

Hybrid Approach of Holography and Augmented-Reality Reconstruction Optimizations for Hyper-Reality Metaverse Video Applications

Hyoung Lee[✉], Hakdong Kim, Taeheul Jun, Wookho Son, Cheongwon Kim, and MinSung Yoon[✉]

Abstract—In this study, we offer a new type of hyper-realistic holographic display system that simultaneously displays both holographic 3D images and AR images. The proposed ultra-realistic display combines 360-degree computer generated holographic (CGH) 3D content to be reconstructed from an SLM and AR content (2D image) to be spatially projected from a microdisplay, allowing users to watch a clearly blended movie without crosstalk. To validate the proposed display's hyper-realistic 3D image characteristics with varied depths, a ruler-based spatial depth measurement method was used to show that images appearing in the real 3D space were clearly reconstructed at different depths. Furthermore, we demonstrate that CGH content synthesized using a deep learning model that extracts high-precision depth maps from RGB color images can be successfully applied to the proposed display system through numerical and actual optical reconstruction experiments. Thus, it is possible to provide users with the effect of maximizing three-dimensional (3D) expression and natural immersion by using the new hybrid display for both free depth expression and clear hyper-realistic 3D expression. Furthermore, the accommodation effect and free-depth control characteristics demonstrated in the proposed system allow viewers to enjoy super-realistic 3D content comfortably, implying that eye fatigue may be overcome even when watching metaverse content for extended periods.

Index Terms—Hyper-realistic holographic display, AR display, deep learning, depth estimation, optical arrangement, hybrid display.

Manuscript received 21 March 2023; revised 17 July 2023; accepted 11 August 2023. Date of publication 8 September 2023; date of current version 11 December 2023. This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korea Government (MSIT, XR User Interaction Evaluation and Employing Its Technology) under Grant 2022-0-00137. (Corresponding author: MinSung Yoon.)

Hyoung Lee and Wookho Son are with the CG/Vision Research Section, Content Research Division, Hyper-Reality Metaverse Research Laboratory, Electronics and Telecommunications Research Institute (ETRI), Daejeon 34129, Republic of Korea (e-mail: hyoung0708@etri.re.kr; whson@etri.re.kr).

Hakdong Kim is with the Department of Digital Contents, Sejong University, Seoul 05006, Republic of Korea (e-mail: hdkim0426@gmail.com).

Taeheul Jun is with the Department of Software Convergence, Sejong University, Seoul 05006, Republic of Korea (e-mail: junth92@naver.com).

Cheongwon Kim is with the Department of Software, Sejong University, Seoul 05006, Republic of Korea (e-mail: wikim@sejong.ac.kr).

MinSung Yoon is with the VR/AR Content Research Section, Content Research Division, Hyper-Reality Metaverse Research Laboratory, Electronics and Telecommunications Research Institute (ETRI), Daejeon 34129, Republic of Korea (e-mail: msyoon@etri.re.kr).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TBC.2023.3308328>, provided by the authors.

Digital Object Identifier 10.1109/TBC.2023.3308328

I. INTRODUCTION

THREE-DIMENSIONAL (3D) near-eye display (NED) technologies for the metaverse, a 3D virtual space with activities similar to those found in the real world, have recently emerged as next-generation displays capable of delivering visually realistic 3D experiences [1], [2], [3], [4], [5]. A metaverse environment refers to a display platform that incorporates virtual reality (VR), augmented reality (AR), or holographic reality (HR), as well as a see-through optical performance, such as mixed reality (MR), to blend the real world and the virtual image floating in front of the eye [6], [7], [8], [9], [10]. Because these display technologies for hyper-realistic environments have the common goal of producing high-quality 3D image content for a viewer or user to experience a total immersive environment with comfort, they have shown promising potential for various attractive applications such as educational training, healthcare, entertainment, and manufacturing. In particular, to provide a high-resolution 3D virtual scene with eye comfort, it is crucial to reduce the problem of accommodation-convergence (AC) conflict that arises from inaccurate focus cues [8], [11]. Holography, one of the key candidates for hyper-realistic expression, is a technology that can accurately reconstruct the wavefront of an object or scene in real 3D space using amplitude and phase modulation. Therefore, several display types that use holographic 3D (H3D) methods have been proposed to resolve the AC conflict problem. In particular, holographic fringe data can be synthesized from numerical calculations called computer-generated holograms (CGH) or digital holograms [12], [13], [14], [15]. However, the digital hologram display requires a complex computational process that requires a long time for CGH synthesis and must suppress high background noise that appears near the reconstructed target objects or around the reconstructed scene [16], [17], [18]. Therefore, a new approach for alternative displays and content creation methods must be developed to deliver a comfortable view as well as a high-quality 3D virtual scene.

Previous studies related to realistic 3D-NED devices have been divided into volumetric display methods, light-field-based 3D display methods, and real holographic display methods. First, the volumetric display approach was used to express multiple focal planes [19], [20], [21], [22], [23], [24]. The latest multifocal technologies, such as Maxwellian NED, multifocal NED, vari-focal NEDs, and multiplexed NEDs, were

systematically presented. In addition, solutions to problems that cause performance degradation and visual fatigue in stereoscopic displays based on the volumetric method were studied. However, it is necessary to overcome the difficulty in making accurate focus-variable adjustment of the liquid crystal utilized for vari-focal NEDs, the requirement for high-specification hardware for fast response speed in multifocal NEDs, as well as the structural complexity due to stacking geometry, accurate synchronization, and high-power issues in multiplexed NEDs. Second, for 3D depth expression, the multi-panel-based light-field 3D display method has been studied [25], [26], [27]. This approach incorporates a hybrid concept to expand and control the depth of content, which can be expressed on the basis of various algorithms, diffraction optics, and the additional insertion of optical elements. However, it is necessary to overcome the problems of brightness degradation and crosstalk caused by multipanel applications, including difficulties in precisely aligning and structurally combining multiple panels. Third, the digital holographic approach, which has recently been suggested and developed for NED devices so that AR device users can experience convenient, realistic 3D content, is an ideal candidate to resolve the accommodation-convergence conflict. Holographic 3D display systems are comprised of multiple elements, including holographic or lithographic optical elements for a slim form factor and holographic content generation algorithms for calculation cost reduction based on deep learning models [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40]. Despite these advancements in holographic NED, current research does not, with the exception of challenges in a thin lightweight optics system, offer sufficient key solutions to obtain multifocal and wide-depth cues, and the optical capability to simultaneously express a high-resolution scene that incorporates both 3D objects and colorful backgrounds.

