Retinal 3D: Augmented Reality Near-Eye Display Via Pupil-Tracked Light Field Projection on Retina

CHANGWON JANG, Seoul National University
KISEUNG BANG, Seoul National University
SEOKIL MOON, Seoul National University
JONGHYUN KIM, Seoul National University
SEUNGJAE LEE, Seoul National University
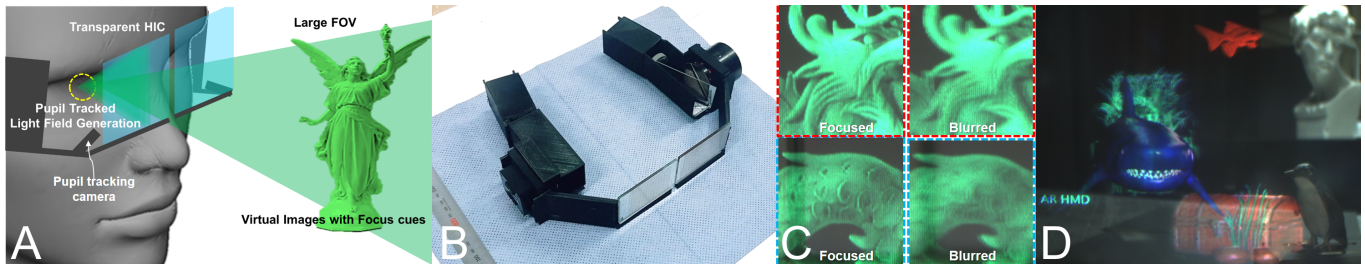BYOUNGHO LEE, Seoul National University

Fig. 1. *A)* Overview of Retinal 3D. The proposed near-eye display provides virtual images with focus cues by generating a localized light field on a pupil with real-time tracking that can then be directly projected onto the retina. Pupil-tracked light field generation can reduce the needed information/computation load as well as provide the eye-box that can overcome the drawbacks of retinal projection-type displays. *B)* A prototype of Retinal 3D. *C)* The experimental results showing the generation of focus cues. *D)* Demonstration of full-color 3D augmented reality.

We introduce an augmented reality near-eye display dubbed "Retinal 3D." Key features of the proposed display system are as follows: Focus cues are provided by generating the pupil-tracked light field that can be directly projected onto the retina. Generated focus cues are valid over a large depth range since laser beams are shaped for a large depth of field (DOF). Pupil-tracked light field generation significantly reduces the needed information/computation load. Also, it provides "dynamic eye-box" which can be a break-through that overcome the drawbacks of retinal projection-type displays. For implementation, we utilized a holographic optical element (HOE) as an image combiner, which allowed high transparency with a thin structure. Compared with current augmented reality displays, the proposed system shows competitive performances of a large field of view (FOV), high transparency, high contrast, high resolution, as well as focus cues in a large depth range. Two prototypes are presented along with experimental results and assessments. Analysis on the DOF of light rays and validity of focus cue generation are presented as well. Combination of pupil tracking and advanced near-eye display technique opens new possibilities of the future augmented reality.

CCS Concepts: •**Hardware** → *Emerging optical and photonic technologies; Displays and imagers;* •**Computing methodologies** → *Mixed/augmented reality;*

Additional Key Words and Phrases: Near-eye display, eye tracking, computational displays, holographic optical element, vergence-accommodation conflict

1 INTRODUCTION

The emergence of augmented reality and near-eye display promises a new era in information production and consumption methods for humans by merging a virtual world with the real-world environment. AR near-eye display is believed to have great possibilities in the education, entertainment, military, and service industries [Azuma 1997]. To open the new era of AR, it is necessary to develop an advanced hardware device that can provide fascinating visual experiences with high performance. Along with growing attention from academia and industry, there have been many efforts in designing optically see-through near-eye displays, or head mounted displays (HMDs) [Kress and Starner 2013]. The ideal AR HMD requires a large field of view (FOV), high resolution, and a large eye-box size with a thin image combiner that has high optical transparency. However, a system that satisfies all the performance factors listed above it yet to be introduced.

In addition, focus cue generation is also an important factor to reproduce a natural visual experience that can reduce the visual

discomfort caused by vergence-accommodation conflict (VAC) [Lambooij et al. 2009]. VAC is a well-known problem in stereoscopic 3D display fields, which is caused by the mismatch between binocular disparity of a stereoscopic image and optical focus cue. In order to mitigate the problem, the display system should be able to provide monocular focus cues that induce the appropriate accommodation of eyes on the convergence plane. As far as we can ascertain, no commercialized AR HMD has yet provided proper focus cues.

In the present study, we introduce an AR near-eye display dubbed "Retinal 3D," which is based on a retinal projection type of display with the novel features. Fig. 1 shows an overview of retinal 3D. Key features of the proposed display system are as follows: A focus cue is provided by tracking the movement of the pupil and generating a light field in the pupil that can then be projected directly onto the retina. The generated focus cues are valid over a large depth range since laser beam is shaped to have a large depth of field (DOF). Pupil-tracked light field generation significantly reduces the needed information/computation load and provides a "dynamic eye-box" which represents a break-through in overcoming the drawbacks of retinal projection types of displays. For implementation, we utilized a holographic optical element (HOE) as an image combiner, which provided high transparency with a thin structure. Compared with current AR displays, the proposed system demonstrates competitive performances of a large FOV, high transparency, high contrast, high resolution, and focus cues within a large depth range. Specific contributions offered by this paper are as follows:

- Pupil-tracked dynamic light field generation with reduced computation load.
- Implementation of a dynamic eye-box for retinal projection displays.
- Combination of full-color holographic optical elements and retinal projection displays with the beam-shaping of a laser source.
- Analysis of the DOF of light rays and validity of accommodation inducing ability.
- Design examples and implementation methods for single/multi-laser scanning projector prototypes with assessments.

2 RELATED WORK

2.1 AR HMDs with Focus Cues

It has been a challenging issue to provide depth information to a monocular eye because this usually requires a large amount of information, a high computation load, and/or a special optical architecture. There are representative display methodologies that can produce focus cues: holographic displays, super multi-view displays, compressive light field displays, and depth-fused displays or multi-focal displays. Holographic displays can physically form the focusing point in mid-air, but the amount of computation load needed and the system requirements are usually very challenging. Yeom et al. [2015] proposed near-eye holographic display, but the resolution and FOV are limited to be practically used. Maimone et al. [2017] have recently demonstrated high-quality holographic display results with large FOV, however it should trade the eye-box size. Alternatively, focus cues can be generated by modulating dense light rays. Super multi-view displays form more than two sampled

view-points entering the pupil [Takaki and Nago 2010]. Hua and Javidi [2014] proposed a novel AR HMD providing focus cues using a freeform optics and a lens-array to generate light field, however the lens-array trades resolution of the image. Takaki and Yamaguchi [2015] have implemented a see-through integral imaging display as a panel type. Tensor displays generate an optimized light field that consists of stacked display panels [Huang et al. 2015; Wetzstein et al. 2011]. Also, in multi-focal displays, the images in multiple discrete depths are intended to induce accommodation between those planes [Liu and Hua 2010a; Narain et al. 2015; Ravikumar et al. 2011] since the accommodation effect is induced dominantly by low-to-middle spatial frequency (4 to 8 cycles/degree, cpd) components [Mathews and Kruger 1994].

2.2 Virtual Retinal Displays

Virtual Retinal Displays. Retinal 3D is based on retinal projection type displays, which is also known as virtual retinal displays (VRDs). VRD is one of the near-eye display technique that is realized by focusing a narrow bundle of collimated rays into the pupil to be projected onto the retina. By forming a focusing point on the pupil of the observer's eye, it can make a broad, uniform illumination at the retina [Westheimer 1966]. This concept can be utilized as a near-eye display method, and it was proposed in the early 1990s [Kollin 1993]. There are some variations to achieve a narrow light ray bundle for VRD systems, as shown in Fig. 2. In this study, a laser scanned by a microelectromechanical mirror (MEMS mirror) is utilized. From the perspective of an AR display, VRDs have distinct advantages. The projection of light rays at the retina guarantees high contrast and a luminance that is appropriate for outdoor use. In addition, a large FOV can be directly achieved by increasing the numerical aperture (NA) of the eye-piece lens. The use of a high NA lens in an imaging system usually causes a severe aberration, or distortion of the image. However, in VRDs, the projection of narrow light rays provides robustness against aberrations from either eye-piece lenses or the user's eye.

Limited Eye-box. Meanwhile, VRDs have the inherent limitation of a limited eye-box. In the optical structure of VRD, the exit pupil must be formed inside the entrance pupil of the user's eye. Consequently, the system cannot allow an eye-box sufficient to accommodate the user's pupil movement, which results in severe vignetting issue to be practically used.

