



# Foveated light-field display and real-time rendering for virtual reality

CHEN GAO,<sup>1</sup>  YIFAN PENG,<sup>2</sup>  RUI WANG,<sup>3</sup> ZHUOPENG ZHANG,<sup>4</sup> HAIFENG LI,<sup>1,\*</sup> AND XU LIU<sup>1</sup>

<sup>1</sup>College of Optical Science and Engineering, Zhejiang University, 38 Zheda Rd., Hangzhou 310027, China

<sup>2</sup>Department of Electrical Engineering, Stanford University, 350 Jane Stanford Way, Stanford, California 94305-9510, USA

<sup>3</sup>College of Computer Science and Technology, Zhejiang University, 38 Zheda Rd., Hangzhou 310027, China

<sup>4</sup>Hangzhou Guangli Technology Co., Ltd., 238 Tianmushan Rd., Hangzhou 310012, China

\*Corresponding author: lihaifeng@zju.edu.cn

Received 1 June 2021; revised 2 August 2021; accepted 2 August 2021; posted 2 August 2021 (Doc. ID 432911); published 22 September 2021

Glasses-free light field displays have significantly progressed due to advances in high-resolution microdisplays and high-end graphics processing units (GPUs). However, for near-eye light-field displays requiring portability, the fundamental trade-off regarding achieved spatial resolution remains: retinal blur quality must be degraded; otherwise, computational consumption increases. This has prevented synthesizing the high-quality light field from being fast. By integrating off-the-shelf gaze tracking modules into near-eye light-field displays, we present wearable virtual reality prototypes supporting human visual system-oriented focus cues. An optimized, foveated light field is delivered to each eye subject to the gaze point, providing more natural visual experiences than state-of-the-art solutions. Importantly, the factorization runtime can be immensely reduced, since the image resolution is only high within the gaze cone. In addition, we demonstrate significant improvements in computation and retinal blur quality over counterpart near-eye displays. © 2021 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

<https://doi.org/10.1364/AO.432911>

## 1. INTRODUCTION

Near-eye displays for extended reality (XR) including augmented reality (AR), virtual reality (VR) and mixed reality (MR) are the promising interface for the next-generation personal computing scenario because of their ability to provide unprecedented immersion and interaction. For near-eye displays, vergence and accommodation are the essential depth cues provided by the physiological response of binocular. While vergence cues are enabled by the angle between the optical axis of both eyes, each eye lens itself provides accommodation cues. State-of-the-art near-eye VR displays mostly support only vergence cues, causing the vergence-accommodation conflict (VAC) [1].

Existing accommodation-supported near-eye display solutions can be sorted into several types: varifocal, Maxwellian view, multifocal (volumetric), holographic, integral imaging, and light-field displays. All of them show advantages as well as limitations. Varifocal [2,3] and Maxwellian-view displays [4] are lightweight in both footprint and computation but require simulating accurate retinal blur images. Multifocal displays passively deliver multiple depths of a 3D scene by multiplexing space, time, polarization, or wavelength, at the expense of form factor, refresh rate, depth range, or color gamut [5,6]. Holographic displays can already generate high-quality 2D holograms [7,8] but are still in their infancy regarding

delivering high-quality dynamic 3D scenes [9]. It is possible to generate holograms based on the light-field approach to enable 3D holographic displays [10,11]. This light-field-based computer-generated holograms (CGH) algorithm's computation complexity is light-field synthesis plus a phase-retrieval method. Integral imaging displays are simple in system configuration but suffer from the trade-off between angular and spatial resolution due to the limited lens per inch (LPI) [12,13]. As such, a near-eye display mechanism with a sizeable eye box, natural depth cues, high spatial resolution, and low computational cost is still in demand.

Compared with the above methods, with simply stacking scattering optical elements and applying the content-adaptive optimization, near-eye light-field displays can reconstruct high-resolution light fields within a specific depth range and an appropriate eye box [14–19]. However, without multiplexing, the challenge of synthesizing high-quality retinal blur images while simultaneously achieving real-time performance remains intractable, presenting a dominating obstruction for making layered light-field displays a practical near-eye display solution.

To this end, in the computer graphics and computational displays communities, foveated rendering [20] has been widely applied to accelerate the computation by reassigning pixels to match the photoreceptor density distribution at the retina without sacrificing the viewing experience. The insight behind

foveated rendering is that the human eye has a pit, called the fovea centralis, in the center of the retina's macula lutea. The fovea centralis is composed of closely packed cone cells and is responsible for the sharpest central vision. Although the recent research work developed a human-vision-based algorithm that could accelerate the light-field reconstruction [21], this acceleration algorithm's practical implementation and comprehensive assessment are still lacking.

This work explores a family of foveated near-eye light-field displays and rendering that integrates the eye-tracking technique into the layered display scheme. As such, this novel combination addresses the following technical insights of enabling the practical display solution for emerging VR applications. The proposed foveated rendering pipeline supports near-correct retinal blur with both foveated and peripheral vision. By leveraging adaptive sampling densities at different visual eccentricities, the computational cost is significantly reduced. The retinal image reconstruction of multiplicative displays via accumulation buffer and blending function enables the simulation for accommodation cues. Moreover, in simulation and experiments, we have comprehensively assessed the retinal blur quality and the computation load between foveated and uniform rendering methods under different parameters.

## 2. RELATED WORK

This work expands from conventional near-eye light-field displays to a variant that exploits the nonuniform visual acuity of the human visual system (HVS). In the following, we review relevant literature across compressive light-field displays and foveated computational displays.

### A. Compressive Light-Field Displays

Compressive light-field displays with multilayer structures have been investigated to support vergence and accommodation cues. Based on the light-field synthesis mechanism, compressive light-field displays can be classified into three types: multiplicative, additive, and hybrid [14–19,22]. Multiplicative light-field displays were first proposed to support vergence cues for glasses-free 3D displays [22] and later applied in near-eye displays to mitigate the VAC [14]. The diffraction effect essentially limits their resolution due to the fact of stacking liquid crystal display (LCD) panels. Therefore, additive light-field displays became prevalent by accumulating light intensity [15–17] of addressable projected pixels. Optical elements realizing additive light-field displays include holographic optical elements (HOEs) [15], scattering polarizers [16], and Pancharatnam–Berry phase lenses [17]. Hybrid light-field displays combine multiplicative and additive manners by delivering more multiplicative layers with time multiplexing [18] or space multiplexing [19], which can extend the depth of field (DoF) of compressive light-field displays [23]. However, adding more layers makes the display system bulky and increases the computational cost exponentially. Seeking a more practical way of overcoming the display hardware bandwidth constraint is of significance.

