



Hybrid multi-layer displays providing accommodation cues

DONGYEON KIM, SEUNGJAE LEE, SEOKIL MOON, JAEBUM CHO,
YOUNGJIN JO, AND BYOUNGHO LEE*

School of Electrical and Computer Engineering, Seoul National University, Gwanak-Gu Gwanakro 1, Seoul 08826, South Korea

*byoungho@snu.ac.kr
<http://oeqelab.snu.ac.kr>

Abstract: Hybrid multi-layer displays are proposed as the system combines additive light field (LF) displays and multiplicative LF displays. The system is implemented by integrating the multiplicative LF displays with a half mirror to expand the overall depth of field. The hybrid displays are advantageous in that the form factor is competitive with existing additive LF displays with 2 layers implemented by a half mirror and two panels, only half of brightness loss is experienced compared to multiplicative LF displays with 2 layers, and no time-division is required to provide images for multi-layer displays. The images for presentation planes are processed by light field factorization and optimized with the presented algorithm. Retinal images are reconstructed based on various accommodation states and display types to check the accommodation response and utilized to compare the proposed displays with existing displays. With ray tracing method, retinal images generated by the proposed displays can be obtained. To verify the feasibility of the system, a prototype of hybrid multi-layer displays was implemented and display photographs were captured with different accommodation states of camera. With the simulation results and experimental results, this system was confirmed to support accommodation cues in a range of 1.8 diopters.

© 2018 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

OCIS codes: (080.2740) Geometric optical design; (100.6890) Three-dimensional image processing; (330.7322) Visual optics, accommodation.

References and links

1. "Microsoft hololens," <https://www.microsoft.com/en-us/hololens> (2014). Accessed: 2018-04-23.
2. "Htc vive," <https://www.vive.com/us/product/vive-virtual-reality-system/> (2016). Accessed: 2018-04-23.
3. V. Krishnan, S. Phillips, and L. Stark, "Frequency analysis of accommodation, accommodative vergence and disparity vergence," *Vis. Res.* **13**, 1545–1554 (1973).
4. S. J. Watt, K. Akeley, M. O. Ernst, and M. S. Banks, "Focus cues affect perceived depth," *J. Vis.* **5**, 7–7 (2005).
5. K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks, "A stereo display prototype with multiple focal distances," *ACM Trans. Graph.* **23**, 804–813 (2004).
6. 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**, 33–33 (2008).
7. M. Emoto, T. Niida, and F. Okano, "Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television," *J. Disp. Technol.* **1**, 328 (2005).
8. M. Lambooij, M. Fortuin, I. Heynderickx, and W. IJsselsteijn, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *J. Imaging Sci. Technol.* **53**, 30201–1 (2009).
9. T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, "The zone of comfort: Predicting visual discomfort with stereo displays," *J. Vis.* **11**, 11–11 (2011).
10. H.-J. Yeom, H.-J. Kim, S.-B. Kim, H. Zhang, B. Li, Y.-M. Ji, S.-H. Kim, and J.-H. Park, "3D holographic head mounted display using holographic optical elements with astigmatism aberration compensation," *Opt. Express* **23**, 32025–32034 (2015).
11. J.-H. Park, "Recent progress in computer-generated holography for three-dimensional scenes," *J. Inf. Disp.* **18**, 1–12 (2017).
12. Y. Takaki, K. Tanaka, and J. Nakamura, "Super multi-view display with a lower resolution flat-panel display," *Opt. Express* **19**, 4129–4139 (2011).
13. C. Jang, K. Bang, S. Moon, J. Kim, S. Lee, and B. Lee, "Retinal 3D: augmented reality near-eye display via pupil-tracked light field projection on retina," *ACM Trans. Graph.* **36**, 190 (2017).

14. D. Lanman, M. Hirsch, Y. Kim, and R. Raskar, "Content-adaptive parallax barriers: optimizing dual-layer 3D displays using low-rank light field factorization," *ACM Trans. Graph.* **29**, 163 (2010).
15. G. Wetzstein, D. Lanman, W. Heidrich, and R. Raskar, "Layered 3D: tomographic image synthesis for attenuation-based light field and high dynamic range displays," *ACM Trans. Graph.* **30**, 95 (2011).
16. D. Lanman, G. Wetzstein, M. Hirsch, W. Heidrich, and R. Raskar, "Polarization fields: dynamic light field display using multi-layer LCDs," *ACM Trans. Graph.* **30**, 186 (2011).
17. 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**, 80 (2012).
18. A. Maimone, G. Wetzstein, M. Hirsch, D. Lanman, R. Raskar, and H. Fuchs, "Focus 3D: Compressive accommodation display," *ACM Trans. Graph.* **32**, 153–1 (2013).
19. 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).
20. 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).
21. G. D. Love, D. M. Hoffman, P. J. Hands, J. Gao, A. K. Kirby, and M. S. Banks, "High-speed switchable lens enables the development of a volumetric stereoscopic display," *Opt. Express* **17**, 15716–15725 (2009).
22. C.-K. Lee, S. Moon, S. Lee, D. Yoo, J.-Y. Hong, and B. Lee, "Compact three-dimensional head-mounted display system with Savart plate," *Opt. Express* **24**, 19531–19544 (2016).
23. J.-Y. Hong, C.-K. Lee, S. Lee, B. Lee, D. Yoo, C. Jang, J. Kim, J. Jeong, and B. Lee, "See-through optical combiner for augmented reality head-mounted display: index-matched anisotropic crystal lens," *Sci. Rep.* **7**, 2753 (2017).
24. 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).
25. T. Zhan, Y.-H. Lee, and S.-T. Wu, "High-resolution additive light field near-eye display by switchable panchromatic–berry phase lenses," *Opt. Express* **26**, 4863–4872 (2018).
26. S. Liu and H. Hua, "A systematic method for designing depth-fused multi-focal plane three-dimensional displays," *Opt. Express* **18**, 11562–11573 (2010).
27. K. J. MacKenzie, D. M. Hoffman, and S. J. Watt, "Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control," *J. Vis.* **10**, 22–22 (2010).
28. J. P. Rolland, M. W. Krueger, and A. Goon, "Multifocal planes head-mounted displays," *Appl. Opt.* **39**, 3209–3215 (2000).
29. S. Ravikumar, K. Akeley, and M. S. Banks, "Creating effective focus cues in multi-plane 3D displays," *Opt. Express* **19**, 20940–20952 (2011).
30. B. A. Wandell, *Foundations of Vision* (Sinauer Associates, 1995).
31. X. Hu and H. Hua, "Design and assessment of a depth-fused multi-focal-plane display prototype," *J. Disp. Technol.* **10**, 308–316 (2014).
32. R. Narain, R. A. Albert, A. Bulbul, G. J. Ward, M. S. Banks, and J. F. O'Brien, "Optimal presentation of imagery with focus cues on multi-plane displays," *ACM Trans. Graph.* **34**, 59 (2015).
33. X. Hu and H. Hua, "High-resolution optical see-through multi-focal-plane head-mounted display using freeform optics," *Opt. Express* **22**, 13896–13903 (2014).
34. M. Levoy and P. Hanrahan, "Light field rendering," in "Proceedings of the 23rd Annual Conf. Computer Graphics and Interactive Techniques," (ACM, 1996), pp. 31–42.
35. T. F. Coleman and Y. Li, "A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables," *SIAM J. Optim.* **6**, 1040–1058 (1996).
36. A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the art algorithm," *Ultrason. Imaging* **6**, 81–94 (1984).
37. D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature* **401**, 788 (1999).
38. M. Hirsch, G. Wetzstein, and R. Raskar, "A compressive light field projection system," *ACM Trans. Graph.* **33**, 58 (2014).
39. S. Lee, J. Cho, B. Lee, Y. Jo, C. Jang, D. Kim, and B. Lee, "Foveated retinal optimization for see-through near-eye multi-layer displays," *IEEE Access* **6**, 2170–2180 (2018).
40. B. E. Saleh, M. C. Teich, and B. E. Saleh, *Fundamentals of Photonics* (Wiley New York, 1991).

