou, Y. Li, B. Yu, and H. Jin, "Design of see-through near-eye display for presbyopia," *Opt. Exp.*, vol. 25, no. 8, pp. 8937–8949, 2017.
[36] C. Jang, K. Bang, S. Moon, J. Kim, S. Lee, and B. Lee, "Retinal 3D: Augmented reality near-eye display via pupil-tracked light field projection on retina," *ACM Trans. Graph.*, vol. 36, no. 6, 2017, Art. no. 190.
[37] 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," *ACM Trans. Graph.*, vol. 33, Jul. 2014, Art. no. 89.
[38] H. Strasburger, I. Rentschler, and M. Jüttner, "Peripheral vision and pattern recognition: A review," *J. Vis.*, vol. 11, no. 5, 2011, Art. no. 13.
[39] M. Levoy and R. Whitaker, "Gaze-directed volume rendering," *ACM SIGGRAPH Comput. Graph.*, vol. 24, no. 2, pp. 217–223, 1990.
[40] 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," *J. Vis.*, vol. 10, no. 8, 2010, Art. no. 22.
[41] D. Scharstein *et al.*, "High-resolution stereo datasets with subpixel-accurate ground truth," in *Proc. German Conf. Pattern Recognit.*, 2014, pp. 31–42.
[42] J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed. Greenwoood Village, CO, USA: Roberts and Company Publishers, 2005.
[43] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm," *Ultrason. Imag.*, vol. 6, no. 1, pp. 81–94, 1984.
[44] D. Lanman, M. Hirsch, Y. Kim, and R. Raskar, "Content-adaptive parallax barriers: Optimizing dual-layer 3D displays using low-rank light field factorization," *ACM Trans. Graph.*, vol. 29, no. 6, 2010, Art. no. 163.
[45] G. Wetzstein, D. Lanman, W. Heidrich, and R. Rasker, "Layered 3D: Tomographic image synthesis for attenuation-based light field and high dynamic range displays," *ACM Trans. Graph.*, vol. 30, no. 4, 2011, Art. no. 95.

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

JAEBUM CHO received the B.S. degree in electrical engineering from Korea University in 2013. He is currently pursuing the Ph.D. degree with the OEQE Laboratory, Seoul National University. His research interests cover holographic display, computer generated hologram, deep learning, and simulations for optics.

BYOUNGHYO LEE received the B.S. degree in electrical engineering from Seoul National University in 2015, where he is currently pursuing the Ph.D. degree with the School of Electrical Engineering. His research interests include digital holography, 3-D microscopy, phase retrieval, 3-D display, augmented reality, and near-eye displays.

YOUNGJIN JO received the B.S. degree in semiconductor engineering from Hoseo University in 2015. He is currently pursuing the M.S. degree with the School of Electrical Engineering, Seoul National University. His research interests include 3-D displays, human perception, augmented reality, and near-eye displays.

CHANGWON JANG received the B.S. degree in electrical engineering from Seoul National University, South Korea, in 2013, where he is currently pursuing the Ph.D. degree with the School of Electrical Engineering. His primary research interest is in the areas of 3-D display and digital holography.

DONGYEON KIM received the B.S. degree in electrical engineering from Seoul National University in 2017, where he is currently pursuing the M.S. degree with the School of Electrical Engineering. His research interests include 3-D displays, augmented reality, and near-eye displays.

BYOUNGHO LEE (M'94–SM'00–F'14) is currently a Professor and the Head of the School of Electrical and Computer Engineering, Seoul National University, South Korea. He is a fellow of SPIE and the Optical Society of America. He is a member of the Korean Academy of Science and Technology. He has served on the Board of Directors of OSA, where he is currently the Chair of the Design, Fabrication and Instrumentation Technical Division. He has received many awards, including the Jin-Bo-Jang Medal from the President of South Korea in 2016. He has served as the Vice President of the Korean Information Display Society. He is currently the Vice President of the Optical Society of Korea.