### B. Foveated Computational Displays

Verified unperceivable in VR [24], foveated rendering permits further optimization for near-eye displays. We address three core

motivations in the optimization of computational displays with gaze-contingent rendering: to avoid image quality degradation by incorporating the pose of the eye [25,26]; to expand the limited field of view of the display system by surrounding the central accommodation-support vision with accommodation-lack displays [27–31], and to reduce computation cost for real-time rendering [32,33]. Accordingly, foveated rendering methods for varifocal [34], Maxwellian view [27], multifocal [35], holography [32], integral imaging [29], and their hybrids [28,33] have been reported.

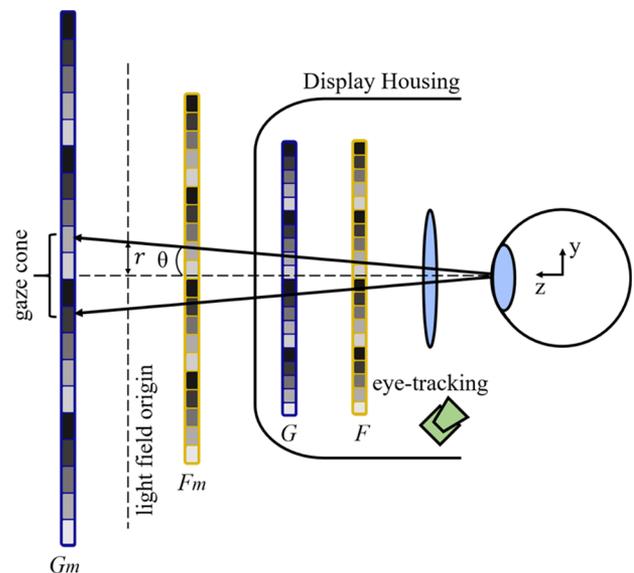
For compressive light-field displays, a table-top prototype was built to show that the viewing-position-dependent weight optimization delivers a reconstructed light field of a higher peak signal-to-noise ratio (PSNR) when the region of interest than uniform optimization [36]. Recently, Sun *et al.* have conducted two psychophysical experiments and found that both the blur and depth discrimination thresholds increase monotonically along with the visual eccentricity [37]. To our best knowledge, we are not aware of any prior work that realizes dynamically foveated near-eye light-field displays.

## 3. PRINCIPLE

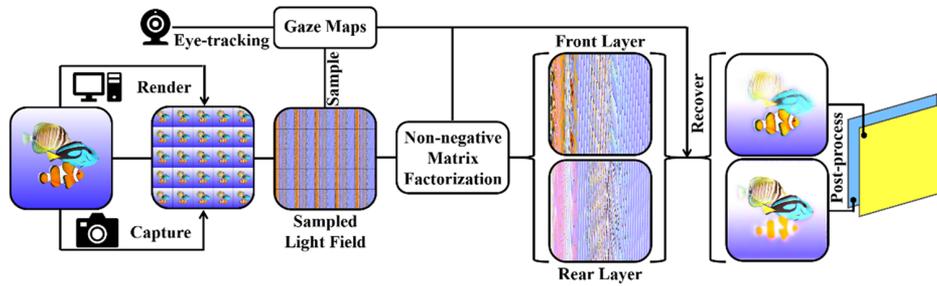
We elaborate on the basic principle of synthesizing a foveated light field in a near-eye layered display system in Fig. 1. By importing two-plane parameterization, the reconstructed light field  $L$  can be expressed as the Hadamard (element-wise) product of the dual-layer patterns as below,

$$L = F_m \circ G_m, \quad (1)$$

where  $F_m$  and  $G_m$  are the virtual images of front LCD  $F$  and rear LCD  $G$ , respectively.  $\circ$  is the Hadamard product. The optimization objective is to minimize the weighted variance between the target light field  $L_t$  and reconstructed light field  $L$ .



**Fig. 1.** Schematic diagram of foveated near-eye light-field displays. Two LCDs are mounted inside the display housing. The dual-layer LCDs modulate the backlight (not shown) in a multiplicative manner. A gaze cone is generated based on the gaze point acquired by the eye-tracking module to determine the dual-layer LCDs' foveated rendering area. A foveated light field is reconstructed by dual-layer patterns when observed through a magnifying lens.



**Fig. 2.** Foveated rendering pipeline for near-eye light-field displays. The target light field is rendered, then sampled according to foveated weight matrix. Foveated dual-layer patterns are obtained by applying nonnegative matrix factorization to the sampled target light field. These patterns are recovered, postprocessed, and then displayed using the dual-layer near-eye display, emitting a reconstruction of the foveated target light field.

That is,

$$\min_{\arg} \|W^{\circ}(L_t - L)\|^2. \quad (2)$$

The weight matrix  $W$ 's value denotes the density and intensity of rays intersecting at dual layer. The value 1 indicates the corresponding ray has the highest density and intensity, while the value 0 indicates no rays. In our case, a gaze cone of a visual eccentricity angle  $\theta$  and a radius of  $r$  is generated based on the gaze point to determine the dual-layer LCDs' foveated rendering region. A foveated weight matrix is calculated by setting different sampling steps within and beyond the gaze cone.

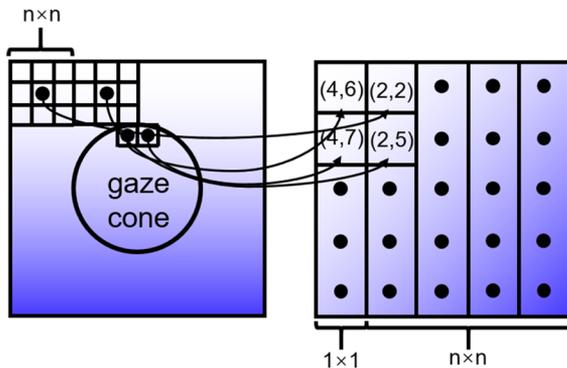
We also propose a foveated rendering pipeline based on the above display system (Fig. 2). The pipeline includes updating the gaze point from the eye-tracker, capturing and sampling the target light field, factorizing the sampled target light field into dual-layer patterns, and recovering and postprocessing the factorized dual-layer patterns.