1. Introduction

Nowadays, people are fascinated by augmented reality (AR) and virtual reality (VR) devices that several companies start commercializing. The standard of consumer demand is getting higher since consumers want more realistic 3D experiences through the devices. To meet the demands, the devices should provide images with binocular disparity while not inducing vergence-accommodation conflict (VAC). Other factors, such as resolution, frame rates, field of view (FOV) and depth of field (DOF), also affect overall quality of computer-generated volumetric scenes.

However, even the 3D scenes that are generated by the state-of-the-art devices [1, 2] are mainly based on binocular disparity, so that VAC still occurs. Here, VAC means difference in depths resulted by vergence and accommodation. When viewing an object in natural conditions, direction of sight is adjusted by changing the angle between optical axes of both eyes (vergence). At the same time, eye automatically reacts to create a sharp retinal image by adjusting the focal power of eye lens (accommodation). These responses are found to be neurally coupled [3] and it makes them spontaneous. However, this consistency is not maintained when watching stereoscopic 3D displays. Conventional stereoscopic 3D displays provide different images to each eye to produce binocular disparity which makes the system provide proper vergence cues. In this situation, both eyes attempt to converge to the distance where a virtual 3D object is presented. However, two eyes accommodate to the different depth where the actual light comes out. This discrepancy leads to perceptual distortions [4] and makes the simultaneous response of focusing and fusing images difficult [5]. These overall effects cause visual discomfort and fatigue to users [6–9].

Although several types of 3D displays have been proposed to resolve the VAC problem, it still remains as one of the difficult challenges to overcome. Holographic displays [10, 11] have limitations as the system requires equipments with high specification and demands computation with large complexity. Within the category of light field displays, super multi-view displays require multiple tiled displays with high resolution which makes it impractical for most near-eye displays [12, 13]. Further works have developed light field displays in multi-layer structure with compressive techniques [14–17]. Based on the architecture, accommodation cues are supported [18] and a prototype for head-mounted display (HMD) was introduced [19]. The system requirements and the amount of computation load for multi-layer displays are considered to be relatively practicable. Among those candidates, multi-layer displays are considered as the most feasible approach for 3D display.

Layered displays are dominantly classified into two representative types based on the light field synthesis mechanism: additive and multiplicative. Additive light field displays are usually implemented by using optical combiners such as a half mirror [5], holographic optical elements [20] or by temporal multiplexing [21–25]. Meanwhile, multiplicative light field displays dealing with transmittance of each attenuator are implemented by stacking liquid crystal (LC) attenuators [19]. The system has a merit in compact size while loss in brightness is pointed out as a shortcoming.

Regardless of how light fields are synthesized, light field displays in a layered structure are known to present an object which is optically placed between the layers. With a small gap between adjacent layers, the expressible depth range diminishes. On the other hand, the system with a distant spacing cannot support proper accommodation cues since the medium-depth objects are not reconstructed. For these reasons, identifying the optimal gap has been considered as an important subject [5, 26–29]. The depth range that can be covered by two layers is much smaller than the DOF that can be perceived by the human visual system [30], so building multi-layer structure is inevitable. However, as each method of arranging multiple layers has its own limitations, a new design is necessary to achieve a wider DOF.

In this work, we propose hybrid multi-layer displays by integrating two multiplicative displays with a use of half mirror. With the proposed system, some significances can be highlighted in this paper. First, the proposed design allows to expand DOF while supplementing the limitations of other schemes when implementing multi-layer displays. An algorithm for optimizing the layer images is also presented and the feasibility of the algorithm is verified. Lastly, retinal image reconstruction via ray tracing enables the analysis about accommodation not only for the additive displays but also for the multiplicative displays. In section 2, the concept of hybrid multi-layer displays will be introduced with the limitations of conventional layered displays and the attainable improvements. Then in section 3, implementation process of the proposed displays will be discussed. In section 4, analysis about accommodation response will be taken based on

the reconstructed retinal images via ray tracing method. In section 5, experimental results will be demonstrated.

2. Concept

Prior to proposing a way to build light field displays in multi-layered architecture, it is necessary to elaborate on the preceding displays that support accommodation cues. Among them, the following subsection will focus on multi-focal-plane displays and compressive light field displays.