To overcome these restrictions, we worked on a hyper-realistic, immersive NED system that would provide users with a hyper-realistic, immersive 3D experience by using a hybrid approach or a combination of a holographically reconstructed 3D scene and an AR-floated 2D scene, each represented at different depths with a wide range in real free space. Furthermore, we propose an advanced holographic-content-generating method based on the modification of a deep learning technique called the HDD model [40] to demonstrate the possibility of real-time processing or real-time interaction of 360-degree multi-viewed dynamic content suitable for the proposed hyper-reality display system. A multi-view CGH for a movie or dynamic scene is synthesized from the FFT algorithm using a color image and depth map prepared from the proposed deep learning model. This experimental confirmation using the proposed hybrid 3D system demonstrates that we can show 360-degree, colorful, and hyper-realistic H3D video scenes.

II. DISPLAY DESIGN AND CONFIGURATION

The hyper-reality holographic 3D display that we demonstrate in this study is characterized by a physical combination of a holographic display (a 3D volume image with a depth

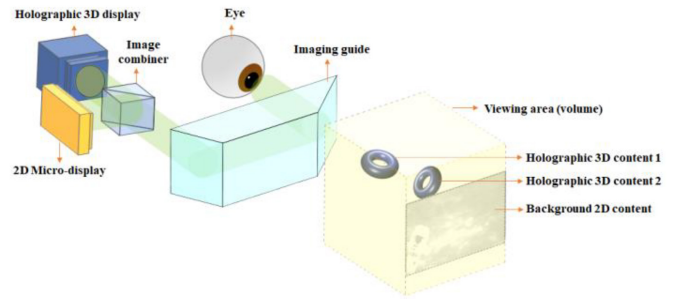


Fig. 1. Schematic of the hyper-reality holographic 3D near-eye display system.

expression from CGH) and an AR display (a 3D spatial projection of a 2D background image). Fig. 1 shows a schematic of the overall concept of the proposed hyper-reality holographic 3D display. Using this proposed approach, we demonstrated that high-quality holographic 3D objects and vivid 2D background images can be located at different spatial depths so that each user can have hyper-realistic, immersive 3D experiences [41], [42]. The components of this system are as follows: First, a laser module with three wavelengths (red = 633 nm, green = 532 nm, blue = 488 nm, MatchBox Laser Series) was used for illumination. The collimated beam was prepared to illuminate the active area of the SLM [43] by passing the RGB laser beams through a fiber-coupled collimator (model: C80APC-A, Thorlabs) and a beam extender (model: WBEF-A1029-V, WIKI Optics).

Second, the two types of devices used for image display are as follows: An LCoS-SLM (model: IRIS-U62, May Display) was used for optical reconstruction of the digital hologram content. The reflective SLM had a resolution of 3840×2160 pixels (an active area of 0.62 inches), a pixel size of $3.6 \mu\text{m}$ with a fill factor of 90%, and a frame rate of 120 Hz. A single field lens (lens 1, focal length $f = 500 \text{ mm}$) was placed next to the SLM panel to view a holographic 3D image that was optically reconstructed using the viewing window approach as a holographic restoration technique. The 2D background image was displayed using an OLED microdisplay (model: BT-300, Epson). The 2D panel had a resolution of 1280×720 pixels (an active area of 0.43 inches) and a diagonal viewing angle of 23° . To combine both images, an image combiner (or half mirror) is positioned at the intersection of a light path line from the SLM and a light path line from the microdisplay (Fig. 2(d)). The 2D background image is placed at a depth on the surface of the SLM when the image combiner is located at the intersection and the 2D micro-display is placed at a distance equal to the distance between the combiner and the SLM. In particular, when a convex lens is inserted between the microdisplay and half-mirror, the observed background image can be magnified according to the lens formula. Through the image combiner, the depth of its 2D background image that will be floated in real space as well as the size of its projected image can be adjusted using a convex lens (Lens 2) with the microdisplay. Experimental observation of the reconstructed holographic 3D objects and the background AR image was performed using a DSLR camera (model: EOD 5D Mark III,

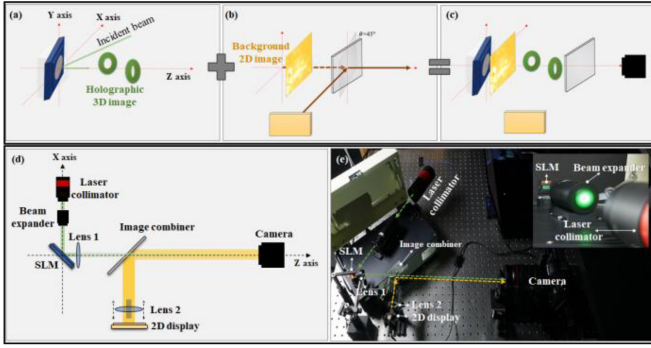


Fig. 2. Configuration of the prototype apparatus for experimental demonstration of the proposed hyper-reality holographic 3D near-eye display. (a): Component of holographic 3D display. (b): Component of 2D micro-display for AR floating. (c): Hyper-realistic hybrid display realized by combining two components. (d): Geometry of the proposed integrated system (top view). and (e): System setup for demonstration (perspective view).