Focus Cues in VRDs. To insure robustness against aberration, VRDs have a large DOF since narrow rays are insensitive to focus changes in the eye lens compared with usual optical imaging based on near-eye displays. Some studies have suggested that retinal-projection types of displays could mitigate VAC by providing all-in focus (or quasi all-in focus) images [von Waldkirch et al. 2003]. When the focus cue is eliminated, the conflict between the vergence and accommodation could be mitigated. Nevertheless, eliminating focus cues would not be the ultimate solution for ensuring visual comfort. Vergence and accommodation are coupled to accelerate each other [Polak and Jones 1990]. When a focus cue is not provided, the accommodation might or might not be tuned to the desired depth since a strong focus cue is not provided.
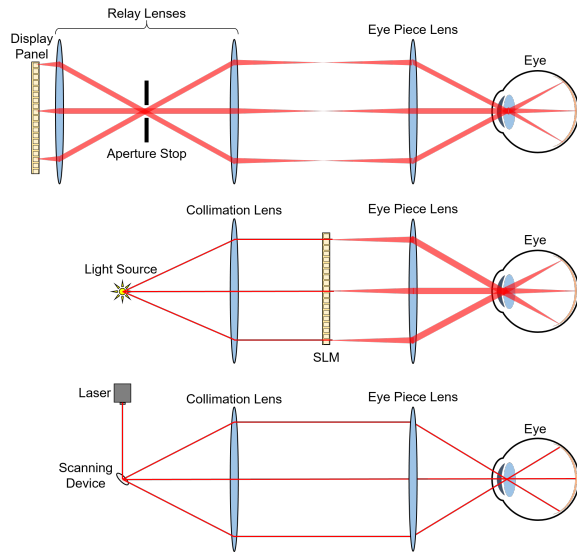
Fig. 2. Basic configurations of VRDs. VRDs can be realized by generating a narrow ray bundle, which enters the pupil of the observer. The ray bundle "draws" an image on the retina and virtual image is shown to the observer. In order to generate narrow light rays, a display panel relayed by a 4-f system with an aperture stop (top), or a collimated light source passed through an SLM (middle), or mechanically scanned laser beams (bottom) can be used.

Furthermore, especially for stereoscopic AR applications, the inconsistency between the blurring of natural scene and virtual image could be problematic. When all the virtual images are sharpened independently of focus distance of eye, out-of-focus virtual images cause sharp-doubled vision, which could facilitate ambiguous depth perception and/or visual discomfort. Indeed, some studies have shown that intentionally inducing out-of-focus blur in a stereoscopic display can reduce visual discomfort compared with doubled vision [Blum et al. 2010; Leroy et al. 2012]. Therefore, providing the appropriate focus cues with a retinal projection display could be a novel solution that would ensure a more realistic visual experience as well as a sufficiently large DOF compared with other methods of generating focus cues. Related concepts have been proposed by [McQuaide et al. 2013; Watanabe et al. 2003], which used deformable membrane mirrors or piezo actuators to form image planes at different depths. Kim et. al. [2011] have proposed a temporal multiplexed multi-focal system, but the system has a bulky form factor to be applied as a wearable AR device.

2.3 Holographic Optical Elements

In this study, we utilized HOEs to build a transparent holographic image combiner (HIC). HOEs are made of holographic recording materials and function as volume gratings or volume holograms [Close 1975; Coufal et al. 2000]. They can be used as traditional optical elements such as lenses or mirrors, by recording corresponding optical waves with a reference wave. Characteristics of volume hologram allows the higher levels of selectivity and optical transparency compared with other image combiner candidates. In addition, the thin structure and high diffraction efficiency of HOEs are well-suited

characteristics for the image combiner. There have been efforts to utilize various types of HOE as guiding optics or eye-pieces of near-eye AR display [Maimone et al. 2017; Yeom et al. 2015]. Also a lens-array HOE can be used to build transparent light field display that can provide a motion parallax [Hong et al. 2014; Jang et al. 2016; Lee et al. 2016].

3 SYSTEM OVERVIEW OF RETINAL 3D

The goal of Retinal 3D is to provide a synthetic image that can induce an appropriate accommodation effect with an enlarged eye-box. To demonstrate an example of hardware implementation, we utilized a HOE as an image combiner. The robustness against the optical aberration is an important feature for the implementation of the system.

3.1 Hardware configuration

The overall configuration of the proposed system is introduced with our design example in Fig. 3. Mechanically scanned laser beams from laser diodes (LDs) are utilized as image sources. The output laser beam is scanned by a set of two scanning mirrors: First, by a MEMS mirror (M1), and second by an electrically controlled fast steering mirror (M2) after passing beam-shaping lenses (BLs, L1), dichroic mirrors (DMs), and a collimation lens (L2). L1 allows a freedom for beam-shaping, while L2 collimates beams. Attenuation and color balancing filters (AF, CF) are used to control the power of laser sources. M2 shifts the exit pupil of the system to the pupil position while the location of the pupil is detected by a small sized pupil-tracking device in real time. The optical path is folded by using an anti-reflection coated half mirror (HM). After being scanned by M2, the beam is obliquely incident on the HIC and converged into the pupil. Lenses, HM, and CF are customized to match specifications such as focal lengths, reflectance, and transmittance respectively.

3.2 Holographic Image Combiner

Herein we present the concept of HIC as well as the basic characteristics of HOEs for clarification. Basically, a HOE is a thin, transparent film that can diffract light as predefined purpose. In this case, it is designed to operate as a "flat" parabolic mirror, only reacts to light of a certain incidence angle and wavelength (which is called a probe beam). Otherwise, it is optically transparent. In this study, we used a photopolymer as the material for a HOE. This photopolymer is optically fabricated with a spherical wave having high NA as a signal beam and an obliquely incident plane wave as a reference beam. Three HOEs were fabricated with different wavelengths and were stacked together to function as a full-color holographic image combiner. The operation of the HIC used in this study is different from a conventional lens or a concave half mirror in two ways: first, it operates with obliquely incident lights. Second, it does not float the image to another plane as in optical imaging. Rather, it directly projects light rays through the user's pupil.

Angular Tolerance of a Reflective HOE. The volume hologram only reacts to specific conditions for a probe beam and a diffracted beam in terms of the "k-vector," which describes the direction and wavelength of light [Hsieh and Hsu 2001]. However, when there is only slight change in the k-vector of probe beam, the HOE can still diffract
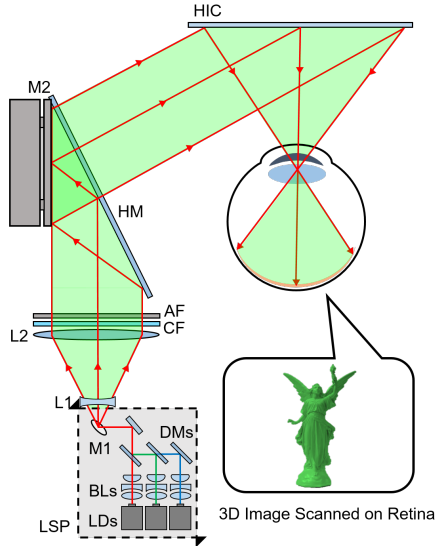
Fig. 3. The configuration of Retinal 3D. Each red line shows beam path of a scanned laser beam which fills green area. The structure of laser beam projector part will be indicated as LSP for simplification. Only a monocular part is depicted in the Figure.
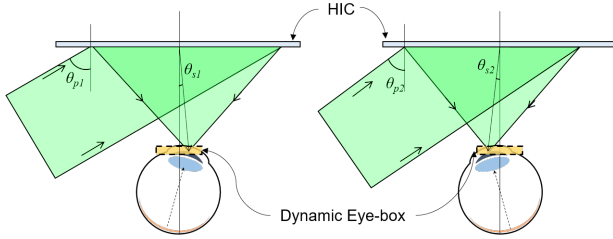

Fig. 4. Dynamic eye-box generation by tilted probe beam incident on HIC. With different incidence angle of probe beams θ_{p1} and θ_{p2}, signal beams with different diffraction angles are generated, forming shifted focusing points.

light with some tolerance, but at a different diffraction angle and efficiency. This effect is addressed as phase mismatch of k-vector. In this study, we induce the phase mismatch to control the diffraction angle by modulating the probe beam angle, so that it can provide a eye-box and focus cues, which will be explained in the following sections. The degree of diffraction angle and efficiency change is primarily dependent on the characteristics of a photopolymer such as thickness, refractive index modulation, degree of shrinkage, or configuration of the HOE. Since we take advantage of the property, we refer to it as angular tolerance instead of angular selectivity. Generally, reflection type HOE with thin thickness has large angular tolerance, which is suitable for the purpose.