2.1. Related works

2.1.1. Multi-focal-plane displays

Multi-focal-plane displays are one of the promising display technologies to visualize volumetric scenes. A use of image combiners such as beam splitters [5] and switchable lenses [21] allows to construct a continuous volume with additively superimposed planes. However, multi-plane displays with half mirrors suffer from form factor problem. In addition, refresh rate of LCD is pointed out as a limitation for the system employing switchable lenses. In most multi-focal-plane displays, the images for the presentation planes are processed based on linear depth-blending [5,29] or nonlinear depth-fusion functions [31] to achieve minimal depth difference between simulated depth and perceived depth. In image rendering process, only one fixed viewpoint is considered, which makes the system inappropriate in a situation where position of eye varies as in the actual case.

2.1.2. Compressive light field displays

With compressive techniques, light field displays in multi-layer structure have been introduced [14–17]. Later, television-type light field displays [18] and a prototype of VR HMD [19] have been independently reported as focus-supporting displays. Among the proposed systems, attenuation-based displays implemented by stacking LC attenuators [19] synthesize light field by multiplication. The overall enclosure is more compact than any other displays that support two layers while the distant optical spacing between attenuated layers induces degradation in resolution and contrast. Also, brightness diminishes due to transmittance of 0.3 for each LC panel. When rendering each layer image that makes up the displays, it is optimized with different view images, so a certain viewing area is guaranteed.

2.2. Schematic proposal

It is widely known that two displays placed at different depths can reconstruct objects of intermediate depth. In order to express a wide range of depth, the spacing between planes and the number of planes were the factors in utmost priority. Research on the optimal distance between adjacent planes has been constantly carried out to express a wide range of depth with minimal equipment. In the middle-range of angular frequency which is from 4 cycles per degree (cpd) to 8 cpd, optical spacing of around 0.6D (dioptric units, reciprocal of metric distance from eye) was found to be an optimal value for additive layers [5,26,27,29]. Besides, by separating presentation planes as 0.6D, differences are believed to be indistinguishable by human visual system that makes the value an optical gap for additive layers providing "nearly correct focus cues" [32]. However, accommodation range of 0.6D is insufficient as the maximum range of accommodation with a human eye is 8D [30]. As the 2-layered displays failed to express sufficient depth range, the researchers' attention has shifted to increasing the number of planes [5,21,33]. Yet, the way of implementing the multi-layer displays using only one of the existing methods is not practical due to the inherent limitations of each.

To supplement the forementioned limitations in building multi-layer structure, a system combining additive LF displays and multiplicative LF displays is proposed. Hybrid multi-layer

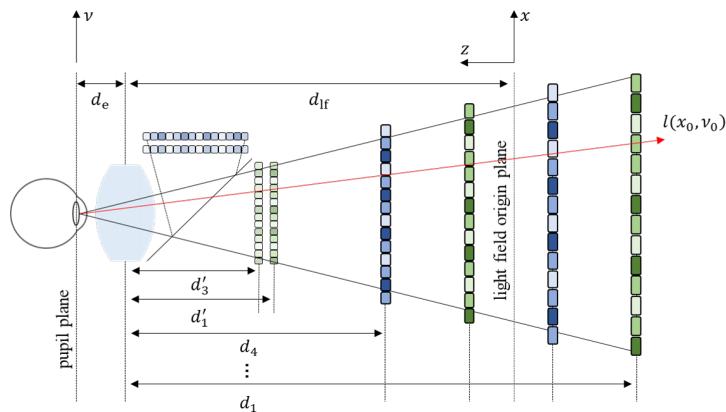


Fig. 1. Schematic diagram and optically equivalent model of the proposed system. The physical displays are presented relatively transparent to the corresponding virtual displays. The half mirror synthesizes the images from two multiplicative displays. Each multiplicative set of layers is colored in a similar tone to show that consecutive depth planes are responsible by the displays placed in the different portions of the system. For simplicity, the gap between the panels, and the spacing between adjacent planes are described as identical respectively.

displays are implemented by integrating two sets of multiplicative displays with a use of half mirror as shown in Fig. 1. Figure 1 shows a schematic diagram and optically equivalent model of the proposed system. By placing physical panels within the focal length of main lens, images presented on LCDs are virtually floated at the discrete depths (0.4D, 1D, 1.6D and 2.2D). Without time multiplexing, this system can provide volumetric scenes with proper accommodation cues to monocular eye for an optical gap of 1.8D. The size of this system is competitive with the dual-focal-plane displays utilizing a half mirror. Only half of brightness is reduced due to the presence of a half mirror compared to the 2-layered multiplicative displays. As it does not require time division to provide images, the system is free from the refresh rate issues. Although the pixelated structure of LC panel makes light diffracted to cause image quality degradation for attenuation-based displays, utilizing panels with dots per inch (DPI) lower than estimated diffraction-limited resolution [19] makes the system less susceptible to diffraction effect.

Even though a half mirror, which is a commonly used image combiner in multi-focal-plane displays, is adopted to integrate two different sets of displays, the images to be displayed on panels are obtained in a compressive manner. Unlike conventional multi-focal-plane displays, which only consider a fixed viewpoint, compressive light field displays tend to be more robust to displacement of eye since several view images are engaged in optimization process. In the next section, implementation of hybrid multi-layer displays will be dealt with light field factorization and optimization process.

3. Implementation

3.1. Light field factorization

To computationally process the images suitable for the proposed system, we have to know the relationship between display patterns and light fields. Since the proposed system as shown in Fig. 1 contains a floating lens and display panels, the imaging formula has to be discussed. Placement of physical panels at a distance of d'_k (k : number of displays) within the focal length f of main lens allows the corresponding virtual image be formed at the distance of d_k with a magnification factor of M_k . Those parameters can be determined by thin-lens approximated equation as Eq. (1):

$$d_k = \frac{1}{\frac{1}{f} + \frac{1}{d'_k}}, \quad M_k = \frac{d_k}{d'_k} = \frac{f}{f + d'_k}. \quad (1)$$

Assuming that the physical stop of lens is wide enough to see the virtual image of n_p pixels in one dimension from display k , the FOV θ_k formed by the virtual image is then,

$$\theta_k = 2\tan^{-1} \left(\frac{M_k p_p n_p}{2(d_k + d_e)} \right), \quad (2)$$

where p_p is the pixel pitch of a micro display and d_e is the eye relief. The FOV increases as the physical panel gets near the main lens. As the maximum angular frequency is determined by the given FOV, the minimum value is chosen to represent the entire system. Therefore, the view frustum of the system has to be built based on the smallest FOV which is θ_1 (constructed by the farthest image from the eye).