### A. Target Light-Field Sampling

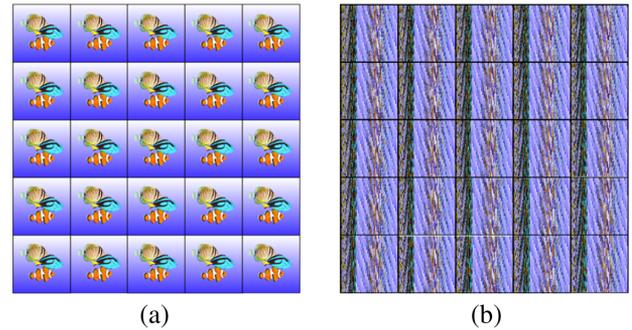
Figure 3 shows the generation of the foveated weight matrix. We assume that the initial gaze cone is at the display area's center. The radius of the gaze cone  $r$  can be calculated by Eq. (3),

$$r = d_{g \rightarrow V_0} \tan \theta, \quad (3)$$

where  $d_{g \rightarrow V_0}$  represents the distance between the pupil center  $V_0$  and the gaze point  $g$  on each layer. The pixel index inside the gaze cone is stored first, and that outside the gaze cone is stored

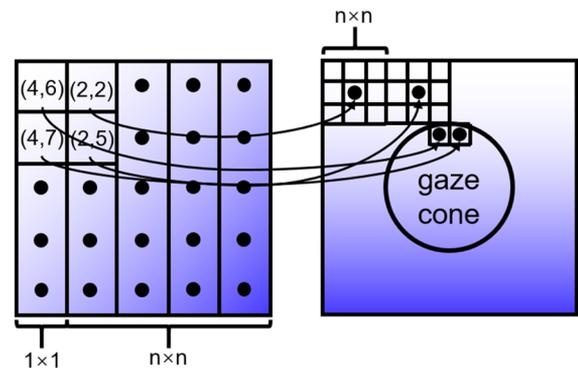


**Fig. 3.** Generation of the foveated weight matrix. The pixel index inside the gaze cone is stored first, and that outside the gaze cone is stored later, column by column.

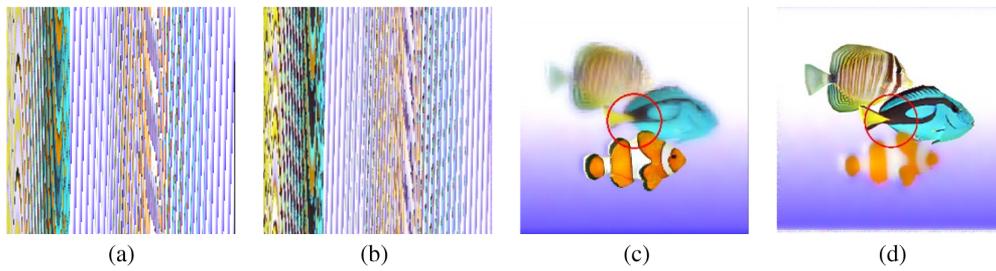


**Fig. 4.** Target light field (a) before and (b) after sampling by the foveated weight matrix when the gaze cone is on the center. After sampling, the target light field's resolution beyond the gaze cone periphery is significantly reduced (the sampled target light field is resized for visualization). Sampled target light field is stored as a texture array in the formation of the foveated weight matrix.

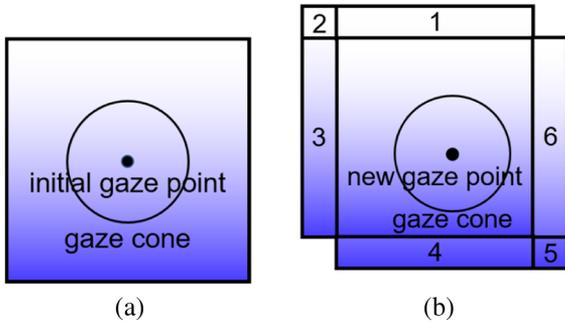
later, column by column. We set an  $n \times n$  grid to sample the area out of the gaze cone. The value of  $n$  should be larger than one to achieve foveated rendering. After being sampled by the foveated weight matrix, the target light field's resolution in the periphery is significantly reduced. Figure 4 shows one target light of a tropical fish scene before and after sampling when the gaze cone is in the center.



**Fig. 5.** Remapping sampled dual-layer pixels to their physical positions to do ray tracing during iterations. The mapping relationship is stored as a texture to utilize GPU acceleration.



**Fig. 6.** Factorized patterns are in the formation of the foveated weight matrix and need to be recovered by the compressed index map. (a)–(b) Factorized and (c)–(d) recovered dual-layer patterns when the gaze cone is in the center.



**Fig. 7.** (a) Initial gaze maps when gazed at the center and (b) updated gaze maps when gazed at the right bottom. The updated foveated weight matrix and the compressed index map could be obtained from the difference with the initial ones.

$$F_C = F_C \circ \frac{\sum_i^V L_C^i \circ G_C}{\sum_i^V (F_C \circ G_C) \circ G_C},$$

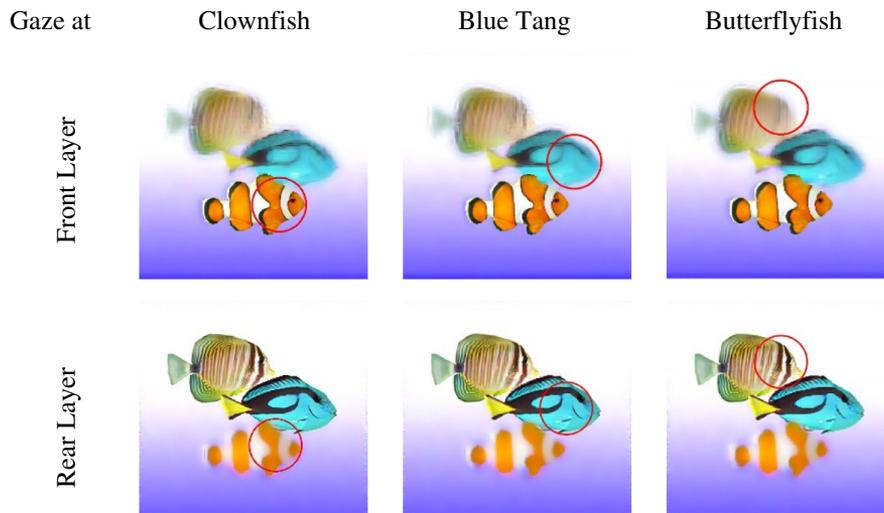
$$G_C = G_C \circ \frac{\sum_i^V L_C^i \circ F_C}{\sum_i^V (F_C \circ G_C) \circ F_C}, \quad (4)$$

where  $\circ$  denotes Hadamard product,  $F_C$  and  $G_C$  are factorized dual-layer patterns.  $L_C^i$  is  $i$ th sampled target light field, and  $V$  is the number of viewpoints in the sampled target light field. Because the sampled target light field only contains the sampled information, its pixels' indices have nothing to do with their physical positions. The key modification is that a look-up texture called a compressed index map is precalculated during iterations to retrieve the pixel indices where sampled light rays intersect at the dual layer. Figure 5 shows the process of remapping sampled dual-layer pixels to their physical positions during iterations.

After several times of iteration,  $F_C$  and  $G_C$  would converge to almost unchanged patterns. We set the iteration number of times as five, since the optimized patterns do not show evident improvement after five iterations in most cases. The compressed index map is used again to recover  $F_C$  and  $G_C$  once iterations are finished. Figure 6 shows the factorized and recovered dual-layer patterns when the gaze cone is in the center after five iterations. After recovering, the spatial resolution beyond the gaze cone periphery (the red circle) has been reduced.