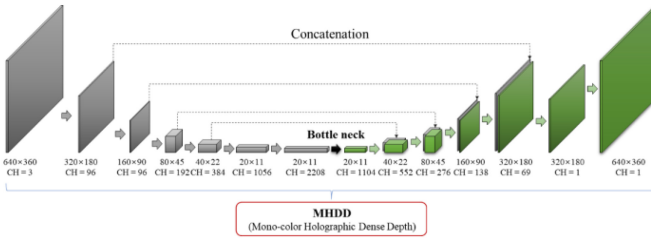


Fig. 3. Pipeline of the proposed deep learning model (MHDD) for estimation of depth map information from the mono-color image where a set of the color and depth map image is used as an input in hologram synthesis.

Canon) placed at the focal point of the positive field lens. Still images and videos displayed for demonstration were captured using a DSLR camera (36×24 mm CMOS sensor, 5760×3840 resolution). The experimental setup is shown in Fig. 2. Figs. 2(a), (b), and (c) show the schematic for a hyper-reality holographic 3D display and the conceptual diagrams of each display component. Figs. 2(d) and (e) show the geometry and a photograph of the experimental setup for the demonstration system, respectively.

III. CGH GENERATION METHOD FOR HOLOGRAPHIC 3D CONTENT

In this section, we describe the processes for creating and reconstructing 360-degree multi-viewed CGHs suitable for the proposed display system. We discuss a deep learning method that was developed to extract high-precision depth maps for multi-viewed CGHs. One of the techniques that can overcome the computational cost problem of hologram synthesis and the time delay problem when the user interacts with hologram content is the deep learning-based depth map estimation method. We adapted a deep learning model called Holographic Dense Depth (HDD), which helps in synthesizing the 360-degree multi-viewed CGHs at high speed [40]. Unlike the previous HDD model, which is optimized for RGB color and depth map input, the advanced deep learning model that we designed for this study is a monochromatic-HDD (MHDD) model dedicated to monochromatic principal color and depth map input.

This advanced MHDD model applied to depth map image estimation consists of an encoder structure that extracts a feature map from an RGB image and a decoder structure that extends the feature map that has been reduced through the up-sampling process [38]. Fig. 3 shows the basic structure diagram of the proposed MHDD model. In the MHDD model, the encoder uses the existing pre-trained DenseNet-161 model and extracts the feature map by performing a downsampling process on a single green color image, which is the input data of the model [44]. In the decoder, the feature map of the shallow layer is concatenated to the feature map of the deep layer by a skip connection, and the size of the previously reduced feature map is expanded again through the upsampling layer. Finally, the output layer outputs one depth map image estimated to have the same size as the input color image through a bilinear interpolation process. The model was trained by reducing the quality difference between the depth map image estimated using the loss function and the ground truth depth map image. The model structure used in the study was the same as that used in [36], and RGB images of green mono-color images were used as input data.

The key steps to perform the task are as follows: First, we acquired a training data set that consists of green color and depth map images (resolution 640×360) where green corresponds to monochromatic principal color [45]. Subsequently, we applied the CNN-based MHDD model, which can quickly and precisely estimate depth maps from new green color images at various missing viewpoints, as a deep learning model for extracting depth map images using multi-viewed green color images as input data [44]. Second, through the resolution conversion process, the original RGB color image and the predicted depth map image are converted into a 4K (3840×2160) resolution corresponding to the SLM's resolution used in the proposed display system. Third, CGH synthesis is performed using the Fast Fourier Transform (FFT) algorithm, wherein the original color image and depth map image estimated by the HDD model are utilized as input data. Finally, a holographic 3D scene restoration experiment in a real 3D space was performed using the proposed system to monitor the quality of 3D images reconstructed from CGHs.

For the acquisition of 360-degree holographic 3D content, we prepared a dataset of multi-viewed original RGB color and depth map images (image resolution: 640×360) using the 3D rendering tools provided by Maya Software [46]. The dataset comprises 1,024 viewpoints, which correspond to a rotation of 360° for each of the four groups of objects (shapes of cone, cube, sphere, and torus). We created two identical 3D objects for each shape in each scene that were placed near the origin in Maya SW. After a virtual camera was set to be rotated in a circular orbit to capture these two objects, we obtained 360-degree multi-viewed RGB color and depth map images for each scene. Here, the virtual camera was used to shoot the RGB color and depth map images while moving by 0.351° per step around the rotational axis at the origin.

The grayscale depth map image acquired in 256 gray level formats is required for both the CGH synthesis and deep learning processes for depth map extraction because it provides

phase information, which is an essential element in hologram data. Phase information corresponds to the distance between the position of the camera and the surface of the target (object). We conducted the training and testing process using the MHDD model using the dataset obtained from four types of shapes, which were made of only green-colored 3D objects for each shape. From 4,096 data pairs, where a pair of images consisted of a green color and a depth map image, we used 60% of the data pairs as a training dataset for the deep learning model and the remaining 40% as a test dataset to measure the performance of the trained model.

The CGH image per viewpoint was synthesized through the fast Fourier transform (FFT) algorithm using a set of RGB color and depth map images as input per viewpoint. Typically, the hologram function $H(x, y)$ contains complex numbers, which can be expressed as

$$H(x, y) = |H(x, y)|e^{i\Phi(x, y)} \quad (1)$$

where $|H(x, y)|$ and $\Phi(x, y)$ are the amplitude and phase, respectively, of the hologram function. The FFT principle of CGH generation for the proposed holographic display can be explained on the basis of a cascaded Fresnel transform [32], [47]. Here the optical field at the holographic display panel with the size of $X \times Y$ and that of retinal plane of the observer are denoted as $H(x_1, y_1)$ and $G(x_2, y_2)$, respectively, for the proposed system in Fig. 2(d). With an assumption that the distance (D_1) between the holographic display and the eye lens is equal to the focal length (f_1) of the field lens (Lens 1), the size of user's eyeball and the focal length of eye lens are D_2 and f_2 , respectively. It is also assumed that the hologram plane (x_1, y_1) near the field lens and the retinal plane (x_2, y_2) of the user are located in parallel with the eye lens plane (u, v). The optical field relation from the cascaded Fresnel transform are given by