For simplicity, 2D parameterization of the light field can be applied to reduce the entire dimension [34]. The 2D light field synthesized by a half mirror can be expressed as

$$\tilde{l}(x, v) = \sum_{m=1}^2 \tilde{l}_m(x, v), \quad (3)$$

where \tilde{l}_1 denotes the light field physically emitted from one side and \tilde{l}_2 from the other side of a half mirror as shown in Fig. 1. x is the coordinate on the light field origin plane and v denotes the coordinate on the pupil plane. And then each \tilde{l}_m synthesized by stacking attenuators with a transmittance of t_m and t_{m+2} is given by

$$\tilde{l}_m(x, v) = \prod_{k=1}^2 t_{2(k-1)+m} (p_{2(k-1)+m}(x, v)) = \prod_{k=1}^2 t_{2(k-1)+m} \left(\frac{1}{M_{2(k-1)+m}} \left(x + \frac{(x-v)(d_k - d_{lf})}{d_{lf} + d_e} \right) \right), \quad (4)$$

where d_{lf} is the distance between the pupil plane and the light field origin plane. The mapping function $p_k : \mathbb{R}^2 \rightarrow \mathbb{R}$ contains the intersection of a light field (x, v) and a display patterns of display k . By discretizing light field presented in Eqs. (3) and (4), it is possible to make a simple equation that stands for the 4D light field as

$$\tilde{l} = \tilde{l}_1 + \tilde{l}_2 = P_1 t_1 \circ P_3 t_3 + P_2 t_2 \circ P_4 t_4. \quad (5)$$

In Eq. (5), $\tilde{l}, \tilde{l}_{1,2} \in \mathbb{R}^L$ are vectors that stand for fully synthesized light field and partially synthesized light field, respectively. Here, partially synthesized light field denotes the light field emitted from each set of multiplicative displays. $t_{1,2,3,4} \in \mathbb{R}^N$ are also vectorized forms of display patterns with the index representing the placing order of panels (1 for the farthest, and 4 for the nearest from the main lens). Projection matrices $P_{1,2,3,4} \in \mathbb{R}^{L \times N}$ which are sparse and binary, map display patterns and light field. Structure of individual matrix is $P_k = [(P_k^{(1,1)})^T \cdots (P_k^{(n_H, n_V)})^T]^T$ (n_H, n_V : number of views in horizontal and vertical direction, respectively), where $(P_k^{(i,j)})^T$ is a mapping matrix for discrete viewpoints (i, j) on the pupil plane. The Hadamard product \circ is employed for element-wise multiplication of matrices.

3.2. Optimization

Based on a compressive manner, images for light field displays in a layered structure have been optimized in various ways. Wetzstein et al. [15] minimized the least-squared error between a target light field and an emitted light field. There have been several algorithms applied to minimize the given error and obtain an optimal set of display patterns [19, 35–38]. As the way

of light field synthesis varies depending on the system, different algorithms are adopted for the optimization process. While for multiplicative displays, non-negative update rules have been dominantly used [19, 37, 38]. Rather, some works optimized images which are formed on retina by tracing the rays incident to pupil [32, 39].

We also minimize the l^2 -norm error between target light field and an emitted light field to obtain an optimal set of display patterns. The simplified formula for optimization is provided as

$$\arg \min_{\{t_1, t_2, t_3, t_4\}} \|l - (P_1 t_1 \circ P_3 t_3 + P_2 t_2 \circ P_4 t_4)\|_2^2, \quad 0 \leq t_{1,2,3,4} \leq 1, \quad (6)$$

where every display pattern holds a normalized value ranging from 0 to 1. As shown in Eq. (6), the light field emitted by the proposed system includes both addition and multiplication in light field synthesis. In addition, we have to acquire an optimal set of 4 different display patterns unlike the existing 2-layered displays. Due to these reasons, it is difficult to apply the widely-used algorithms as they are.

We develop non-negative update rules with an empirical approach to optimize a set of display patterns. The simplest method of error minimization is to make the error itself as zero. In our case, the optimal display patterns can be acquired when a sum of light field emitted by each multiplicative displays equals target light field.

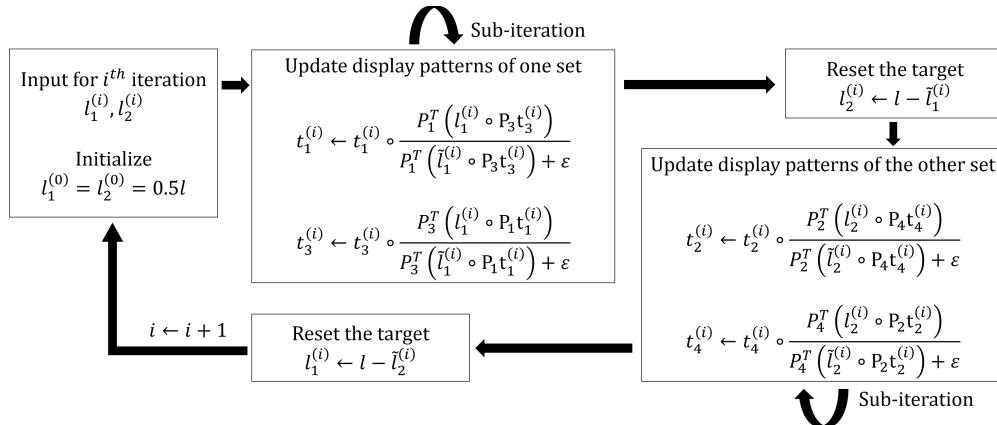


Fig. 2. Flow chart of the display pattern optimization algorithm

Figure 2 shows the flow chart of the given optimization algorithm. As shown in Fig. 2, we iteratively update the display patterns which are responsible for a set of multiplicative displays with a target as difference between the original target light field and the estimated light field from the counterpart set of multiplicative displays. For an initial condition, the target for a set of multiplicative displays is given a half value of the original target light field. Sequential updating of display patterns for individual multiplicative set is conducted until the number of iterations reaches to the designated value. In this algorithm, two things should be confirmed: the number of iterations required for the sub-loops and the entire loop, and convergence of proposed algorithm.