### B. Factorization of Sampled Light Field

We implement a modified nonnegative matrix update rules [38] in OpenGL shading language (GLSL) as Eq. (4),



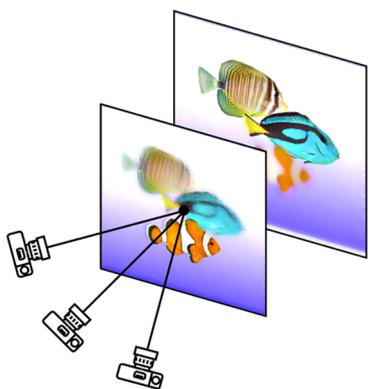
**Fig. 8.** Dual-layer patterns when shifting foveated rendering region. The spatial resolution beyond the gaze cone periphery (the red circle) reduces.

### C. Shifted Foveated Rendering

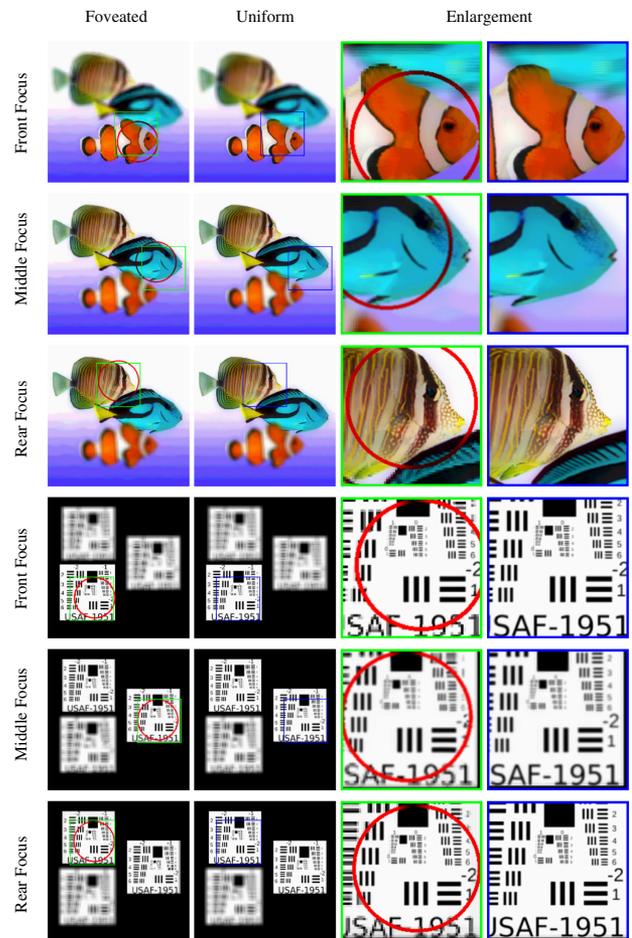
The foveated rendering technique aims to display a high-resolution image only in the fovea region. Notice that generating these two gaze maps during every frame requires time to transfer data between the central processing unit (CPU) and graphics processing units (GPU), inevitably causing rendering latency. To adapt to the gaze point change of the eye without uploading gaze maps to the GPU every frame, we have developed a method to shift the foveated rendering area only with the initial gaze maps. Specifically, we utilize the display area's symmetry to update the gaze cone. For example, as shown in Fig. 7, when the gaze point moves toward the bottom right, the pixel's coordinates in areas 4, 5, and 6 out of initial gaze maps could be obtained from the left-up areas 1, 2, and 3 of initial gaze maps. Figure 8 contains recovered dual-layer patterns when the gaze cone is shifted from the front clownfish, the middle blue tang, and the rear butterflyfish. Here we set the gaze cone angle  $\theta = 5^\circ$ , and the sampling grid outside the gaze cone is  $3 \times 3$ .

## 4. SIMULATION

We validated our method by simulating the retinal blur image (i.e., DoF effect) using an accumulation buffer [39] and the multiplicative blending function. An accumulation buffer is one method that can simulate a realistic DoF effect, and it is very easy to execute. Figure 9 shows the principle of our simulation. A virtual 3D scene includes a background, two display layers, and a camera that simulates the human eye. The scene's background is set to white to simulate a uniform backlight. The blending function of these two layers is set multiplicative so that the light from the background passes through these two layers successively in a multiplicative manner. A reconstructed light field of one viewpoint is captured with the camera, just as Eq. (1). Here we apply an ideal camera model with no optical aberration. The simulation results are obtained by setting every viewpoint's focal plane the same, accumulating the captured images together, and dividing the accumulated image by accumulation time (equal to total viewpoint number). Theoretically, the simulated DoF appears more closely to realistic over more accumulation time. However, limited by the numerical precision of commercial graphics cards, the simulated DoF result tends to saturate dimly, being divided by large numbers. This simulation method's



**Fig. 9.** Simulation principle of the DoF effect. The simulation results are obtained by setting every viewpoint's focal plane the same, accumulating the captured images together, and dividing the accumulated image by accumulation time (equal to total viewpoint number).



**Fig. 10.** Simulated retinal blur images reconstructed for near-eye light-field displays of foveated rendering and uniform rendering with three different focus states of the eye (front focus, middle focus, and rear focus). Enlargements are shown on the right side of the figure.

drawback is that it cannot simulate the diffusion and diffraction light rays of pixels, partly causing deviation from experimental results (Section 5.D). The reconstruction results with a resolution of  $2048 \times 2048$  after 25 times of accumulation are shown in Fig. 10. We used the factorized dual-layer patterns in Fig. 8 for simulation. A scene of USAF-1951 resolution charts is also demonstrated to clearly show the performance. As expected, the simulation results show that a high-resolution image is restored in the foveal region. In contrast, the images in the peripheral area exhibit a low resolution compared to the previous uniform rendering method [14].

## 5. EXPERIMENT

### A. Hardware Implementation

As shown in Fig. 11, we build one monocular and one binocular prototype with an eye relief of 20 mm to verify the proposed foveated rendering method. These prototypes use dual-layer LCD panels (Sharp LS060R1SX01), and the magnifying eyepieces are biconvex lenses with a focal length of 50 mm and a diameter of 50 mm. For the binocular prototype, two lenses are separated by 64 mm. The physical LCD panels are placed at 38 and 48 mm, whereas the virtual LCD panels are imaged at 158 and 1200 mm, respectively, from the lens. The eye-tracking

modules mounted in monocular and binocular HMDs are 7invensun aGlass I and Droolon F1, respectively. The latency of these eye-trackers is within 10 ms, which is much less than the duration of fixation (200 ms). The display housings are modified from the supplement of Huang *et al.* [14] and fabricated with 3D printing and 3M DP460 epoxy adhesive. Experimental results are captured with a charge-coupled device (CCD) camera (Imaging Source DFK 33UX250) with a resolution of  $2448 \times 2048$ .