$$\begin{aligned} G(x_2, y_2) &= a(x_2, y_2) \iint \left[e^{\frac{i\pi}{\lambda} \left(\frac{1}{D_1} + \frac{1}{D_2} - \frac{1}{f_2} \right) (u^2 + v^2)} \iint h(x_1, y_1, u, v) H(x_1, y_1) dx_1 dy_1 \right] \\ &\times e^{-\frac{2i\pi}{\lambda D_2} (x_2 u + y_2 v)} dudv, \end{aligned} \quad (2)$$

where $a(x_2, y_2) = -e^{\frac{2i\pi(D_1 + D_2)}{\lambda}} \frac{e^{\frac{i\pi(x_2^2 + y_2^2)}{\lambda D_2}}}{D_1 D_2 \lambda^2}$, and $h(x_1, y_1, u, v) = \text{rect}\left(\frac{x_1}{X}, \frac{y_1}{Y}\right) \times e^{\frac{i\pi}{\lambda D_1} (x_1^2 + y_1^2)} e^{-\frac{2i\pi}{\lambda D_1} (x_1 u + y_1 v)}$.

When 1,024 CGHs for each solid shape were prepared using the FFT algorithm, an additional process called Lee's encoding scheme [48] was applied so that they could be represented directly on the amplitude-modulating LCoS-SLM that we used for the study. Lee's encoding decomposes a complex-valued optical field into four components with real and non-negative coefficients $L_m(x, y)$ for each component. At least two of the four coefficients $L_m(x, y)$ are equal to zero. Lee's representation of the hologram function can be written as

$$\begin{aligned} H(x, y) &= L_1(x, y)e^{i0} + L_2(x, y)e^{i\pi/2} \\ &+ L_3(x, y)e^{i\pi} + L_4(x, y)e^{i3\pi/2} \end{aligned} \quad (3)$$

The MHDD model, is similar to the previous HDD model, is made up of an encoder that extracts a feature map while

TABLE I
PERFORMANCE MEASUREMENT RESULTS FROM THE MHDD MODEL
CONCERNING FOUR KIDS OF 3D SHAPES

Indicator Shape	SSIM	PSNR	RMSE	ACC	LRMSE	Abs rel
Cone	0.9999	84.52	0.0006	0.9906	0.0049	0.026
Cube	0.9999	84.39	0.0003	0.9933	0.0050	0.017
Sphere	0.9999	85.03	0.0002	0.9961	0.0047	0.015
Torus	0.9999	84.90	0.0003	0.9922	0.0049	0.022

shrinking the size of the input image data, and a decoder that expands the size of the reduced feature. In the encoder process, down-sampling is performed on the input data image using the widely known Pre-trained DenseNet-161 model, and feature maps are extracted [49]. Subsequently, the size of the reduced feature map is gradually expanded through the upsampling layer in the decoder process. In this process, the feature maps of each upsampling layer in the decoder are concatenated with the feature maps of each layer in the encoder. This operation contributes to accurate depth prediction. Finally, the feature map was expanded to the same size as the input green image through bilinear interpolation in the output layer and output as a single depth map image. The loss function used for training the MHDD model is a combination of structural similarity (SSIM) and mean squared error (MSE), with an MSE to SSIM ratio of 8:2. Here, the coefficients of the loss function for the MHDD were derived by experimentally exploring the best depth estimation task from each missing viewpoint using multi-viewed green color and depth map images as input data for 360-degree holographic content. The loss function is given by

$$\text{Loss function} = (0.8) \times \text{MSE} + (0.2) \times \text{SSIM} \quad (4)$$

Details about the loss function, including six types of image quality and performance measurement indicators, which appear in Table I, can be found in the Supplementary Information.

Table I lists the performance measurement results of the MHDD model trained under the loss function condition using 1,640 test data that were made from four types of 3D shapes. In Table I, each indicator value concerning individual 3D shapes is the average result from 410 test data points for the corresponding indicator. In this case, each performance indicator was used to calculate the image difference between the MHDD-estimated depth map and the ground truth depth map image. From this measurement result, we find that the peak signal-to-noise ratio (PSNR) is greater than 84 dB and the accuracy (ACC) is greater than 0.99 in all four shapes: Here we define the accuracy index (ACC) as

$$\text{ACC} = \frac{\sum_d (I \cdot I')}{\sqrt{[\sum_d I^2][\sum_d I'^2]}} \quad (5)$$

ACC was used to measure the degree of similarity between the original and the estimated depth map image, where I corresponds to the original and I' corresponds to the pixel information of the estimated depth map image [36]. Among

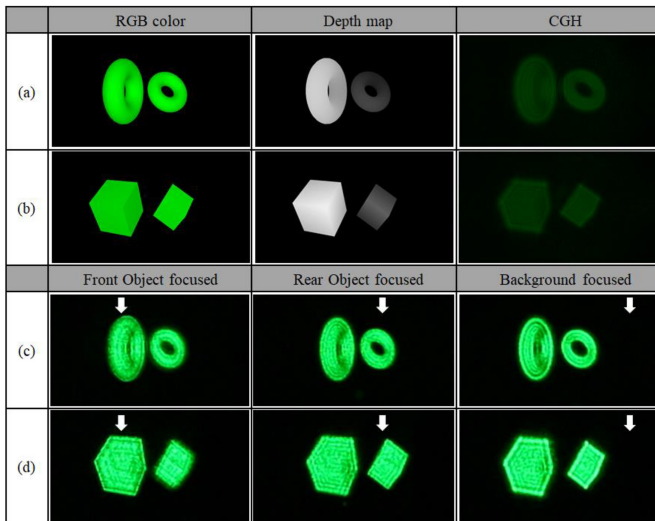


Fig. 4. Color & depth map set and CGH image at the 480th viewpoint case for a torus shape (a) and for a cube shape (b). Optically reconstructed H3D images from the CGHs for the torus shape (c) and for the cube shape (d). In each given scene captured through the DSLR camera, a white arrow marks each focused place among a pair of 3D objects and a background.

the four shapes, the shape of the sphere achieves the best performance in terms of *PSNR*, *ACC*, and *RMSE* index values.