To determine the adequate number of iterations, peak signal-to-noise ratio (PSNR) between target light field and estimated light field, and runtime are measured as the number of iterations in the sub-loop varies. Since the iteration in sub-loop exceeding 3 times does not cause a large PSNR change, we apply 3 iterations for updating display patterns of each attenuated set. Moreover, to determine the number of iterations for the entire loop and check the feasibility of the algorithm, convergence of algorithm should also be investigated. PSNR between target light field and fully synthesized light field is measured in every iteration of the entire loop. PSNR is measured with

two different contents (dice and chess pieces). PSNR curves for both cases are found to be converging. Also, after iterating 30 times, the change in PSNR is less than 0.1 percent of the PSNR measured in the previous iteration. Therefore, 3 times of iteration in both sub-loops and 30 times of iteration in the entire loop are chosen for this algorithm. With this parameter, each iteration for the whole process takes around 25 seconds with MATLAB and convergence of algorithm is confirmed. 3.5 GHz Intel Xeon E5-1620 workstation with a memory of 128GB is employed for this process.

With the stated specifications and hardwares for overall rendering and computation, one iteration takes 4 seconds in rendering the images for 2-plane multiplicative displays with nonnegative update rules. Since one whole iteration of proposed algorithm includes 6 sub-iterations updating 4 different display patterns, it is equivalent as it takes 25 seconds in one iteration. In optimizing images for 2-plane additive displays, rendering the images through trust region algorithm takes 378 seconds. However, the trust region algorithm is a built-in function of MATLAB, and the convergent results appear unlike the results after one iteration of update rules. Since several aspects differ by systems and specifications, it is difficult to provide a simple comparison of rendering speed among the displays. Nevertheless, as 4 planes are concerned in the proposed algorithm, convergence based on PSNR is slightly slower than that of the 2-image-plane systems. To improve the overall rendering speed, parallel processing is recommended. As the paper mainly focuses on presenting the feasibility of the proposed algorithm, rendering through parallel processing is not entailed in this work.

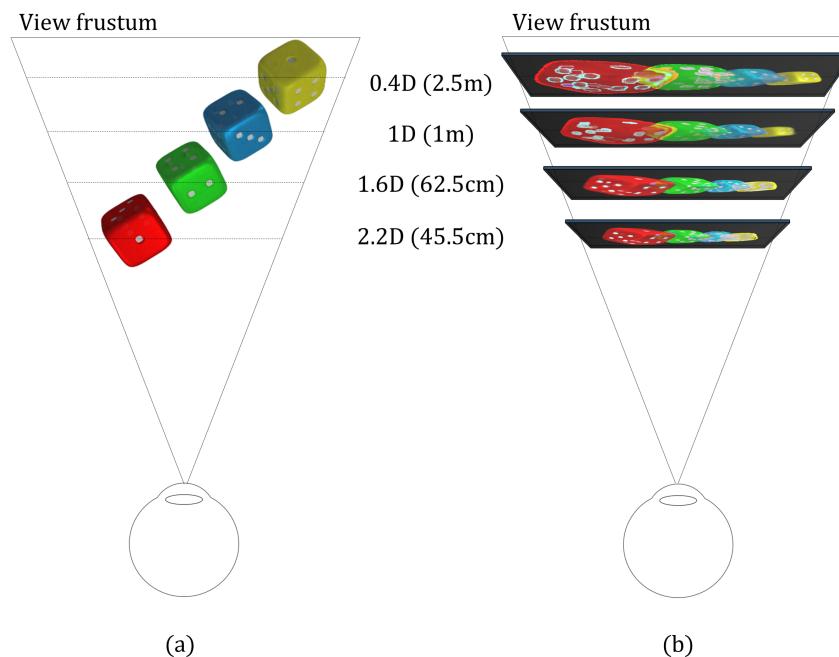


Fig. 3. Concise illustration of hybrid multi-layer displays based on dioptic unit: (a) a top view of rendered volumetric images and (b) a converted model with optimized images virtually floated at the presentation planes. For simplicity, sizes of dice are demonstrated as identical and the gap between adjacent dice as equidistant, which differs in actual case.

Prior to displaying the optimized images that suit the proposed system, target objects have to be constructed and rendered. First, dice are placed at the distance of 0.4D, 1D, 1.6D and 2.2D as

shown in Fig. 3(a). Rendering is processed with the assigned conditions such as view numbers, size of eye-box, FOV, rotation center of the camera to match the requirements of target images that are set by the actual specifications of the system. The detailed specifications of hybrid multi-layer displays will be given in section 5. The whole rendering process is implemented by open-source computer graphics software, Blender. The contents (dice and chess pieces) are captured in 13×13 different views with a resolution of 400×200 and corresponding FOV of $36^\circ \times 18^\circ$. An increase in the number of views indicates dense sampled light field which would enhance the quality of the images. The rendering of different view images with Blender is conducted by lateral shift of camera with a fixed point of sight and a fixed FOV that results in rotation of the reference plane which is not considered in light field synthesis. However, the rotating angle of 0.298° is estimated based on the parameters for rendering process which can be neglected in pickup process.

Hybrid multi-layer displays are implemented by optically placing the optimized images at the presentation planes as Fig. 3(b). This system is expected to induce accommodation response within a range of 1.8D. In the following section, it will be investigated whether this system supports proper accommodation cues according to reconstructed retinal images via ray tracing.

4. Analysis

4.1. Retinal image reconstruction

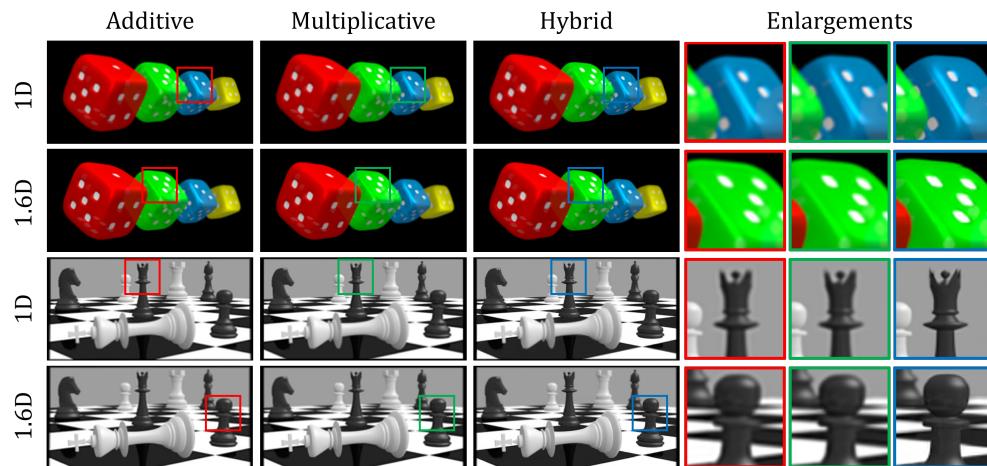


Fig. 4. Retinal images reconstructed for three different displays (additive displays with 2 layers, multiplicative displays with 2 layers, hybrid displays with 4 layers) with two different accommodation states of eye (1D and 1.6D). For 2-layered displays, the spacing between consecutive image planes is 1.8D while 0.6D is guaranteed between adjacent layers for hybrid displays. The black border lines appear due to the sampling parameter in the retina plane. Enlargements are shown on the right side of the figure.