3.3 Dynamic Eye-box

For realizing an AR HMD, a VRD has powerful advantages as described in Section 2.4. However, the size of the limited eye motion
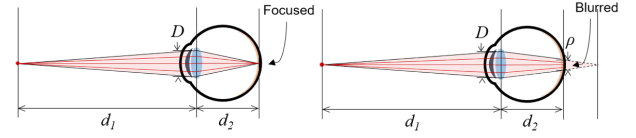

Fig. 5. Conceptual diagram of focus cues generation with light field in human eye. Shadowed area indicate light cone from a point source entering the human eye (natural blurring case) and solid red lines indicate sampled light rays consisting light field. When eye focuses on the point, light rays are focused at a same location in retina as well (left). When eye is focused elsewhere, light rays are not focused but produces blurred image (right).

box has been a crucial drawback, which hindered retinal projection display systems from being practically used. Therefore, enlarging the eye-box of the retinal projection view display without loss of FOV would be an important break-through. In this perspective, we proposed a novel solution by shifting the position of the exit pupil by taking advantage of the optomechanical scanning device and angular tolerance range of the HOE. When the converging wave is recorded as a signal beam, the incidence of tilted probe beam results in a lateral shifting of the focus point of a reconstructed wave, as shown in Fig. 4. By actively controlling this shifted focus point to track the movement of the pupil in real time, we can provide room for the eye to move around. We refer to this eye motion box expansion method as "dynamic eye-box", differentiated from physical exit-pupil size. In the prototype, we used a fast steering mirror (M2) to modulate the phase mismatch of the incident beam on the HIC. Given that the HIC has focal length of h, the focus shift is determined as follows:

$$s = h \tan\left(\sin^{-1}\left(\sin\theta_p - \frac{1+a_n}{1+a_l}\frac{\lambda_s}{\lambda_p}\sin\theta_r\right)\right), \qquad (1)$$

where a_n is refractive index change, a_l is lateral shrinkage, λ_s and λ_p are wavelengths of the recorded and probe beams, θ_p and θ_r are the incidence angles of probe the beam and reference beams, respectively. When we generate a shifted focus point by inducing a phase mismatch, diffraction efficiency drop should also be considered. If the diffraction efficiency is reduced too much, uniform brightness may not be achieved. Therefore the maximum shift should be chosen inside the bandwidth of angular tolerance (see Section B. 4 in the supplemental material). When the maximum shift is provided as s_{max}, the final dynamic eye-box size E can be determined as:

$$E \approx s_{max} + D, \qquad (2)$$

which will allow movement of the pupil, with D denoting the pupil diameter. The intensity of a displayed image can be re-normalized considering the diffraction efficiency to provide uniform brightness. More detailed characteristics of a dynamic eye-box are described in supplementary Section C.

3.4 Focus Cue Generation via the Localized Light Field Scanning

In this section, we introduce one of the key concepts of this work, which is localized light field scanning with pupil-tracking. In light field displays, focus cues are provided by sampling the light rays
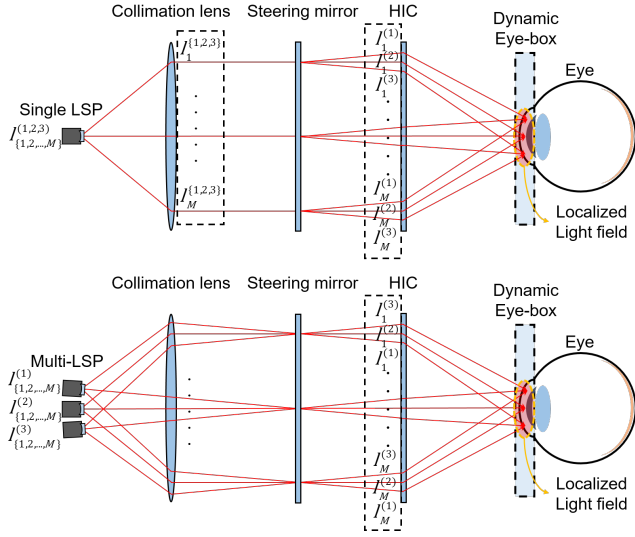
Fig. 6. Schematic of light field generation of Retinal 3D. Light field consists of multiple viewpoints is formed on the pupil location. Each Element, $I_m^{(n)}$, indicates a light ray generated by a single scan of MEMS mirror, and $I_{\{1,2,\ldots,M\}}^{(n)}$ provides the nth viewpoint, where M is the number of light rays generated within a single frame of the LSP. An optical path is unfolded for simplicity. The formation of multiple viewpoints can be achieved either by the temporal multiplexing of a single LSP synchronized with a steering mirror (top), or by using multiple LSPs (bottom).

emanated from an image point as shown Fig. 5. In general, it requires large amount of information or computation load to generate the light field over a large eye-box area. In this study, we produce the light field by forming multiple view-points only inside the pupil, rather than whole eye-box area. This localized light field is cooperated with dynamic eye-box to provide focus cues efficiently.

As shown in Fig. 6, each viewpoint samples a single light ray among the entire light bundle that enters the pupil from a virtual image point. When the observer's eye focuses at the virtual image point, these light rays are projected at the same position on the retinal plane. Otherwise, projected light rays are not focused at the same position which induces blurring of the image. Multiple focusing points can be generated by using the same principle used for the dynamic eye-box. To modulate the incidence angle of a probe beam, the fast steering mirror (FSM) used for a dynamic eye-box can be utilized with the temporal multiplexing, as shown at the top of Fig. 6. The scanning speed of a MEMS mirror decides the total amount of light field information since a single step of the MEMS mirror corresponds to a single light ray. To generate a sufficient number of viewpoints at a sufficient frame rate, the device performance should be high enough. Instead, multiple LSPs can be used to enhance the bandwidth, as shown at the bottom of Fig. 6. Three laser scanning projectors are aligned with slightly different angles to induce a phase mismatch, which formed multiple viewpoints without a temporal multiplexing.

Merits. The concept of focus cue generation is based on super multi-view displays or light field displays. However, in Retinal 3D,
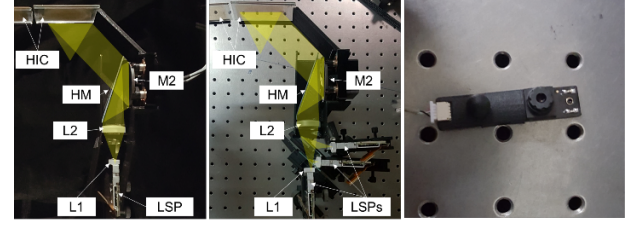

Fig. 7. Pictures of Retinal 3D prototypes. Left) The single LSP prototype, Center) The multi-LSP prototype, and Right) The pupil-tracking camera module.

there are noteworthy differences compared with conventional panel type super multi-view displays. First, since the system is designed as a pupil-tracking scheme, the amount of required information can be significantly reduced. The system proposed here only provides information for a pupil-sized area, while other types of light field displays generate light field information within relatively large area for the eye-box or motion parallax, which will trade the image quality or computation load. In our design, we use only three view images to generate focus cues, without the need for an additional optimization algorithm. Second, since each generated ray in the proposed scheme has a large DOF, the unintended retinal blur effect is much smaller, which ensures a large depth expressible range. These concepts are explained in more detail in Section 6.2.

4 IMPLEMENTATION

Based on the proposed concepts, we designed and built two versions of a Retinal 3D prototype. First, a single LSP prototype was implemented to show the feasibility of an integrated optical system as a compact HMD. Second, a multi-LSP prototype was implemented to demonstrate the performance of the proposed system with high frame rate. Although the single LSP prototype could be designed with a smaller size and a lower cost, a full framerate is difficult to achieve when using the current LSP model. Both systems have the optical path configuration depicted in Fig. 3.

4.1 Display System Implementation

Hardware Design. For the implementation, we used commercialized laser scanning projectors (Picopro, Celluon) as the display sources. The optical engine was detached from the casing and battery part. For M2, a fast steering mirror (OIM5002, Optics In Motion) was used. The input signal for the steering mirror was generated via a DAQ board (PCI-6221, NI) in real-time. The pupil position information was acquired by integrating small pupil-tracking devices (Pupil Labs). When tested by a user or a camera, the prototype and user's head/camera were fixed on the optical table with holders. A pupil-tracking camera module was located off-axis from the eye in order to not disturb the observer's sight. The pupil-tracker was fixed separately at the bottom of the HIC, but the size was small enough to be easily integrated within the frame of a pair of glasses. Specifications of optical components were designed to satisfy the beam-shaping condition, as will be described in Section 6. For the single projector prototype, the focal lengths of lenses L1, L2, HIC

were set as -50 mm, 50 mm, 45 mm, respectively. The optical distance between L1 and L2 was set at 36 mm, while the distance between the HIC and steering mirror M2 was 80 mm. To simplify the design, the optical parameters were optimized for green color only (522 nm). We designed and manufactured the frame using a 3D printer, with the aim of implementing a compact head-mounted prototype. Fig. 7 shows the pictures of Retinal 3D prototypes. The multi-LSP scheme was implemented by adding two more LSPs by using beam splitters and detaching the tube from the front of the L1. Under the assumption of a pupil size of 3 mm, incidence angles for three viewpoints were selected to shift 1.4 mm apart from each other at the pupil domain.

Synchronization. For precise synchronization in a single projector prototype, a VESA signal is used as a reference. When a single view frame ends, the FSM scans beam to the new view position according to the view number. Details of the synchronization and view-image generation are depicted in Supplementary Section C. 4. The frame rate of a single LSP prototype is 10Hz while the multi-LSP prototype operates as 60Hz. When the device performance is enhanced, the amount of the light field data and the frame rate can be further increased.