**B. Software Implementation**

All coding is implemented in C++ and OpenGL. Rendering is done with OpenGL, and the iterations are executed in GLSL. As a result, the total time of rendering, iterations, postprocessing, and displaying is short enough to drive the LCD panels at their fastest refresh rate (60 Hz) for real-time frame rates. The redundant time allows us to seed every frame’s initial state with random values to avoid motion blur [14].

**C. Calibration**

The recovered dual-layer patterns should be further processed to compensate for deviations between the ideal geometrical optics model and the actual display system. These deviations include the orthogonality of polarizers, the misalignment between display panels, and the distortion of lenses. Since LCD polarizers’ polarization states are orthogonal, one of the dual-layer patterns needs to be flipped vertically. Huang *et al.* [14] compensated for the lens aberrations by predistorting the target light field applying the transformation below,

$$L_{\varphi}(x, y) = L(x + x_d, y + y_d), \tag{5}$$

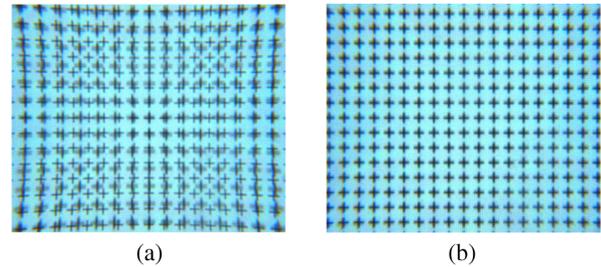
$$x_d = x(k_{0x} + k_{1x}r^2 + k_{2x}r^4), y_d = y(k_{0y} + k_{1y}r^2 + k_{2y}r^4), \tag{6}$$

where  $L_{\varphi}(x, y)$  and  $L(x + x_d, y + y_d)$  are the predistorted and undistorted target light field;  $r^2 = x^2 + y^2$ ,  $(x, y)$  and  $(x_d, y_d)$  are ideal and distorted point coordinates on the image plane of the lens with the lens center as origin, respectively; and  $\{k_{0x}, k_{1x}, k_{2x}\}$  and  $\{k_{0y}, k_{1y}, k_{2y}\}$  are lens-specific distortion coefficients along the  $x$  and  $y$  direction. However, this method cannot be applied in our pipeline. Since predistorting target light field is a pixel-wise operation, pixels inside one sampling grid share the same distortion coefficients, deviating from the actual distortion. We observe that with careful assembly, the radial distortion could be compensated for by separately performing predistortion for each layer. The misalignment between display panels and the eyepiece’s distortion is compensated for by interactively adjusting the alignment offset and distortion coefficients until perspective images of a crosshair array on each layer are coincident. Concrete calibration steps are below:

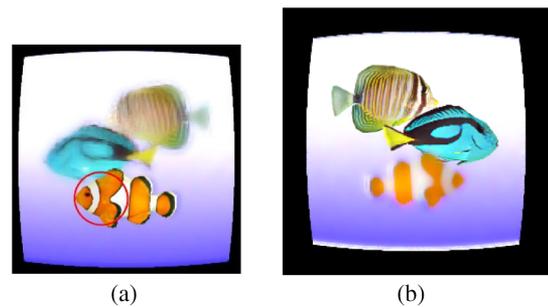
1. Adjust the translation stage of the CCD to coincide with the optical axis of the CCD and lens.
2. Dual-layer images of the crosshair array should be resized for display due to the different magnification factors caused by the panels’ lateral placement.
3. Set the CCD’s F-number to the maximum to obtain the largest DoF. For observation, the gamma value for the



**Fig. 11.** (a) Monocular and (b) binocular prototypes of the foveated display. These prototypes comprise two stacked LCDs and a commercial eye-tracking module, respectively.



**Fig. 12.** (a) Without calibration, we observe strong superposed pin-cushion distortions of dual-layer. (b) After appropriate flipping, predistortion, and alignment, perspective images of a crosshair array on each layer are visually coincident.



**Fig. 13.** (a) Front and (b) rear layer display patterns after flipping and predistorting when gazed at the front’s clownfish. Note that the distortion coefficients (including the scale factors) of dual-layer patterns are visually different, since their distances to the lens differ.

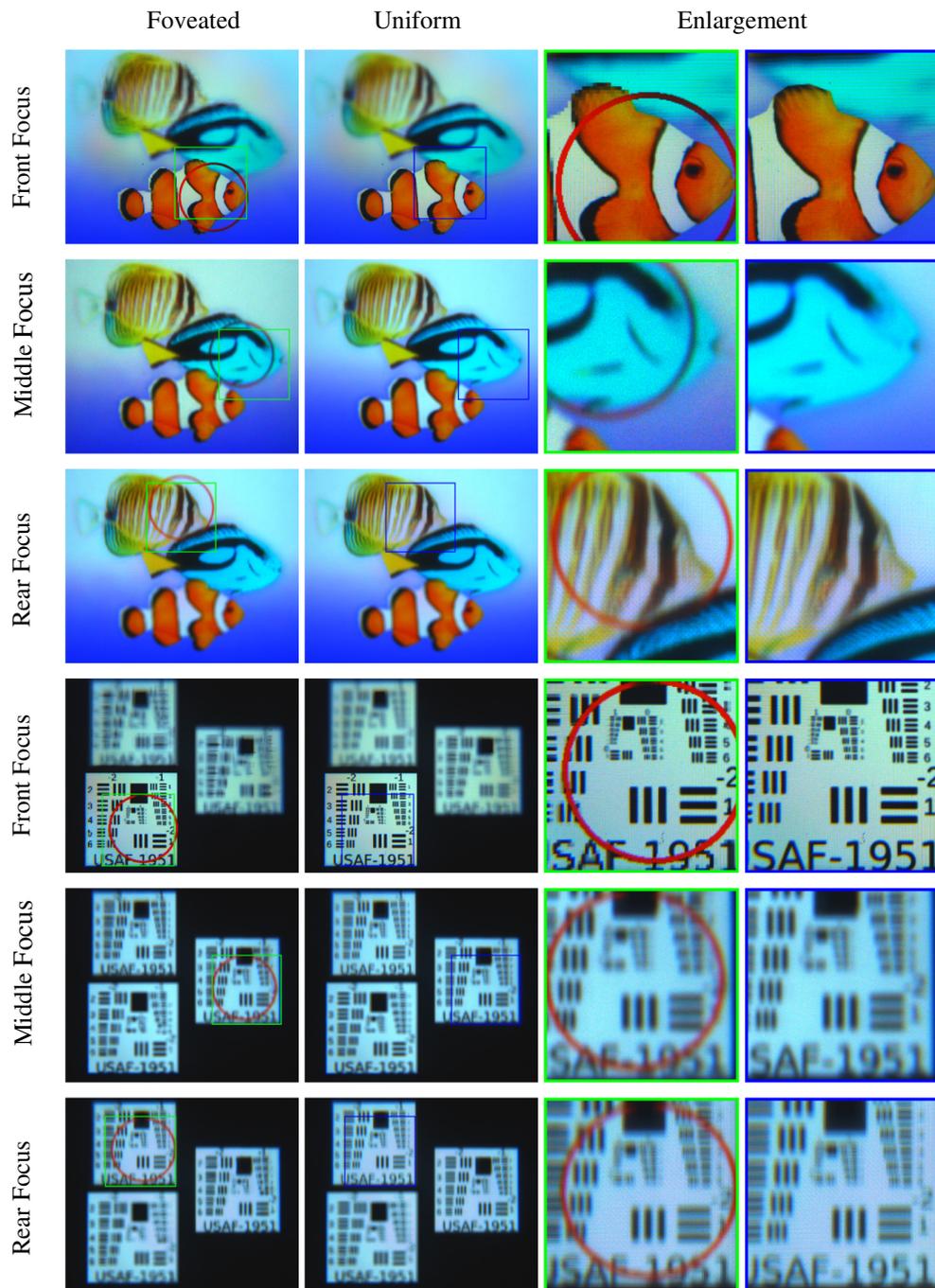
panels was measured as 1.0, the gain was set at 25.00 dB, and the exposure time was set at 1/4 s.