IV. RECONSTRUCTION OF HYPER-REALITY SCENES

A. Evaluations of the Hyper-Reality Holographic 3D Display

Fig. 4 shows the display experiment results for monochromatic holographic 3D image generation. Among the 360-degree, multi-viewed (1,024 viewpoints) CGH content creation results, Figs. 4(a) and (b) show examples of torus and cube shapes at the 480th viewpoint. The first column represents the image corresponding to the green color image, the second column represents the depth map image, and the third column represents the CGH generated using the color and depth maps. Figs. 4(c) and (d) show the camera-captured photographs based on the depth of the focused position in monitoring the H3D scene reconstructed from each CGH. Each photograph was captured by observing object-centered, multi-viewed H3D image movies (30 FPS) without any background image after the laser beam with a wavelength of 532 nm was illuminated on the SLM.

For a pair of 3D objects optically reconstructed from a given CGH, each picture is recorded when the focus of the observational camera is placed on the position of the target object or background.

In observing each given scene with a digital camera, the focused object between a pair of 3D objects is indicated by the white arrow marked in Fig. 4. Here, it is observed that one target that is in focus is clear, whereas another that is out of focus is blurred. In addition, when the focus of the observation camera is placed at the plane of the depth corresponding to the background (the plane where the SLM panel is placed behind these two 3D objects), it can be observed that both objects are blurred. These observational experiments prove that the scenes reconstructed from the CGH support the accommodation effect provided by typical holographic 3D displays.

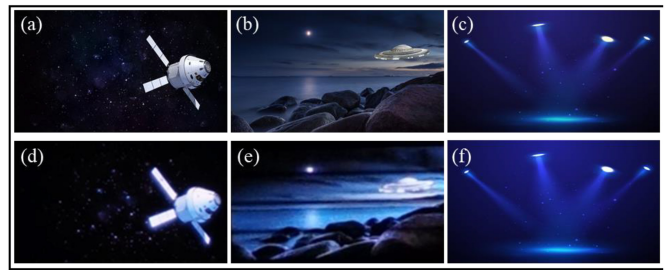


Fig. 5. 2D background images played back by the micro-display for AR floating. (a)-(c): 2D input image content. (d)-(f): Camera-captured photographs of each AR-floated background image replayed from the micro-display and image combiner by using (a)-(c).

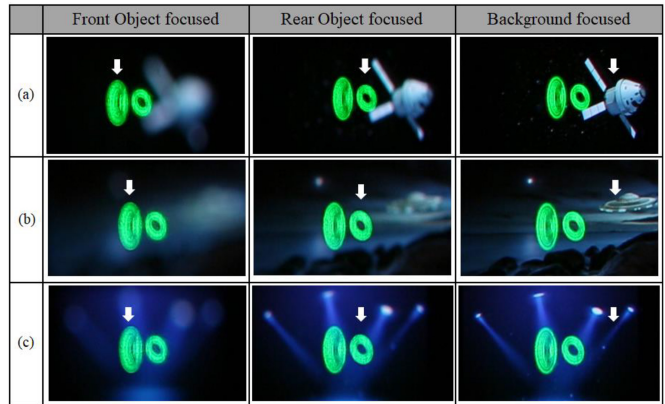


Fig. 6. Experimental results achieved using the hyper-reality holographic 3D display. (a): Torus-shaped H3D image combined with AR-floated background image (case I). (b): Torus-shaped H3D image combined with AR-floated background image (case II). (c): Torus-shaped H3D image combined with AR-floated background image (case III).

Fig. 5 shows the input images (resolution: 1280×720) as 2D background content prepared to be played back for the proposed hyper-reality system. The DSLR camera recorded images to monitor the AR-floated image outputs in real space, which were generated from the micro-display, including the image combiner. The position of the output image can be adjusted by directly setting the position of lens 2, as shown in Fig. 2(d). The 2D images shown in Fig. 4 was selected as still-life images to distinguish them from the 3D object and 2D background scene.

Fig. 6 shows the result of combining the reconstructed holographic 3D image with the background image (2D) projected through the image combiner. The reconstructed H3D image consists of a pair of torus-shaped objects made of monochromatic light (532 nm) reproduced at different depths (Fig. 2(a)) between the front and rear objects. The depth of each background image (satellite and space, UFO and beach, and stage with lighting) was formed at a plane directly adjacent to the SLM panel (Fig. 2(b)). All image quality tests for the H3D scenes were performed under the same camera shooting conditions. In Fig. 6, it can be observed that the in-focus object is clear, whereas the out-of-focus object is blurred. In addition, when the focus of the observation camera is placed at a depth corresponding to the background (the position where the SLM is placed), it can be observed that both objects are blurred. This observational experiment demonstrated that the

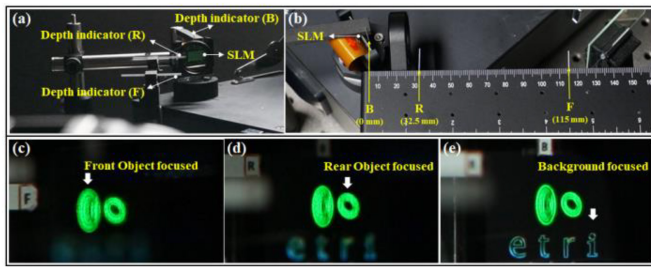


Fig. 7. Experiment to measure actual depth position values of the hyper-reality holographic scene that expresses multi-depths (representing the 3D objects and the 2D background image at different depth planes) realized on 3D physical space. (a): Optical set-up (perspective view) which shows each depth indicator holding its own character (F, R, B). (b): Optical set-up (top view) which indicates the depth position of each target to be measured by using a graduated ruler. (c)-(e): Photographs of each target (objects/background) image being camera-focused to the depth at which the target is located in the hyper-reality 3D scene.

image reconstructed from the CGH was H3D, providing an accommodation effect. Fig. 6 shows that the object or background on which the observation camera is focused is clear, whereas the object or background that is not focused by the observation camera is blurred. Here, the focused 3D object or background scene is indicated by a white arrow. Furthermore, the proposed hyper-reality holographic 3D display system was used to demonstrate the combination of still holographic 3D content with 2D background content and the combination of dynamic holographic 3D content with dynamic 2D background content in this experiment [see Supplementary Movie 1 and Supplemental File Fig. S4.–S5.].