The portion that plays an identical role as the image sensor of the camera in the eye is the retina. To check accommodation response, it is important to consider the image that is actually formed on the retina. Each display forms an image on retina according to the wave-based incoherent point-spread function (PSF) of eye-lens based on the accommodative state and the position of display. For additive displays, retinal images can be simply formulated as a total sum of convolution of PSF and image. On the other hand, it is difficult to present the retinal image formed by the proposed system in the conventional way. By adopting ray-approximated PSF via ray tracing method instead of wave-based PSF, we can easily reconstruct retinal images formed

by proposed displays. Besides, it would lessen the overall computation load as it deals with sampled light rays rather than wave. In retinal image reconstruction, retina is assumed to be planar. Then, 9×9 different views in pupil plane are sampled in the size of $4\text{mm} \times 4\text{mm}$. By the ray transfer equations [40], mapping functions between retina and image planes are obtained. As the reconstruction of retinal images for the proposed system requires direct synthesis of sampled rays, consisting the overall operation in a matrix form reduces overall memory capacity and computation time. For this reason, retina projection matrices consisting of mapping function between rays and display patterns in discrete planes are formed based on the accommodative states and the positions of image planes.

In Fig. 4, reconstruction of retinal images is conducted with the optimized images of hybrid displays placed at 0.4D, 1D, 1.6D and 2.2D respectively. For comparison, retinal images for the conventional 2-layered displays are also provided with the factorized images that are placed 1.8D apart with each other (0.4D and 2.2D). Note that the arrangement of 4 layers in conventional ways (additive,multiplicative) is impractical for forementioned reasons and is therefore excluded from this comparison. The layer images of conventional 2-layered displays are obtained with trust region algorithm [35] and non-negative update rules [37], respectively. There are two different contents (dice and chess pieces) shown and retinal images of the contents are given with two different accommodative states of eye (1D and 1.6D). The object with a small rectangle on it is the one placed at the depth equivalent to the depth of corresponding accommodation state. It can be observed from Fig. 4 that medium-depth objects maintain the high frequency information in hybrid displays. Meanwhile, the objects which are located near the depths where the layer images of 2-layered displays are, are also reconstructed in a similar quality with hybrid displays.

4.2. Accommodation response

To show whether hybrid multi-layer displays support accommodation cues within the given depth range, quantitative analysis is conducted. For the analysis, we plot contrast ratio curves of different layered displays in Fig. 5. Human eyes are known to accommodate to the depth where the maximum contrast ratio is achieved to perceive a sharp retinal image [27]. In this simulation, a 2D image is provided as a 3D target object for light field synthesis. For instance, view images provided for light field factorization are rendered with the target object being placed at the specific depth. With the retinal images that vary with accommodation states and display mechanisms, contrast ratio curves of each angular frequency are estimated in Fig. 5. For generality, some frequencies near the center frequency (2 cpd and 4 cpd) are sampled. The maximum angular frequency that can be achieved with the specifications of the proposed system is limited as 5 cpd.

The target object is placed at the stated optical distances (0.7D, 1.3D and 1.9D). Retinal images are reconstructed depending on accommodation states and display mechanisms. For retinal images simulated in a natural viewing condition, reconstruction is directly done with a raw rendered image of target object. Contrast ratio is derived by Fourier transformed retinal images and normalized by Fourier transformed retinal image obtained in natural viewing condition which is reconstructed with the specific accommodation depth equivalent to the target depth.

As a result, in Fig. 5, hybrid displays achieved the highest contrast ratio at the target depth regardless of the target object's position in 2 cpd range. In 4 cpd section, although the maximum value is not achieved at the target depth, only a slight difference existed. Additive displays with spacing of 1.8D have the maximum contrast ratio at the virtual image plane close to the target depth. Moreover, even in the range of 2 cpd, the depth with the maximum value of the contrast ratio for multiplicative displays is found to be dependent on the depth of the target object. Such a result is obtained without consideration of the image quality degradation problem due to the diffraction effect, which is known as an inherent drawback of the multiplication displays. For this reason, the optimal spacing between adjacent multiplicative layers should be smaller than 1.8D. This result poses a question on the value of 4.45D which was presented as an optical spacing