4. Accommodate the CCD’s focal plane until the crosshair array images of dual-layer are both clear.
5. Adjust the alignment offset and distortion coefficients until perspective images of a crosshair array on each layer are coincident.

Figure 12 shows perspective images of a crosshair array on each layer before and after flipping, predistortion, and alignment. We perform these transformations on the dual-layer patterns after recovery (Fig. 13).

**D. Results**

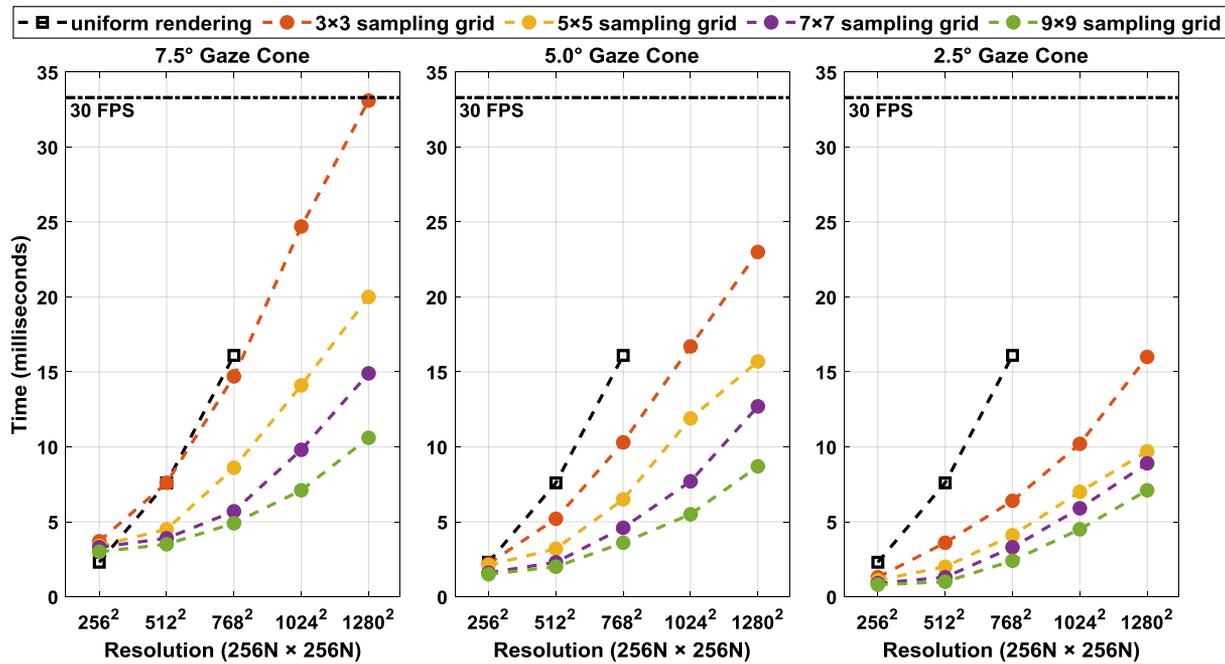
We used the same factorized dual-layer patterns in experiments as those in simulation. To simulate a 4.00 mm pupil diameter of the human eye in the dark, we mount a lens



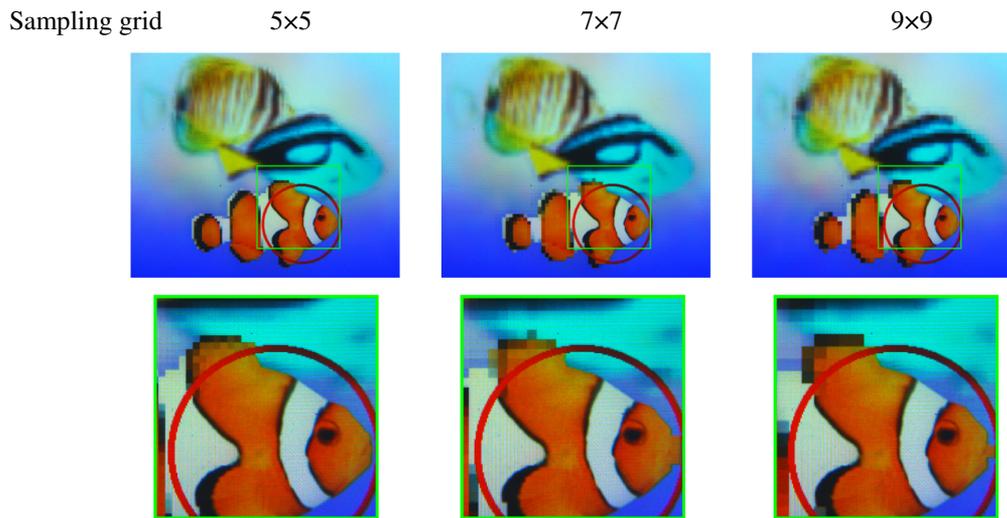
**Fig. 14.** Experimental retinal blur images reconstructed for near-eye light-field displays of foveated rendering and uniform rendering with three different CCD focus states (front focus, middle focus, and rear focus). Enlargements are shown on the right side of the figure. The experiment figures' size is larger than simulation's because we put the camera close to the magnifying eyepiece to avoid capturing the optical frame, with a higher resolution to capture the whole field of view. Videos for the continuous change of the camera's focus position are shown in [Visualization 1](#) and [Visualization 2](#).

with a focal length of 16 mm on the CCD and set the F-number to 4. Figure 14 shows the display photographs captured with the CCD camera. These photographs were captured with an exposure time of 1/60 s, a gain of 36.50 dB, and a gamma of 1.00. By adjusting the focal plane of the camera, accommodation cues with visual acuity fall-off were observed. Only the part of the tropical fish and resolution chart, both at the camera's focal

plane and within the gaze cone, is well-focused, while other parts are blurred. Figure 14 also shows the experimental comparison between the foveated rendering and uniform rendering. The enlargements show the expected resolution difference in the foveal and peripheral regions.