Calibration. Two types of calibrations are needed for the system implementation. First, the pupil-tracking camera's coordinate and the steering system should be calibrated with the physical coordinate of a user's pupil position. Second, the coordinate system of displayed images should be calibrated to compensate for the off-axis aberrations of the HIC caused by pupil swim. Note that image distortion can be handled separately from the resolution degradation in a VRD. The distortion map is measured at multiple points and inverse distortion is applied to the image to compensate. Details of the calibration are presented in Supplementary Section C. 1.

4.2 Fabrication of HIC

Since it would not be a straightforward process to fabricate a full-color HIC that operates in an off-axis scheme, we have devised some HOE recording techniques to compensate chromatic aberrations. Detailed specifications of the fabrication method are presented in Supplementary Section B. The HOE fabrication system was based and built on the optical table. For the recording beam, three colors (red: 660 nm, green: 532 nm, and blue: 473 nm) of lasers were combined and expanded to be used as signal beams and reference beams. A mercury lamp was used for the curing process after the laser beam exposure process. As a result, we fabricated a full-color HIC that consists of three HOEs. HIC has total thickness of 2.6 mm, mostly due to the glass substrate. A single HOE layer shows nearly 90% of transparency over the visible spectrum. In total, 65% of average transmittance was achieved over the visible region (400-700nm), which include reflections and absorptions of stacked HOEs or substrates when measured.

5 EXPERIMENTAL RESULTS

Display Results. We present experimental results as shown in Fig. 8 to verify the feasibility and performance of the prototypes. Mostly, monochromatic (green) results are presented since the optical parameters for a beam-shaping lens are optimized for a green laser.

However, the full-color results are also presented with reasonable image quality. A CCD camera was used to capture the image. Fig. 8 *A)* verifies the large DOF or generated light rays of the prototype. Without focus cue generation, the image is focused over a large range of depths. *B)* and *C)* demonstrate the focus cue results of the multi-LSP prototype. In both results, displayed images are focused or blurred corresponding to the camera focus. Although the alignment mismatch of the three LSPs induces slight resolution degradation, it provides a full framerate. It is also verified that high transparency is acquired with no degradation of the actual scene. To investigate the blur characteristics of Retinal 3D, viewpoint interval was set differently in *B)* (1.4 mm) and *C)* (2 mm). The results shows larger blur effect in *C)*, simulating the blurring of a larger numerical aperture. However, since the sampling rate becomes sparse, a ripple in the image can be noticed with a larger blurring effect. In both cases, the focus cue generation is clearly shown in accordance with the camera focus change. *D)* shows the results of the single LSP prototype, and also with appropriate blurring. The camera frame was set to 10Hz, in accordance with the display framerate. We also present the full-color AR image display results "Ocean World," as shown in *E)*. A dark curtain is hung behind the grass to not confuse the depth information since the grass is located behind the wall. A miniature penguin is located at 3 D, while the sculpture is located at 0.5 D. For each virtual image, valid focusing and blurring effects are observed.

Dynamic Eye-box Results. Fig. 9 shows the results of a dynamic eye-box. Since it is difficult to demonstrate the pupil-tracking results with an ordinary camera, we built an eye-modeling camera that mimics the rotational movement of a human eye and has similar optical structure. It consists of two rotation stages, a CCD sensor, a lens, and a fake pupil aperture. The eye-tracker was set to track the fake pupil aperture attached to the lens. The radius of rotation is designed to be 12.5 mm. Since the FOV of eye-modeling camera is limited, we placed a grid in the back ground for reference. The dynamic eye-box size was designed at 11 mm (horizontally) by 11 mm (vertically). The result verifies that the viewpoints are properly generated to track the pupil in real time (see supplementary video). The real-time tracking results for the human eye are also presented in the supplementary video. Fig. 9 shows that the eye-box can be dramatically enlarged when a dynamic eye-box is adopted. Since the parameters were set for green color, red color shows slight drop in intensity. For results in the vertical direction, see Section D. 6 in the supplemental material.

6 ASSESSMENT AND ANALYSIS

To verify and further discuss the Retinal 3D, we performed additional experiments and analysis. In this section, we summarize the keypoints of assessments while detailed descriptions are provided in the Supplementary.

6.1 Assessment

Resolution. To measure the resolution of the display system, a slanted edge modulation transfer function (MTF) measurement algorithm is used [Burns 2000]. The resolution is measured over the area of a dynamic eye-box and Fig. 10 plots the resultant MTF curves
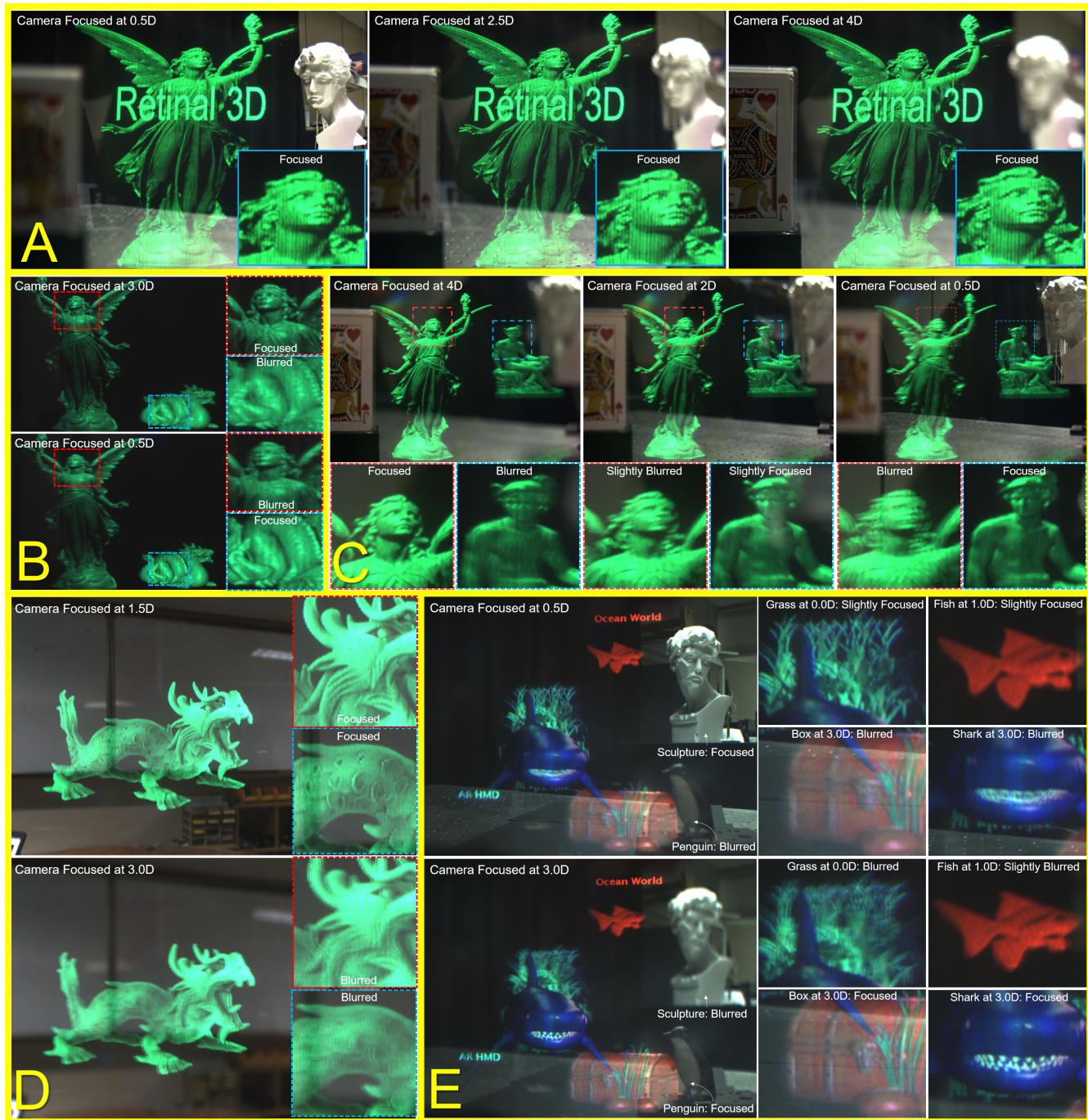
Fig. 8. Photographs of display results. *A)* All-in-focus mode result. To demonstrate the DOF of Retinal 3D, the virtual image is displayed without generating a focus cue to promote comparison with focusing/blurring of real objects: a trump card is located at 25 cm (4 diopters, D) and a sculpture is located at 1.75 m (0.57 D). Regardless of what the camera is focused on, the displayed images and letters are in focus, showing an intrinsically large DOF. *B)* Photographs of experimental results demonstrating the focus cue generation (multi-LSP prototype) were captured in a dark back ground: A statue (Lucy) is displayed at 33 cm (3 D), and a dragon image is displayed at 0.5 D. *C)* The display results with a larger gap between viewpoints which generates larger blurring effect: Lucy is displayed at 3 D, and the Mercury is displayed at 4 D. It is noteworthy that the result shows a larger blur compared with result *B)*, simulating the blurring of a larger numerical aperture. *D)* Captured display results using the single LSP prototype. A dragon is located at 67 cm (1.5 D), showing the appropriate focusing/blurring. *E)* A full-color augmented reality display results "Ocean World," using the multi-LSP prototype. Each image is displayed at different distances: The grass is located at 2 m (0.5 D), the fish is located at 1 m (1 D), and a box is located at 33 cm (3 D) from the observer. A shark swims from infinity distance to a distance of 25 cm (4D). For each object, valid focusing and blurring effects are observed (see supplementary video).
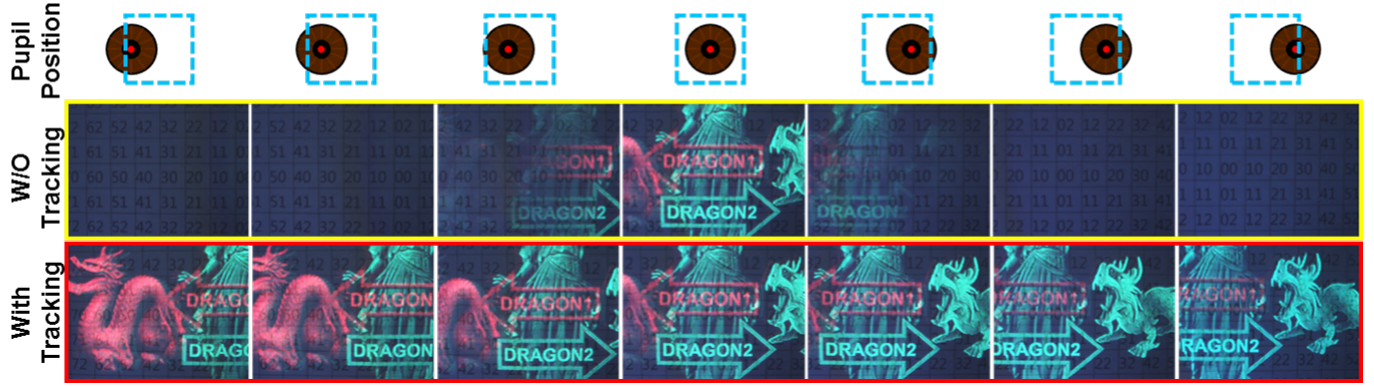
Fig. 9. Experimental results showing the effectiveness of a dynamic eye-box, using the multi-LSP prototype. Without the pupil-tracking, severe vignetting or total image loss occurs when the pupil moves. By adopting a dynamic eye-box with pupil-tracking, the image can be observed in a large eye-box area as shown in the bottom row. Since the used eye-modeling camera had a small FOV, the result only shows the cropped images. The real-time result is provided in the supplementary video clip.
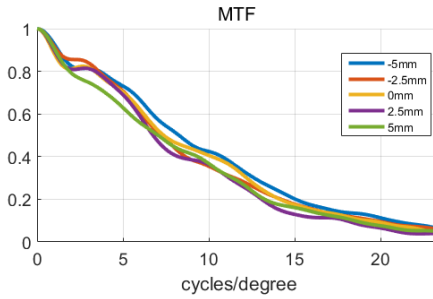