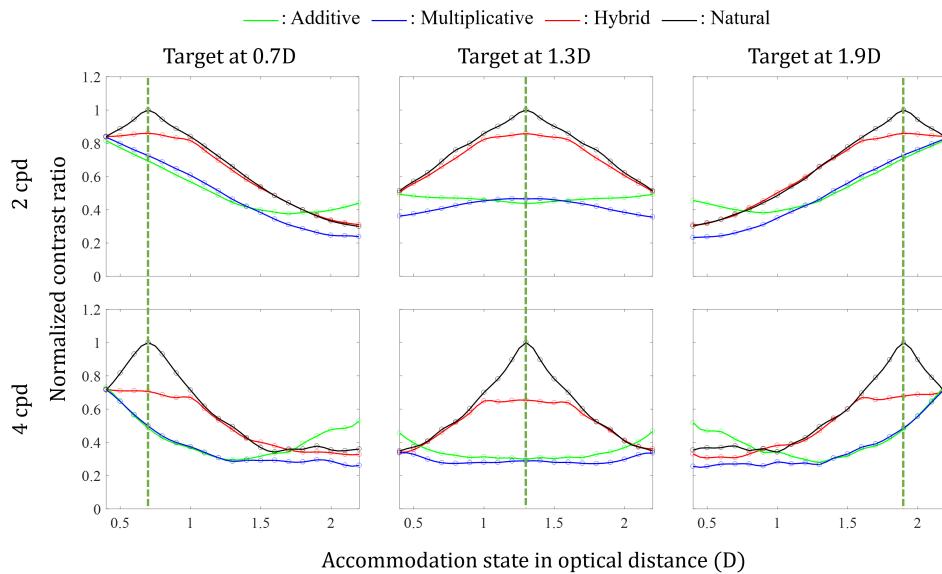


Fig. 5. Normalized contrast ratio curves of multi-layer displays depending on the angular frequency and placement of the target object: hybrid displays (4 layers, red line), additive displays (2 layers, green line), multiplicative displays (2 layers, blue line) and estimated normalized contrast ratio curves in a natural viewing condition (black line). Green dashed line indicates the depth where the target object (light field origin plane in light field synthesis) is located. For additive and multiplicative displays, layers are optically located at 0.4D and 2.2D while four layers are placed at 0.4D, 1D, 1.6D and 2.2D for hybrid multi-layer displays.

between multiplicative layers in the work of other group [19]. In addition, hybrid multi-layer displays have a relatively higher peak value of contrast ratio compared to other displays, which enables the viewers to perceive sharper images.

5. Experiments

To investigate the feasibility of hybrid multi-layer displays and observe whether this system provides accommodation cues to monocular eye, an experimental setup is built as shown in Fig. 6.

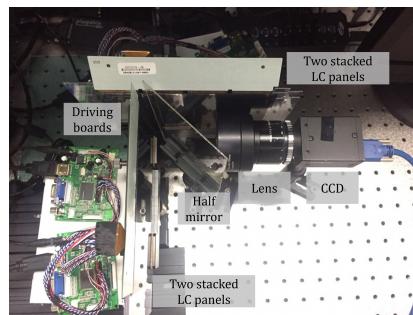


Fig. 6. Experimental setup of hybrid multi-layer displays for monocular eye.

Table 1. Detailed specifications of hybrid multi-layer displays

Specifications	Values
Eye relief	1 cm
Size of eye-box	6 mm (horizontal) × 6mm (vertical)
Number of views	13 × 13
Pixel pitch of LC panel	117 μ m
Focal length of main lens	75 mm
Resolution of image	400 × 200
Optical distances of LC panels from eye	0.4D, 1D, 1.6D, 2.2D
Expressible depth range	1.8D

In the experiment, Chimei Innolux N070ICG-LD1 panels with 217 DPI and a resolution of 1280×800 are utilized as displays. For the front panels of multiplicative sets, backlight units are removed and panels are placed in a reverse direction to match the polarization angle with the corresponding rear panels. Each panel is placed and aligned to match the optical distances of 0.4D, 1D, 1.6D and 2.2D, respectively. Acrylic plates and OHP films which are 1mm and 0.1mm thick respectively, act as spacers for precise and uniform alignment of physical panels which function as multiplicative displays. Since the system size depends on the focal length of main lens, a product of Thorlabs (AC508-75-A-ML) that has focal length of 75mm and diameter of 2 inches is employed as a main lens. The charge-coupled device (CCD) camera (Pointgrey, GS3-U3-91S6C/M-C) which has pixel pitch of 3.69μ m × 3.69μ m and a resolution of 3376×2704 is used to capture the experimental results. Due to the presence of half mirror, low-resolution images are displayed using the side parts of the panels. The f/1.4 c-mount lens with focal length of 16mm is utilized to capture magnified virtual images with a small depth of focus. The size of eye-box in the experimental setup is given as 6mm × 6mm. Eye relief is minimized as 1 cm by placing the c-mount lens at a distance close enough to the main lens so as not to affect overall image fidelity. The detailed specifications of hybrid multi-layer displays are shown in Tab. 1.

For calibration, gamma value for the panels was measured as 2.4 and we post-processed optimized images based on the measured value. In addition, several steps were taken to align the images before being displayed. First of all, an image with the white line on the edge and the same resolution as the processed image is displayed on the farthest panel and placed in the center of CCD. If the image deviates much from the center of the CCD, the distortion caused by lens becomes noticeable. Then, the sizes of images should be manipulated to match the border line with the same image for the other displays due to the different magnification factors caused by the lateral placement of the panels. During this process, symmetry of the images should also be considered since the front panels of multiplicative sets are oriented in the reverse direction and displayed images are combined with a half mirror.

After the calibration process, we display optimized images with hybrid type of multi-layer displays. With the prototype as shown in Fig. 6, display photographs of hybrid multi-layer displays are captured as shown in Fig. 7. The results are captured by modifying the focal plane of CCD camera to focus at the depth where the objects are located. Retinal blur is clearly observed from the captured images based on the focal depth of CCD camera and placement of objects as the results are shot with small depth of focus lens. The support of accommodation cues is confirmed with the enlarged photographs, which are demonstrated on the right side of Fig. 7. To show the image fidelity of the proposed system, one is mainly rendered with RGB color while the other with black and white color.