**Fig. 15.** Performance comparison of our foveated rendering method under different sampling set with uniform rendering method. When uniform rendering is implemented under reconstructed resolution larger than  $1024 \times 1024$ , the iteration time becomes unacceptable for real-time display.



**Fig. 16.** Experimental retinal images reconstructed for near-eye light-field displays of foveated rendering when different sampling grids are applied (front focus). Enlargements are shown on the bottom side of the figure.

## 6. DISCUSSION

### A. Computational Performance

We have tested our algorithm for the tropical fish scene with different gaze cone angles and sampling grids under several reconstructed resolutions ( $256N \times 256N$ ,  $N = 1, 2, 3, 4, 5$ ) to evaluate the computation time (Fig. 15). The gaze cone angles are  $2.5^\circ$  (equal to the foveal portion of the retina),  $5^\circ$  (slightly larger than the paracentral portion), and  $7.5^\circ$  (close to the macular portion). The sampling grids are  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ , and  $9 \times 9$ . For a fair comparison, we have also assessed the computation cost for the conventional method where the full resolution light field is rendered, which is implemented for five iterations

on a PC with an Intel Core i7-4790 CPU, 16 GB RAM, and an NVIDIA GeForce GTX 1050Ti GPU.

We note that under low reconstructed resolutions, our method is slightly slower than the uniform rendering due to the extra sampling procedure. However, the computation time that our method saves compared with the earlier study overgrows when the reconstructed resolution increases. For example, the calculation time is 16.1 ms in the conventional method for  $768 \times 768$  image pixels. In comparison, under a  $3 \times 3$  sampling grid, the time in the proposed method is 14.7, 10.3, and 6.4 ms for gaze cone angles with  $7.5^\circ$ ,  $5.0^\circ$ , and  $2.5^\circ$ , respectively. Furthermore, our approach can support higher reconstructed resolutions (e.g., above  $1024 \times 1024$ ) that might overload

GPU usage when applying the conventional uniform rendering scheme.

## B. Diffraction Effect

We observe that the foveation of the rear-focus image is not as sharp as that of the front-focus image due to the diffraction effect of the front layer panel. This degradation could be partially compensated for by acquiring the point spread functions (PSFs) of a single pixel on both the front and rear screens, modeling the diffraction as convolution with those PSFs, and incorporating them into the factorization algorithm [40]. In addition, the diffraction effect could be significantly eliminated with layered diffuser HOEs [15]. Our insight into using gaze maps can be modified and implemented in additive and hybrid light-field displays to achieve a diffraction-free compressive light-field display with low computation.

## C. Aliasing Artifact

We observe tiny aliasing artifacts on the edges due to the rectangular corners of the sampling grid, and it becomes more noticeable when a larger sampling grid is applied (see Fig. 16). The transition between the foveal and peripheral display regions could be feathered using a Gaussian mask to eliminate the resolution discontinuity [27]. A quantity merit function to evaluate and choose the level of the artifact based on human eyes' resolution in peripheral vision [41] is described below. The resolution  $R$  perceived by the retina should be larger than the corresponding eccentricity's visual acuity (VA), i.e.,

$$R = \frac{d_{g \rightarrow v_0} \text{radians}(1^\circ)}{2D \times P} \geq \text{VA}, \quad (7)$$

where radians() is a function to convert a degree to its radian number,  $D$  is the sampling grid's one dimension,  $P$  is the magnifying pixel pitch, VA is measured in cycle per degree (cpd). The product of  $D$  and  $P$  denotes the perceived minimum half-period of a displayed rectangular grating. Evaluated by the above merit, our prototypes' hardware cannot reach the VA of eccentricity  $10^\circ$  ( $\sim 10$  cpd) with a  $3 \times 3$  sampling grid unless display panels of an  $11 \mu\text{m}$  pixel pitch are applied.

## 7. CONCLUSION

In summary, we have implemented an efficient foveated rendering scheme for near-eye light-field displays. Specifically, foveated dual-layer patterns are generated using the nonnegative matrix factorization with two gaze maps, and the foveated rendering region is updated in real-time. This method provides continuous focus cues and near-correct retinal blur as that of state-of-the-art research, while at the same time with much less computation load and natural foveated and peripheral vision that many prior systems lack. Two near-eye display prototypes are built to verify the proposed foveated rendering scheme. We envision our method significantly advance the practical use of near-eye light-field displays in real-time XR applications.

**Funding.** Hangzhou Leading Innovation and Entrepreneurship Team (std013); Zhejiang University Education Foundation Global Partnership Fund

(100000-11320); National Key Research and Development Program of China (2017YFB1002900).