Fig. 10. Resolution of the system over the eye-box area, as measured using the slanted edge method.

along the x-axis. In the best situation, intensity reached 0.5 at 8 cpd and 0.1 at 20 cpd. We believe the performance is competitive against other current prototypes, particularly when compared with the focus cue providing HMDs. For simple visualization, the total area under the MTF curves is plotted according to the locations measured in Supplementary Section D. 1. There is a tendency toward asymmetric resolution degradation along the x-axis, which is natural since the aberration of a HOE is not symmetric. However, the results show that the resolution change due to steering is quite robust over the area of a dynamic eye-box since the worst case shows only a 25% degradation of the MTF area compared with the maximum. In other words, the system is robust to off-axis aberrations, which guarantees the validity of the eye-box expansion using an HIC. Note that the shape of the image could be distorted as we compensated for this with the distortion map as described in Section 4. 1. Nevertheless, the resolution degradation itself is not severe.

Framerate. The multi-LSP prototype can utilize the full framerate of the device (60Hz) since three projectors are used for displaying three view images. When temporal scanning is adopted for view images using a single device, the framerate is reduced to less than one-third of the original framerate because additional scanning of FSM is needed in the intervals of each frame for the view image. In the single LSP prototype, we allocate an entire single frame (16.7ms) to the view transition time, which reduced the framerate to 10 Hz. We believe the improvement of device performance will allow a practical framerate. This will be discussed further in Section 6. 4.

Latency. The delay between the pupil movement and the generation of the voltage signal for FSM (td) was 11.7 ms. The speed of the scanning mirror should be faster than normal eye movement speed to prevent vignetting and a total loss of the image. According to related research [Abrams et al. 1989], the average velocity of saccadic pupil movement varies from $92°/s$ to $174°/s$ when the movement is small ($3° − 9°$), but for the large movement ($60°$), the peak velocity could reach $720°/s$. The maximum scanning speed of a FSM corresponds to $6000°/s$ of eye movement, providing sufficient speed to surpass that of the eye. However, the latency requirements become more important with larger saccadic movements. The total latency between the pupil movement and the display of a single image should be fast enough to prevent a total loss of the image by satisfying the following criteria,

$$T_{tot} < \frac{D}{l\omega}, \tag{3}$$

where $T_{tot} = td + ts$. ω is the angular speed of saccadic eye movement and ts is the scanning delay of a FSM. Assuming the pupil size D =3 mm and the eye diameter l =17 mm, the constraint is 110 ms for $\omega = 92°/s$, and 14 ms for $\omega = 720°/s$. The total delay was 19.7 ms (ts = 8 ms) for the multi-LSP prototype and 28.4 ms (ts = 16.7 ms) for the single-LSP prototype. However, in practice, the single LSP prototype shows an even larger latency of greater than 45 ms, because of the sequential scanning timing (see Section C.4 in the supplemental material).

In conclusion, the steering speed and latency of the multi-LSP prototype can sufficiently cover a slow eye movement or a relatively small saccadic eye movement without noticeable frame loss. However, when there is a large saccadic eye movement, the system
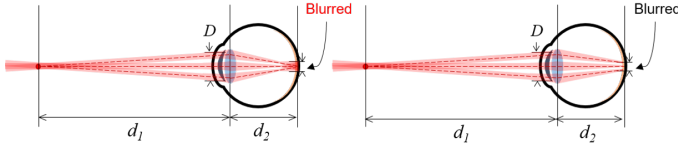
Fig. 11. Conceptual diagram that shows the unintended retinal blur of real light rays. Unlike the ideal case of Fig. 5, light rays have a physical beam width and a diverging angle. In this case, even though the light rays are intended to be focused, changing of the beam width at the retinal plane could cause unintended blurring (left).
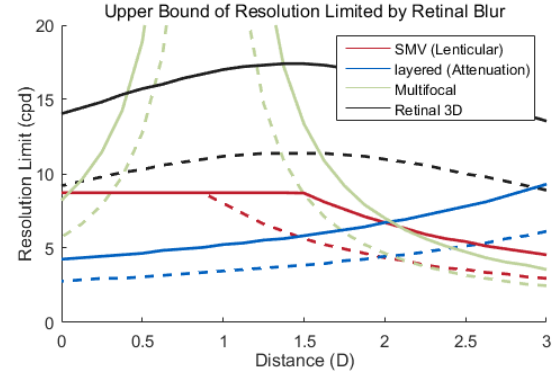

Fig. 12. The upper bounds of resolution limited by retinal blur, for an attenuation type layered display, a lenticular SMV display, a multi-focal display and Retinal 3D. The light ray generated by each display system is simulated to calculate the resolution limit. The angular resolution limit curves are plotted where MTF shows contrast intensity of 0.5 (dotted line) and 0.2 (solid line) respectively, varying with the eye focus distance.

may show vignetting or a potential loss of image frames due to the limited latency of the current prototype. For example, during the eye moves as 720°/s, the system delay (19.7 ms) cannot satisfy the requirement (14 ms), and a dynamic eye-box will be formed outside the pupil, causing temporal image loss. Although, we believe the latency can be further reduced. Since currently td accounts for a majority of the delay, enhancement of the tracking camera performance and the image processing algorithm could reduce the delay. Also, a mechanical delay of FSM can be reduced as well.

6.2 Analysis on Depth Expressible Range

Due to the diffraction of light, an ideal light ray which is perfectly collimated and sharpened does not exist in the real world. Every light ray generated by a light field display system has its diverging angle and beam width. Therefore, each light ray is blurred at the retina depending on the focal state of the eye in practice. This retinal blur of a light ray can conflict with the intended focus cues, as shown in Fig. 11, and therefore limits the depth expressible range, or DOF. This may be not crucial when the intended depth range is small, but it becomes important for near-eye displays, which need to cover a large depth range. From this perspective, Retinal 3D can provide an extremely large DOF by directly shaping the Gaussian beam as described below. Also, we investigated the effect of retinal blur on DOF and resolution limits and compared the results with other types of displays that provide focus cues.