The corresponding depth is different for each piece of content monitored through a direct observation experiment of hyper-reality holographic 3D images. Fig. 7 shows the experiment in which the actual depth value was measured for a pair of 3D objects or a 2D background reproduced at different depth positions along the z -axis in 3D real space by the hyper-reality holographic 3D display system.

To measure the actual depth at which each target represented in space is placed, we used three types of depth-marked (F, R, and B) indicators and a graduated ruler, as shown in Fig. 7, including the DSLR camera. In this experiment, the F (front focus), R (rear focus), and B (background focus) indicators were installed at the true depth positions corresponding to make each object and background clear when observing the hyper-reality 3D scene through the camera. As shown in Fig. 7, the depth position (indicator B) where the 2D background AR image is placed is $z = 0 \text{ mm}$ (SLM's surface position). The depth position (F indicator) of the front object between two reconstructed H3D objects is $z = 32.5 \text{ mm}$, and the depth position (F indicator) of the front object between them is $z = 115 \text{ mm}$. As a solution for CGH's acquisition of SW for 360-degree holographic content uploaded onto the proposed HW system, we proposed a deep learning-based MHDD model to perform high-precision depth value estimation from each missing viewpoint using multi-viewed green color and depth map images as input data. The depth map data set used to calculate CGH was extracted with ACC greater than 99.0%, as listed in Table I. The proposed CNN-based

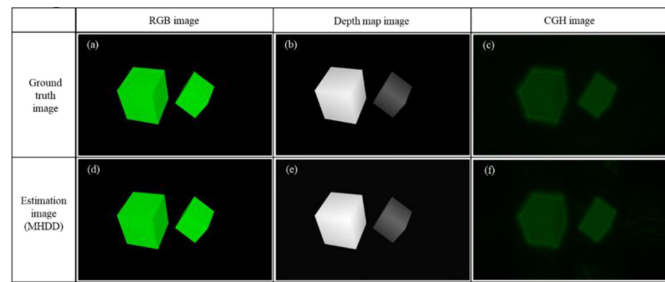


Fig. 8. Numerical observation of mono-colored (532 nm) CGH content of the torus shape. (a)-(c): Images of ground truth. (d)-(f): Images of estimation.

MHDD technique is expected to improve processing speed and high-quality CGH content generation for hyper-reality holographic NED applications. This capability indicates that the high-precision depth-map estimation algorithm developed through deep learning can be applied to high-speed, real-time computational processing for 360-degree holographic 3D content. Fig. 8 shows an example of holographic 3D images reconstructed from the 480th viewpoint using monochromatic (green) light ($\lambda = 532 \text{ nm}$). These images were obtained by the numerical simulation/computational reconstruction method of holographic 3D images using two types of CGHs: one was synthesized from the ground truth depth map and green color image, and the other from the depth map estimated by the proposed MHDD and green color image.

The visual numerical observation results confirm that the estimated holographic 3D image through the MHDD model is quite similar to the ground truth holographic 3D image.

B. Conventional and Advanced Image Combiners

The optical component used for the image combination was a commercial thin beam splitter (Edmund Optics, model: #35-946, with a thickness of 1.0 mm). This conventional thin plate had a reflection/transmission ratio of 50:50 near an incident angle of 45° in the wavelength range of 400–700 nm. Interference fringes appeared in the reconstructed H3D images passing through the commercial thin plate. Constructive interference occurs if the transmitted coherent laser beams are in phase after passing through the two surfaces of the thin-beam splitter. Assuming that the transmitted beams with the two greatest intensities contribute dominantly to the constructive interference, we can consider these two beams as follows [50]: one beam is directly transmitted through the thin plate, and the other beam is internally reflected twice before being transmitted. When the angle of the incident beam toward the beam splitter is θ , the phase difference (δ) between these two beams is given by $\delta = 2knl\cos\theta$, where n is the refractive index of the beam combiner, l is the thickness of the thin beam splitter, and $k = 2\pi n/\lambda$ is the wavenumber inside the thin plate [50], [51]. Therefore, when the coherent laser light passing through the thin plate used in this study satisfies the above-mentioned condition, an interference pattern is localized. This fringe phenomenon was confirmed by watching the video that captured the H3D images passing through the beam splitter [Supplementary Movie 3]. In addition, the brightness

of the H3D image and the projected AR background were reduced to approximately 50% by the thin beam splitter, as shown in Fig. 11(a).