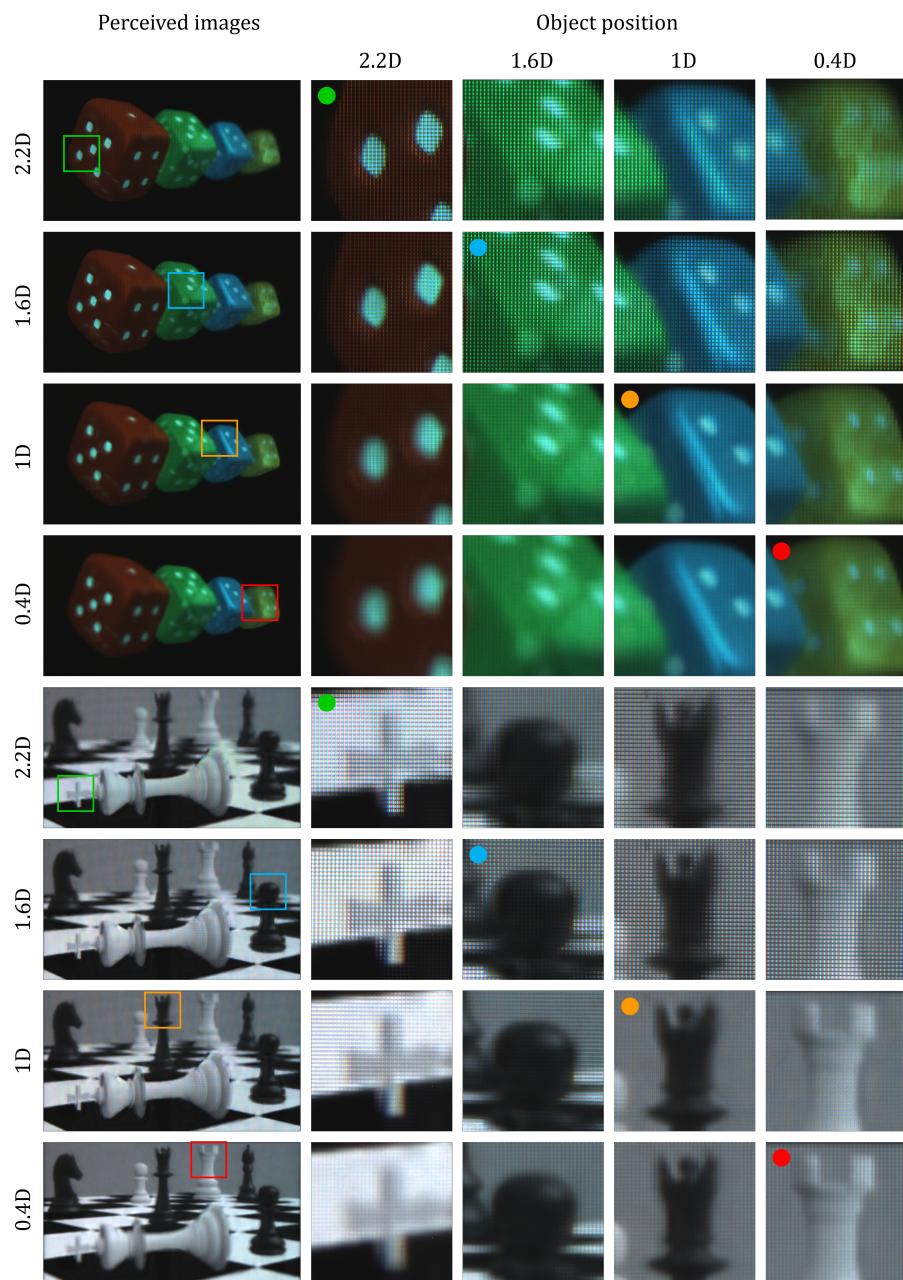


Fig. 7. Display photographs of hybrid multi-layer displays taken by modifying the accommodation depths of CCD. To confirm the focus cues, some parts of the images with different accommodation depths are enlarged. Enlarged image with a colored dot in the upper left side indicates that the accommodation distance of the camera is matched with the depth of rendered object. Retinal blur is observed from all results since the images are captured with small depth of focus lens. Sub-pixels of the displays are observed in the perceived images with the camera's focal distance of 1.6D and 2.2D.

In the results, accommodation cues were confirmed within a depth range of 1.8D. The contrast differences were noticed in RGB colored objects as well as black objects such as chess pieces of queen and pawn. As the objects are magnified due to employment of c-mount lens with large focal length and main lens with minimized focal length, pixelated structures of LC panels are observed in the perceived in Fig. 7. Due to this phenomenon, it is difficult to figure out the images with best quality. It can be alleviated if c-mount lens with small focal length is employed to capture the result. Moreover, by utilizing a main lens of larger focal length, high resolution images can be displayed in hybrid multi-layer displays.

In summary, a prototype of hybrid multi-layer displays is built with the given specifications. Prior to display the optimized images, calibration is processed. As the processed images are displayed in the prototype, it is verified with the experimental results that the system can provide adequate accommodation cues to monocular eye in a range of 1.8D.

6. Conclusion

In this work, a hybrid multi-layer display is proposed as it combines additive LF displays and multiplicative LF displays. The system is implemented by integrating two sets of multiplicative displays with a half mirror. As the optimal distance between adjacent planes is found to be 0.6D for additive layers, images are virtually floated at the optical distance of 0.4D, 1D, 1.6D and 2.2D to cover a range of 1.8D. To present the images for the displays, light field is factorized in the compressive manner and optimized with the proposed algorithm which includes non-negative update rules. In order to ensure that the proposed displays support adequate accommodation cues, retinal images with various accommodation states and display mechanisms are reconstructed via ray tracing method. Retinal images and contrast ratio curves are obtained based on the optimized images generated according to each synthesis mechanism to compare the proposed displays with the existing displays. From this simulation, hybrid multi-layer displays were verified to induce proper accommodation response regardless of target object's position in an optical interval of 1.8D. To back up the given simulation results, a prototype of hybrid multi-layer displays is built and display photographs of two different contents taken by changing the focal distance of the camera are provided. Hybrid multi-layer displays were confirmed to support accommodation cues for a range of 1.8D.

Although the brightness reduction is experienced due to the presence of a half mirror, this can be alleviated if the half mirror is replaced with a polarizing beam splitter (PBS) and half wave plates (HWP). Since the LC panels are linearly polarized, if we match a linearly polarized state of a set of LC panels to one of the PBS's polarizing states and employ HWP to match the other side, the brightness reduction can be theoretically assumed to be zero. However, transmission characteristics of PBS are dependent on the incident wavelength. This can be pointed out as a drawback of applying PBS as an image combiner of the given setup. In addition, as the paper does not provide the comparison with the experimental results of conventional additive and multiplicative displays, thorough comparison and quantitative analysis with the experimental results may remain as one of the future works.

The proposed displays have the greatest significances in that the DOF of system is expanded to 1.8D, the form factor is competitive to the existing additive displays with 2 layers, only half of brightness degradation is appeared compared to the multiplicative displays with 2 layers, and no time division is required in building multi-layer architecture. Moreover, we discovered an algorithm to optimize the layer images for the proposed system. The accommodation response was confirmed by reconstructed retinal images via ray tracing method. With this approach, we were able to analyze the accommodation response based on the optical gap between additive layers as well as multiplicative layers.

Funding

Korea Government (MSIT) (Development of vision assistant HMD and contents for the legally blind and low visions) under Grant 2017-0-00787.

Acknowledgments

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