Gaussian Beam as a Light Ray. Since the light generated from a laser diode can be assumed to be a Gaussian beam, we can calculate the beam spot size at the retina according to designed beam parameters. In this study, the display was intended to cover a depth range from 0-3 D. Therefore, the spot size at the retina is designed to be minimized when the eye is focused at middle of the depth range, 1.5 D (R=2/3m). We set the beam waist of a Gaussian beam entering the pupil at near w_{ref}, which is as follows:

$$w_{ref} = \sqrt{\frac{R\lambda}{2\pi}}. \quad (4)$$

Theoretically, the beam width at the retinal plane will retain stable values only showing small variation (9-12 μm) compared to intended blur effect while eye focus changes from 0 D to 3 D, which would not conflict with the focus cues. In this case, an achievable angular resolution of Retinal 3D is calculated around 30 pixels per degree, assuming an eye focal length of 17 mm. Since the resolution is calculated for ideal conditions, it can be degraded by the beam

quality in a practical situation. In addition, because the HIC operates obliquely, the Scheimpflug principle is used to form a waist with the regular beam parameter. Details are presented in Supplementary Section C. 6.

DOF Limited by the Retinal Blur of Light Rays. Here, we present the simulation results of the light ray DOF and its effect on resolution from retinal blur, which takes diffraction into consideration. Similar analysis on the spot size of a light ray at the retina can be applied to other depth-expressible displays: attenuation-type layered displays, super multi-view displays using lenticular lenses, and multi-focal displays. In the attenuation-type layered displays, diffraction in the front panel is usually a main factor that induces the blurring of each light ray component [Huang et al. 2015]. Light ray is modeled to be diffracted as airy pattern at the retina, determined by a pixel pitch of front panel. In the super multi-view display, a beam width determined by a lenticular lens induces retinal blur of the light ray and maximum resolution is limited by the lenticular lens pitch. Meanwhile, multi-focal displays have the same DOF as a natural scene since radiant light is spread from the physical panel. When a display panel is added, an identical graph is plotted at the shifted distance to cover the larger depth. It is claimed a 0 - 3 D range can be covered using 6 planes [Liu and Hua 2010b].

Corresponding point spread functions (PSFs) are numerically calculated based on diffractive optics and are Fourier transformed to obtain lateral MTFs. In the simulation, the pupil diameter is set as 3 mm with a diffraction limited assumption. In order to set the upper limit, possible degradation factors such as aberrations of the optical components, or of eyes are assumed to be ideal. More Simulation details are described in Supplementary Section D. 2. Although defining the DOF of a Gaussian beam is not a straightfoward proposition, slowly varying curves can be understood to have a large DOF [von Waldkirch et al. 2003]. This result shows that spatial frequencies between 4 and 8 cpd, which are dominant for focus cue inducing, can be fully covered with a 0.5 contrast value in Retinal 3D.

In conclusion, Fig. 12 shows that, compared with other types of displays, Retinal 3D can theoretically support a larger depth expressible range, with robust imaging resolution regardless of the displaying depths. For a fair comparison, we point out that the maximum resolution of attenuation type layered display was achieved outside the simulation region since the distance of front panel was set very close to the observer's eye (5.17 D). Still, Retinal 3D shows wider depth range. It is also true that there is the maximum resolution loss compared to multi-focal displays. For valid focus cue generation, however, large DOF is important to prevent unintended retinal blur of a light ray and the degradation of the resolution. Also, in other displays, the optical structure have fixed trade off relation between light ray density and DOF. For example, to generate more dense light rays, a pitch of a panel or lenticular lens should become smaller, which will increase the diffraction effect. On the other hand, in Retinal 3D, a Gaussian beam can be separately shaped to produce large DOF independent of the ray density.

6.3 Artificial Blur Function and Validity of Focus Cues

The feedback process between vergence and accommodation couples each other to find correct focus plane [Polak and Jones 1990]. It would be challenging to consider all the physiological factors of a human visual system to validate the system. Previous evidence, however, supports the accommodation-inducing ability of super multi-view displays [Takaki and Nago 2010]. Also, some studies have validated the capability for appropriate focus cue generation by considering the physiological characteristics of the human visual system [Ravikumar et al. 2011]. Based on a previous study, we analyzed the validity of the accommodation inducing principle by considering both intended blur function for focus cues and unintended retinal blur of light rays.

Retinal Blur Function of the System. In Fig. 5, the shadowed area shows the natural retinal blurring of a human eye that is caused by defocusing. The object point source located at distance d_1 is blurred at the retinal plane located at distance d_2 from the eye-lens with a focal length of f_{eye} and a pupil diameter of D. The amount of focusing error is expressed as follows:

$$\varepsilon = \frac{1}{d_1} + \frac{1}{d_2} - \frac{1}{f_{eye}}. \tag{5}$$

Based on diffractive optics theory, we can calculate the PSF of natural blur for a monochromatic source as follows [Saleh and Teich 1991]:

$$PSF_N(x, y) = \left| h_0 P_1 \left(\frac{x}{\lambda d_2}, \frac{y}{\lambda d_2} \right) \right|^2, \tag{6}$$

where

$$P_1(v, u) = F\{p_1(x, y)\}, \ p_1(x, y) \exp\left(-j\pi\varepsilon \frac{x^2 + y^2}{\lambda} \right) \tag{7}$$

and $p(x, y)$ is a pupil aperture function. Similarly, we can describe the retinal PSF of the proposed system as follows:

$$PSF_R = \sum_{k=1}^{n} \{\delta(x - x_k, y - y_k)\} * \left| g_{f_{eye}}(x, y) \right|^2, \tag{8}$$

where

$$(x_k, y_k) = \left(\frac{1}{2}\varepsilon d_2 s_{x_k}, \frac{1}{2}\varepsilon d_2 s_{y_k} \right). \tag{9}$$
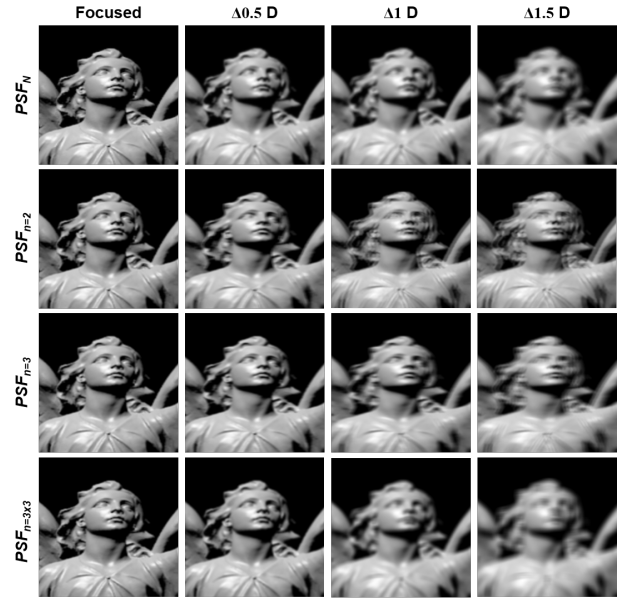

Fig. 13. Simulated image blur corresponding to types of PSFs. Top row shows a natural blur (PSF$_N$) while the others shows the results of PSF$_R$ varying the number of viewpoints (n). n=3 shows the blurring of implemented prototypes (third row).

s_{x_k}, s_{y_k} denote the location of kth sampled view point decided by focus shift in Equation 1 while $g_{f_{eye}}$ denotes amplitude distribution of a Gaussian beam at the retina with an eye focal length f_{eye}. Finally, the image at the retinal plane I_r can be represented as a convolution of the original image and PSF_R as follows:

$$I_r = I_o * PSF_R, \tag{10}$$

where I_o denotes the intended image displayed at a depth of d_1 without blurring. Fig. 13 shows the simulated retinal image with calculated PSFs for natural cases and for the proposed display. The simulation conditions are identical to previous descriptions. In general, the behavior of artificial blurring is in consistent with the natural blurring case and it can be confirmed that the more views provides the more natural blurring. However, additional analysis is needed to tell how many views are needed for valid focus cues.

Validity of Focus Cues: How Many Views are Needed? [Ravikumar et al. 2011] have analyzed the validity of proper accommodation inducing ability based on artificial PSFs of the multi-focal display system. The accommodation is induced towards the depth of where the perceived image contrast is maximized, and is driven by the gradient of the image contrast curve. The perceived image contrast C is defined as the integration of the MTF curve multiplied by the neural contrast sensitivity function (CSF), which is a weighting function that describes human sensitivity to spatial frequencies as follows:

$$C = \int MTF(f) \times CSF(f) \times 1/f \, df, \tag{11}$$

where f denotes a spatial frequency. The factor $1/f$ is multiplied to model the spatial frequency of a usual image source and a neural
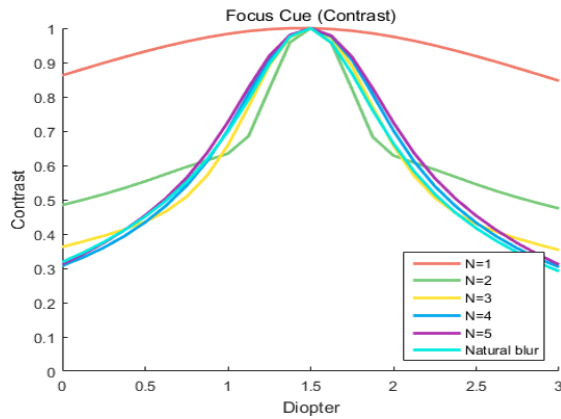
Fig. 14. Image contrast according to the number of viewpoints of the generated light field. Using calculated PSFs, image contrast is calculated for the intended displaying depth (1.5 D) and focal lengths of the eye in diopter.