To address both the interference fringe and brightness reduction problems, we chose a holographic optical element (HOE) plate as an alternative for the image combiner. The fabricated HOE sample, as shown in Fig. 11(d), is a volume Bragg grating that was holographically recorded in a reflective-type using a layer of photopolymer (model: Bayfol HX200, Covestro), a photosensitive material with a thickness of $240\ \mu\text{m}$ [52], [53]. A $170\ \mu\text{m}$ thick glass substrate was laminated with the photopolymer film. Two interference beams with incident angles of 0° and 45° were irradiated on the film from the laser (model: Cobolt's Samba $532\ \text{nm}$) used for recording, with intensities of 850 and $710\ \mu\text{W}$, respectively. Prior to applying the UV beam to cure the recording, there was a $12.5\ \text{s}$ exposure time. The experimental results of using a DSLR camera to observe holographic 3D images from thin beam splitter and the advanced HOE-based image combiner are shown in Figs. 9(b) and (c). Figs. 9(a)–(c) compare the results of the reconstructed holographic 3D image with and without the image combiner. In addition, Figs. 9(e) and (f) compares the light distribution of the reconstructed holographic 3D objects in relation to the two types of image combiners. Here, the beam profiler's instrument is the CinCam CMOS-Nano-1.001 Laser Beam Profiler from CINOGY Technologies. The beam-profiling measurements are shown in Figs. 9(d)–(f), $28\ (\text{nW}/\text{px})^2$, $16.5\ (\text{nW}/\text{px})^2$, and $23\ (\text{nW}/\text{px})^2$ were measured as the peak values, where the x-axis represents the measured pixel position and the y-axis represents the measured beam intensity. The experiments revealed that, in contrast to a clean H3D image using the HOE plate, the H3D image passing through the thin beam splitter showed a significant reduction in brightness and displayed vertical stripe noise. The photograph of Fig. 9(b) shows the vertically striped phenomenon. In addition, we observed that the brightness of the H3D image in the case of the HOE was improved by 23.2% compared with the case of the thin beam splitter. Therefore, this experimental verification confirmed that the HOE plate effectively removed interference fringes and improved brightness.

In addition, the light efficiency characteristics of the beam splitter and the HOE plate were evaluated. Fig. 10(a) shows the measurement result of light efficiency using a monochromatic laser beam ($532\ \text{nm}$) to illuminate each sample (incident beam intensity: $1.333\ \text{mW}$), with each sample rotated by 0.1° . The light efficiency is defined as the ratio of the intensity of the reflected light (in the beam splitter) or first-order diffracted light (in the reflective HOE plate) compared to the intensity of the incident light irradiated into the sample. A Gentec-EO's power-meter (model: Maestro) was used to measure the light intensity. The intensity of the reflected light at 45° was $0.6378\ \text{mW}$ for the beam splitter, and the absolute light efficiency compared to the incident beam was 47.84% . In contrast, for the reflective HOE plate, the maximum diffraction light intensity of $0.9704\ \text{mW}$ is measured at 44.8° , and the absolute diffraction efficiency value at this angle is 72.78% ;

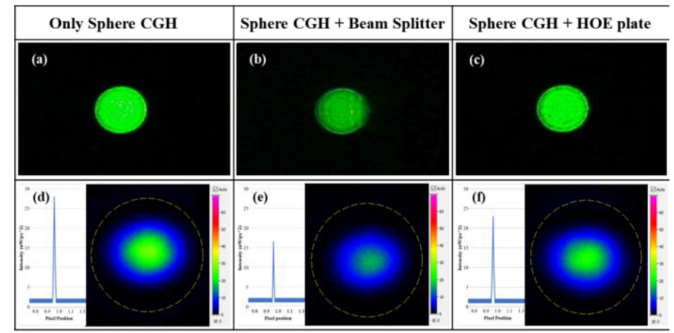


Fig. 9. Comparison of H3D image quality for a green sphere shape using the conventional thin beam splitter and the HOE plate. Camera-captured photographs of each reconstructed H3D image. (a): Case without the beam combiner. (b): Case with the thin beam splitter. (c): Case with the HOE plate. Beam intensity profile of each reconstructed H3D image. (d): Case without the beam combiner. (e): Case with the thin beam splitter. (f): Case with the HOE plate.

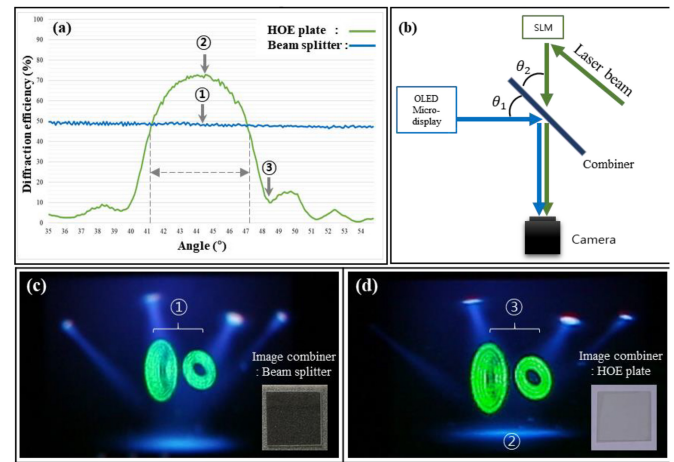


Fig. 10. Comparison of hyper-reality holographic 3D images using a thin beam splitter or a HOE plate as an image combiner. (a): light efficiency (reflectance) of the scene beam splitter and diffraction efficiency of the reflective-type HOE plate. (b): Geometric configuration for each image combiner. (c): Observation result by using the conventional thin beam splitter of which photograph is shown as inset. (d): Observation result by using the HOE plate of which photograph is shown as inset.

the diffraction efficiency of the HOE sample is improved by 24.95% compared with the maximum efficiency of the thin beam splitter. The full width at half maximum (FWHM) of the HOE diffraction efficiency curve is 6.6° (ranging from 41.1° to 47.7°). The HOE shows a larger benefit over the beam splitter in terms of both efficiency and brightness characteristics, as well as having a degree of freedom, *i.e.*, angular selectivity, in the alignment precision of the reflected light as much as the value of FWHM, when it is used as an image combiner.

Furthermore, we realized that the replacement of such a reflective HOE plate results in brightness improvement when an alternative component is applied to the proposed hyper-reality 3D display system. From the measurements, it is confirmed that the optimal conditions for the beam splitter are $\theta_1 = \theta_2 = 45^\circ$, whereas $\theta_1 = 46.3^\circ$ and $\theta_2 = 43.7^\circ$ for the HOE plate. Fig. 10(b) shows a schematic of the geometric configuration corresponding to the aforementioned conditions of each image combiner derived from Fig. 10(a). The hyper-reality holographic 3D images obtained under these

conditions corresponding to each case are shown in Figs. 10(c) and (d), respectively.