**Acknowledgment.** The authors thank Gordon Wetzstein for the fruitful discussions.

**Disclosures.** The authors declare that there are no conflicts of interest related to this paper.

**Data Availability.** Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

## REFERENCES

1. D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," *J. Vis.* **8**(3), 33 (2008).
2. K. Akşit, W. Lopes, J. Kim, P. Shirley, and D. Luebke, "Near-eye varifocal augmented reality display using see-through screens," *ACM Trans. Graph.* **36**, 189 (2017).
3. X. Xia, Y. Guan, A. State, P. Chakravarthula, K. Rathinavel, T. J. Cham, and H. Fuchs, "Towards a switchable AR/VR near-eye display with accommodation-vergence and eyeglass prescription support," *IEEE Trans. Visual Comput. Graph.* **25**, 3114–3124 (2019).
4. T. Lin, T. Zhan, J. Zou, F. Fan, and S. T. Wu, "Maxwellian near-eye display with an expanded eyebox," *Opt. Express* **28**, 38616–38625 (2020).
5. Q. Chen, Z. Peng, Y. Li, S. Liu, P. Zhou, J. Gu, J. Lu, L. Yao, M. Wang, and Y. Su, "Multi-plane augmented reality display based on cholesteric liquid crystal reflective films," *Opt. Express* **27**, 12039–12047 (2019).
6. K. Rathinavel, H. Wang, A. Blate, and H. Fuchs, "An extended depth-at-field volumetric near-eye augmented reality display," *IEEE Trans. Visual. Comput. Graph.* **24**, 2857–2866 (2018).
7. P. Chakravarthula, E. Tseng, T. Srivastava, H. Fuchs, and F. Heide, "Learned hardware-in-the-loop phase retrieval for holographic near-eye displays," *ACM Trans. Graph.* **39**, 1–18 (2020).
8. S. Choi, J. Kim, Y. Peng, and G. Wetzstein, "Optimizing image quality for holographic near-eye displays with Michelson holography," *Optica* **8**, 143–146 (2021).
9. W. Song, X. Li, Y. Zheng, Y. Liu, and Y. Wang, "Full-color retinal-projection near-eye display using a multiplexing-encoding holographic method," *Opt. Express* **29**, 8098–8107 (2021).
10. N. Padmanaban, Y. Peng, and G. Wetzstein, "Holographic near-eye displays based on overlap-add stereograms," *ACM Trans. Graph.* **38**, 1–13 (2019).
11. Z. Wang, L. M. Zhu, X. Zhang, P. Dai, G. Q. Lv, Q. B. Feng, A. T. Wang, and H. Ming, "Computer-generated photorealistic hologram using ray-wavefront conversion based on the additive compressive light field approach," *Opt. Lett.* **45**, 615–618 (2020).
12. W. Song, Q. Cheng, P. Surman, Y. Liu, Y. Zheng, Z. Lin, and Y. Wang, "Design of a light-field near-eye display using random pinholes," *Opt. Express* **27**, 23763–23774 (2019).
13. J. Zhao, Q. Ma, J. Xia, J. Wu, B. Du, and H. Zhang, "Hybrid computational near-eye light field display," *IEEE Photon. J.* **11**, 1–10 (2019).
14. F. C. Huang, K. Chen, and G. Wetzstein, "The light field stereoscope: immersive computer graphics via factored near-eye light field displays with focus cues," *ACM Trans. Graph.* **34**, 60 (2015).
15. 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.* **35**, 60 (2016).
16. S. Moon, C. K. Lee, D. Lee, C. Jang, and B. Lee, "Layered display with accommodation cue using scattering polarizers," *IEEE J. Sel. Top. Signal Process.* **11**, 1223–1231 (2017).
17. T. Zhan, Y. H. Lee, and S. T. Wu, "High-resolution additive light field near-eye display by switchable Pancharatnam-Berry phase lenses," *Opt. Express* **26**, 4863–4872 (2018).
18. M. Liu, C. Lu, H. Li, and X. Liu, "Bifocal computational near eye light field displays and structure parameters determination scheme for bifocal computational display," *Opt. Express* **26**, 4060–4074 (2018).

19. D. Kim, S. Lee, S. Moon, J. Cho, Y. Jo, and B. Lee, "Hybrid multilayer displays providing accommodation cues," *Opt. Express* **26**, 17170–17184 (2018).
20. A. Patney, M. Salvi, J. Kim, A. Kaplanyan, C. Wyman, N. Benty, D. Luebke, and A. Lefohn, "Towards foveated rendering for gaze-tracked virtual reality," *ACM Trans. Graph.* **35**, 1–12 (2016).
21. M. Liu, C. Lu, H. Li, and X. Liu, "Near eye light field display based on human visual features," *Opt. Express* **25**, 9886–9900 (2017).
22. G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting," *ACM Trans. Graph.* **31**, 1–11 (2012).
23. M. Xu and H. Hua, "Systematic method for modeling and characterizing multilayer light field displays," *Opt. Express* **28**, 1014–1036 (2020).
24. C. F. Hsu, A. Chen, C. H. Hsu, C. Y. Huang, C. L. Lei, and K. T. Chen, "Is foveated rendering perceivable in virtual reality? Exploring the efficiency and consistency of quality assessment methods," in *Proceedings of the 25th ACM International Conference on Multimedia* (2017), pp. 55–63.
25. O. Mercier, Y. Sulai, K. Mackenzie, M. Zannoli, J. Hillis, D. Nowrouzezahrai, and D. Lanman, "Fast gaze-contingent optimal decompositions for multifocal displays," *ACM Trans. Graph.* **36**, 1–15 (2017).
26. S. Lee, J. Cho, B. Lee, Y. Jo, C. Jang, D. Kim, and B. Lee, "Foveated retinal optimization for see-through near-eye multi-layer displays," *IEEE Access* **6**, 2170–2180 (2017).
27. J. Kim, Y. Jeong, M. Stengel, K. Aksit, R. Albert, B. Boudaoud, T. Greer, W. Lopes, Z. Majercik, P. Shirley, J. Spjut, M. McGuire, and D. Luebke, "Foveated AR: dynamically-foveated augmented reality display," *ACM Trans. Graph.* **38**, 99 (2019).
28. J. S. Lee, Y. K. Kim, M. Y. Lee, and Y. H. Won, "Enhanced see-through near-eye display using time-division multiplexing of a Maxwellian-view and holographic display," *Opt. Express* **27**, 689–701 (2019).
29. G. Koo, D. Shin, J. C. Leong, and Y. H. Won, "Foveated high-resolution light-field system based on integral imaging for near-eye displays," *Proc. SPIE* **11304**, 1130417 (2020).
30. S. Lee, M. Wang, G. Li, L. Lu, Y. Sulai, C. Jang, and B. Silverstein, "Foveated near-eye display for mixed reality using liquid crystal photonics," *Sci. Rep.* **10**, 16127 (2020).
31. A. Cem, M. K. Hedili, E. Ulusoy, and H. Urey, "Foveated near-eye display using computational holography," *Sci. Rep.* **10**, 14905 (2020).
32. Y. G. Ju and J. H. Park, "Foveated computer-generated hologram and its progressive update using triangular mesh scene model for near-eye displays," *Opt. Express* **27**, 23725–23738 (2019).
33. C. Chang, W. Cui, and L. Gao, "Foveated holographic near-eye 3D display," *Opt. Express* **28**, 1345–1356 (2020).
34. N. Padmanaban, R. Konrad, T. Stramer, E. A. Cooper, and G. Wetzstein, "Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays," *Proc. Natl. Acad. Sci. USA* **114**, 2183–2188 (2017).
35. G. Tan, Y. H. Lee, T. Zhan, J. Yang, S. Liu, D. Zhao, and S. T. Wu, "Foveated imaging for near-eye displays," *Opt. Express* **26**, 25076–25085 (2018).
36. D. Chen, X. Sang, X. Yu, X. Zeng, S. Xie, and N. Guo, "Performance improvement of compressive light field display with the viewing-position-dependent weight distribution," *Opt. Express* **24**, 29781–29793 (2016).
37. Q. Sun, F. C. Huang, L. Y. Wei, D. Luebke, A. Kaufman, and J. Kim, "Eccentricity effects on blur and depth perception," *Opt. Express* **28**, 6734–6739 (2020).
38. X. Cao, Z. Geng, M. Zhang, and X. Zhang, "Load-balancing multi-LCD light field display," *Proc. SPIE* **9391**, 93910F (2015).
39. T. T. Yu, "Depth of field implementation with OpenGL," *J. Comput. Sci. Coll.* **20**, 136–146 (2004).
40. M. Hirsch, G. Wetzstein, and R. Raskar, "A compressive light field projection system," *ACM Trans. Graph.* **33**, 1–12 (2014).
41. L. Frisen and A. Glansholm, "Optical and neural resolution in peripheral vision," *Invest. Ophthalmol. Visual Sci.* **14**, 528–536 (1975).