Fig. 15. Calculated image contrast when n=3 according to the focus depth of the eye lens, normalized by maximum value. Intended depth is labeled with different color while x-axis indicates real focus distance of eye lens.

CSF is adopted from [Michael et al. 2011]. We adopted the same analysis to show the validity of focus cue generation. Fig. 14 shows the image contrast curve according to the number of lateral viewpoints as well as cases of natural blurring. With only a single view, the accommodation cannot be induced since the curvature is very slow. When the number of viewpoints is increased, the curve converges close to natural blur as also shown in Fig. 13. When the number of viewpoints is increased by three or more, the image contrast shows quite similar curvature to a natural blur. Although an accommodation is expected to be induced even when n=2, the curve shows a distinctively different shape compared with the natural blur. Especially with the large blurring, it shows the larger contrast and smaller gradient. This would suggest either an unnatural blur or the perception of a doubled image, which is observed in Fig. 13. Since the shape of a curve shows only slight difference from n=3, we chose to generate a minimum of three views to gain valid focus cues. Each curve in Fig. 15 indicates the image contrast for n=3 as our prototype, which varies with the eye focus distance when the virtual image is intended to be located at the labeled distance. This result shows the highest contrast value at the intended depth with a smooth gradient, and suggests that accommodation can be properly induced over a 0 - 3 D depth range.

6.4 Discussions

Display Performance. Retinal 3D prototype shows competitive display performances compared to other AR HMDs prototypes. We emphasize that the combination of a retinal projection method and a HIC enables a large FOV with high transparency as well as offering a large DOF for the focus cue range as shown in the experimental results. The achieved FOV is 55°(horizontal) by 40°(vertical) with a transparency of 65%, and an eye-box size of 11 mm by 11 mm. The MTF showed 0.5 intensity for 8 cpd, which represented a degradation of 30 - 40% from the theoretical value. The brightness/contrast was sufficiently high, and suitable for AR applications. Additionally, although the details are not emphasized in the paper, the proposed display naturally provides a vision correction function [Huang et al.
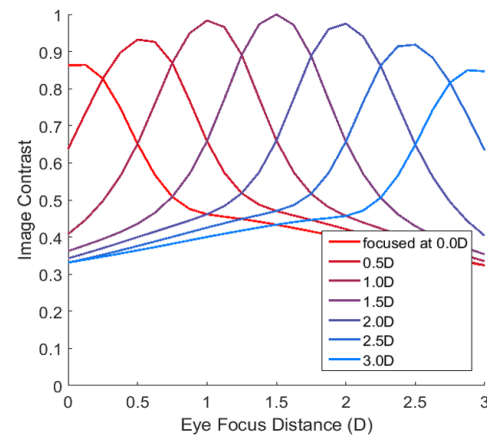
2014] as a Maxwellian view display. We expect this property will enable some interesting applications such as 3D movies with all-in focused subtitles, or to be used as an information display for legally blind or weak sighted people (e.g., with severe aberration caused by corneal transplantation).

Merits of HIC. By substituting the lens for an HIC, the system obtains superior optical transparency with a thin form factor that cannot be obtained with either a partial mirror or a concave half-mirror. Also, the oblique operation of HOEs enables a compact structure with no guiding optics as well as an efficient optical path for wearable devices. Usually, the oblique alignment of an optical element causes aberration issues, which are complicated to address. In particular, HOEs as imaging components usually show astigmatism and comm that degrade the image quality. However, via the beam-shaping of light rays, we could effectively suppress high frequency components that mainly contribute to optical aberrations. Therefore, the HIC is a well-suited choice for an AR HMD with a simple form factor and without severe degradation of image quality.

Reducing the form factor. Current prototype HMD has a form factor that is a bit larger than commercialized HMDs. Nevertheless, we expect further compact versions can be implemented with additional optical design techniques. For example, here we suggest some possible solutions for minimization. First, a wedge-light guide or free-form optics can be used to replace optical elements and reduce the optical path. Also, we can fabricate a HOE that operates with a spherical reference wave without collimation. When a collimation lens is not needed, the required optical path length and aperture size of FSM can be reduced. For now, the FSM is the bulkiest part of the system. We expect that it will be possible to build a compact version by adopting large MEMS mirror (7-10 mm is available in the market) instead of current FSM. Furthermore, scanning mirror could be replaced by a non-mechanical beam steering element such as a tunable liquid lens or LCoS SLM [Zohrabi et al. 2016]. We leave minimization of the HMD to future works.

Trade-off of Framerate and Number of Light Field Views. There is a trade-off relationship between the framerate and amount of information needed by the image in the LSP. Our LSP has a resolution of 1280×720 with a framerate of 60Hz, which generates 55296k of light rays per second. Using a raster scanning mode, the MEMS mirror should scan 43.2k lines per second and oscillate at 27 KHz with an 80% duty cycle [Hofmann et al. 2012]. Fig. 16 shows the trade-off curves for the total information amount of light field (indicated as the number of generated viewpoints) and fps decided by the device performance (oscillation frequency of MEMs mirror) and settlement time of the FSM (*ts*) when the resolution is fixed as 1280×720. In practice, the number of viewpoints should be larger than 3 to ensure appropriate accommodation, and the framerate should be larger than at least 15 fps for displaying a movie. Under these conditions, a 162 KHz MEMS mirror will enable full sequential light field scanning over eye-box area (with 8×3 viewpoints in 10 mm × 10 mm eye-box) with a single LSP, with assuming the *ts* is nearly zero using non-mechanical scanning device or using the optimized scanning sequence. Meanwhile, localized light field with pupil-tracking enables the scanning with a dramatically reduced information amount and hardware requirements. The requirements for light field scanning can be relieved with light field optimization or by adopting a prediction algorithm for eye movement.

More Natural Blurring. In this work, the feasibility of focus cue generation is demonstrated via the generation of three horizontal viewpoints. We confirmed that further natural blurring could be achieved by generating more viewpoints by simulations. An increase in the number of viewpoints fundamentally requires an increase of needed information amount. Thus far, the bottleneck is the framerate of MEMS mirror as shown above. When the scanning device's performance is enhanced, a more natural blurring can also be achieved. Alternatively, multiple LSPs can be used with temporal multiplexing at the same time.

Pupil-Tracking Reliability and Stability. In consistent lighting, the tracking operates quite stable and shows less than 0.5 mm error in the pupil domain. Provided that the detection error is smaller than the pupil diameter, the displayed image does not disappear or move from intended position, and only vignetting may occur. However, in practice, the lighting condition and rapid movement of pupil makes detection errors causing temporal missing of tracking information stream. When tested by users, vignetting or instability (missed image frame) was occasionally noticed with the rapid eye movement speed. In addition, there could be eye-relief variation in practical use, which will cause a calibration error. We expect the advanced tracking algorithms such as gaze prediction/3D position detection can enhance the reliability. Also, it is needed to design a stable hardware structure to inhibit the eye-relief variation when worn by users.

Artifacts. The full-color results showed a slight degradation in resolution of the image. This is partly because optical parameters are optimized for a green laser only, because there were limitations in modifying the optical engine module. However, optimizing for the every wavelength would not be very challenging in the mass production stage. Also, there is an artifact caused by the LSP's own
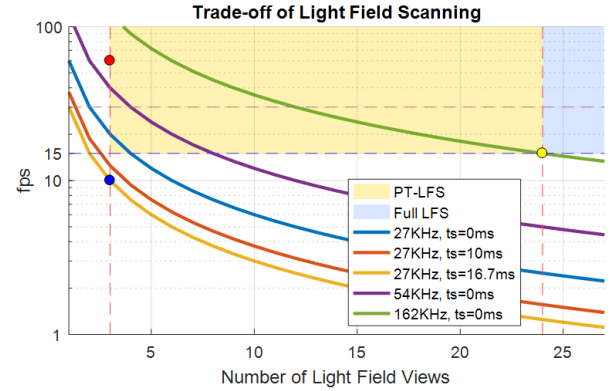

Fig. 16. Trade-off relationship between framerate and generated light field information amount. Each line shows the trade-off curve corresponding to scanning device performance and system latency. The yellow area shows the region where pupil-tracked light field scanning is available (PT-LFS) when 15 fps is satisfied with more than three viewpoints, which supports valid focus cues. The blue area shows the region that enables full light field (Full LFS) scanning over the eye-box. The red point indicates the multi-LSP prototype while the blue point indicates the single LSP prototype. The yellow point shows the minimum requirements for full light field scanning over the eye-box area.

color representing method. For example, to represent red color, green and blue lasers are turned on at the same time as well as a red laser. Since three HOE layers are fabricated separately, it causes the color separation looks like a crosstalk. However, it is not originated from a crosstalk of HOE itself, and therefore can be simply solved when only a single laser is used for the corresponding R, G, B HOE. The beam quality degradation caused by optical components or the laser performance is another artifact of the prototype which affects the overall resolution. Lastly, the single LSP prototype generated continuous noise of the FSM.

7 CONCLUSIONS

The emergence of mixed/augmented reality and near-eye display promises a new paradigm of user experiences. By bringing the displays in front of the eyes, a whole new form of interaction can be realized between the virtual world and real world environments. In particular, we believe pupil-tracking will be an essential part of future AR. As the gaze is one of the most intuitive user interface that can be used even without learning, it can be utilized as a efficient intermediary between the user and the device.

We demonstrated how the pupil-tracking can be used to overcome both computational and optical constraints of near-eye displays. Localized light field scanning can reduce the required computation load efficiently as well as provide the eye-box for retinal projection type displays. We built the prototype AR HMDs using a holographic image combiner, achieving competitive display performances such as a high FOV, high resolution, a sufficient eye-box, and focus cues as well as high transparency and a thin structure. We hope Retinal 3D is a meaningful approach towards the realization of the future AR displays.

ACKNOWLEDGMENTS

REFERENCES

Richard A. Abrams, David E. Meyer, and Sylvan Kornblum. 1989. Speed and accuracy of saccadic eye movements: characteristics of impulse variability in the oculomotor system. *Journal of Experimental Psychology: Human Perception and Performance* 15, 3 (1989), 529–543.

R. T. Azuma. 1997. A survey of augmented reality. *Presence: Teleoperators and virtual environments* 6, 4 (1997), 335–385.

T. Blum, M. Wieczorek, A. Aichert, R. Tibrewal, and N. Navab. 2010. The effect of out-of-focus blur on visual discomfort when using stereo displays. In *In Proceedings of the IEEE Science and Technology International Symposium on Mixed and Augmented Reality*. IEEE, Los Alamitos, CA.

Peter D. Burns. 2000. Slanted-Edge MTF for digital camera and scanner analysis. *Proc. IS&T 2000 PICS Conference* (2000), 135–138.

D. H. Close. 1975. Holographic optical elements. *Opt. Eng.* 14, 5 (1975), 145402.

Hans J. Coufal, Demetri Psaltis, and Glenn T. Sincerbox. 2000. *Holographic Data Storage*. Springer-Verlag.

U. Hofmann, J. Janes, and H. J. Quenzer. 2012. High-Q MEMS resonators for laser beam scanning displays. *Micromachines* 3 (2012), 509–528.

K. Hong, J. Yeom, C. Jang, J. Hong, and B. Lee. 2014. Full color lens-array holographic optical element for threedimensional optical see-through augmented reality. *Opt. Lett.* 39, 1 (Jan. 2014), 127–130.

Mei-Li Hsieh and Ken Y. Hsu. 2001. Grating detuning effect on holographic memory in photopolymers. *Opt. Eng.* 40, 10 (2001), 2125–2133.

Hong Hua and Bahram Javidi. 2014. A 3d integral imaging optical see-through head-mounted display. *Opt. Express* 22, 11 (June 2014), 13484–13492.

Fu-Chung Huang, Kevin Chen, and Gordon Wetzstein. 2015. The light field stereoscope immersive computer graphics via factored near-eye light field displays with focus cues. *ACM Trans. Graph.* 34, 4 (2015), 60.

F.-C. Huang, G. Wetzstein, B. A. Barsky, and R. Raskar. 2014. Eyeglasses-free display: Towards correcting visual aberrations with computational light field displays. *ACM Trans. Graph.* 33, 4 (2014), 59:1–59:12.

C. Jang, C.-K. Lee, J. Jeong, G. Li, S. Lee, J. Yeom, K. Hong, and B. Lee. 2016. Recent progress in see-through three-dimensional displays using holographic optical elements. *Appl. Opt.* 55, 3 (Jan. 2016), A71–A85.

D.-W. Kim, Y.-M. Kwon, Q.-H. Park, and S.-K. Kim. 2011. Analysis of a head-mounted display-type multifocus display system using a laser scanning method. *Opt. Eng.* 50, 3 (2011), 034006.

Joel Kollin. 1993. A retinal display for virtual-environment applications. *SID International Symposium Digest of Technical Papers* 24 (1993), 827.

Bernard Kress and Thad Starner. 2013. A review of head-mounted displays (HMD) technologies and applications for consumer electronics. *In Proceedings of SPIE Defense, Security, and Sensing, International Society for Optics and Photonics* 87200A (2013).

M. Lambooij, Marten. Fortuin, Ingrid Heynderickx, and Wijnand IJsselsteijn. 2009. Visual discomfort and visual fatigue of stereoscopic displays: a review. *J. Imaging Sci. Technol.* 53, 3 (2009), 30201–1.

Seungjae Lee, Changwon Jang, Seokil Moon, Jaebum Cho, and Byoungho Lee. 2016. Additive Light Field displays: Realization of Augmented Reality with Holographic Optical Elements. *ACM Trans. Graph.* 35, 4 (July 2016), 60.

Laure Leroy, Philippe Fuchs, and Guillaume Moreau. 2012. Real-time adaptive blur for reducing eye strain in stereoscopic displays. *ACM Transactions on Applied Perception (TAP)* 9, 2, Article 9 (June 2012).

Sheng Liu and Hong Hua. 2010a. A novel prototype for an optical see-through head-mounted display with addressable focus cues. *IEEE Transactions on Visualization and Computer Graphics* 16, 3 (2010), 381–393.

Sheng Liu and Hong Hua. 2010b. A systematic method for designing depth-fused multi-focal plane three-dimensional displays. *Opt. Express* 18, 11 (2010), 11562–11573.

Andrew Maimone, Andreas Georgiou, and Joel S. Kollin. 2017. Holographic Near-Eye Displays for Virtual and Augmented Reality. *ACM Trans. Graph.* 36, 4 (July 2017), 85:1–85:16.

Steven Mathews and Philip B. Kruger. 1994. Spatiotemporal transfer function of human accommodation. *Vision Research* 34, 15 (Aug. 1994), 1965–1980.

S.C. McQuaide, E. J. Seibel, J. P. Kelly, B. T. Schowengerdt, and T. A. Furness. 2013. A retinal scanning display system that produces multiple focal planes with a deformable membrane mirror. *Displays* 24 (2013), 65–72.

R. Michael, O. Guevara, M. de la Paz, J. A. de Toledo, and R. I. Barraquer. 2011. Neural contrast sensitivity calculated from measured total contrast sensitivity and modulation transfer function. *Acta ophthalmologica* 89, 3 (2011), 278–283.

R. Narain, R. A. Albert, A. Bulbul, G. J. Ward, M. S. Banks, and J. F. O'Brien. 2015. Optimal presentation of imagery with focus cues on multi-plane displays. *ACM Trans. Graph.* 34, 4 (Aug. 2015), 59.

N. A. Polak and R. Jones. 1990. Dynamic interactions between accommodation and convergence. *IEEE Trans on Biomed Eng.* 37, 10 (Oct. 1990), 1011–1014.

Sowmya Ravikumar, Kurt Akeley, and Martin S. Banks. 2011. Creating effective focus cues in multi-plane 3D displays. *Opt. Express* 19, 21 (Oct. 2011), 20940–20952.

Bahaa E. A. Saleh and Malvin C. Teich. 1991. *Fundamentals of Photonics*. New York: Wiley.

Yasuhiro Takaki and Nichiyo Nago. 2010. Multi-projection of lenticular displays to construct a 256-view super multi-view display. *Opt. Express* 18, 9 (April 2010), 8824–8835.

Yasuhiro Takaki and Yuta Yamaguchi. 2015. Flat-panel see-through three-dimensional display based on integral imaging. *Opt. Lett* 40, 8 (April 2015), 1873–1876.

Marc von Waldkirch, Paul Lukowicz, and Gerhard Tröster. 2003. Defocusing simulations on a retinal scanning display for quasi accommodation free viewing. *Opt. Express* 11, 24 (2003), 3220.

M. Watanabe, T Haruhisa, A. Nobuaki, M Riki, and Y. Shoji. 2003. A retinal scanning display with a wavefront curvature modulator. *Journal of the Society for Information Display* 11, 3 (2003), 511–515.

Gerald Westheimer. 1966. The maxwellian view. *Vision Research* 6, 11-12 (Dec. 1966), 669–682.

G. Wetzstein, D. Lanman, W. Heidrich, and R. Raskar. 2011. Layered 3d: Tomographic image synthesis for attenuation-based light eld and high dynamic range displays. *ACM Trans. Graph.* 30, 4 (July 2011).

H. J. Yeom, H. J. Kim, S. B. Kim, H. Zhang, B. Li, Y. M. Ji, S. H. Kim, and J. H. Park. 2015. 3D holographic head mounted display using holographic optical elements with astigmatism aberration compensation. *Opt. Express* 23, 25 (Dec. 2015), 32025–32034.

Mo Zohrabi, Robert H. Cormack, and Juliet T. Gopinath. 2016. Wide-angle nonmechanical beam steering using liquid lenses. *Opt. Express* 24, 21 (2016), 23798–23809.