Each reconstructed H3D image (content: twin torus shape) was displayed in combination with the AR background image (content: illumination stage). The observation camera focuses on the rear-positioned objects of the two 3D objects. The DSLR camera captures the image in Fig. 10(c) when the H3D light and background AR light pass through the thin beam splitter under condition ①, as shown in Fig. 10(a). Fig. 10(d) is an image captured by the same camera when the H3D light is passing via the HOE plate on condition ③ indicated in Fig. 10(a) and the background AR light is passing via the HOE plate on condition ②, as shown in Fig. 10(a). These configuration characteristics lead to improved image brightness in the proposed NED system because the HOE plate is superior to the conventional beam splitter in transmitting H3D light and reflecting background AR light. In addition, an examination of the background image in detail (Fig. 10(d)) shows that color separation occurs around each white illumination light source. This phenomenon occurs because the HOE fabricated in this study is a reflective-type volume Bragg grating produced by monochromatic light (recording light wavelength: 532 nm), and diffraction in blue or red light is different from that in green light (recording wavelength). The diffraction angle θ reflected at the lattice planes of the volume Bragg grating (lattice plane period: p) depends on the wavelength (λ) according to the equation $\sin\theta = \lambda/2p$ [50], [53]. Therefore, when white light is incident on the volume hologram sample, the degree of matching of the red and blue colors is inadequate, with the green color being centered.

C. Reconstruction of Multi-Colored, Hyper-Reality Video Scenes

The strategy for obtaining and displaying multi-colored CGHs is as follows: A depth map is estimated and extracted through MHDD using only one representative color (green) among three elementary colors representing the object's scene. Based on the predicted depth map, three CGHs corresponding to the three primary colors were synthesized using the FFT algorithm. Finally, multicolored CGHs are uploaded to be displayed on the proposed NED module, where each color's H3D image is spatially matched through a mutual position adjustment process in real space. Thus, we can observe a complete H3D image reconstructed from these multicolored CGHs. Fig. 11 shows examples of hyper-reality 3D movies, which we displayed using this strategy on the proposed hyper-reality holographic 3D device by combining a holographic 3D video reconstructed from two-colored CGHs with an AR background video. After each torus-shaped green (532 nm) and blue (470 nm) color CGH were reconstructed within the same depth range, they were adjusted to achieve mutual spatial matching in real space, so that a cyan-colored H3D image could be displayed, as shown in Fig. 11(c). In addition, cube-shaped CGHs of green (532 nm) and red (633 nm) colors were reconstructed to be spatially matched, so that a yellow-colored H3D image could be observed, as shown

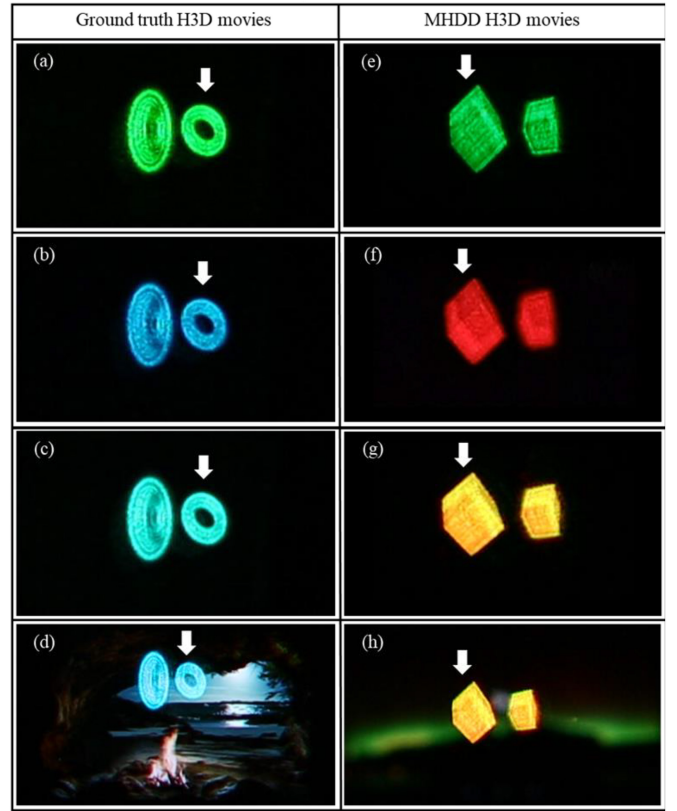


Fig. 11. Optical observation of the multi-colored, hyper-reality video scene obtained by combining both the H3D movie reconstructed from multi-colored CGHs with the AR 2D background movie. (a): H3D movie of torus shape in 532 nm where the rear green torus is focused. (b): H3D movie of torus shape in 470 nm where the rear blue torus is focused. (c): H3D movie of torus shape in both 532 nm and 470 nm where the rear cyan torus is focused. (d): Example to demonstrate the 360-degree, hyper-reality video scene by combining H3D movies with the AR 2D background movie where the rear cyan torus is focused. (e): H3D movie of cube shape in 532 nm where the front green cube is focused. (f): H3D movie of cube shape in 633 nm where the front red cube is focused. (g): H3D movie of cube shape in both 532 nm and 633 nm where the front yellow cube is focused. (h): Example to demonstrate the 360-degree, hyper-reality video scene by combining H3D movies with the AR 2D background movie where the front yellow cube is focused.

in Fig. 11(g). The captured scenes displayed in moving pictures by the proposed hyper-reality holographic 3D system are shown in Figs. 11(d) and (h) (360-degree, fully-viewed video scenes with 30 FPS are shown in Supplementary Movie 2). From the experimental demonstration shown in Fig. 11, we verified that the proposed hyper-reality holographic 3D system could be displayed in the complete video mode by using multi-colored CGH movies as well as the monochromatic CGH movie, each combined with the 2D background AR movie.

V. CONCLUSION AND OUTLOOK

In this study, a new 3D/2D hybrid display system for hyper-reality, holographic NED, was proposed by merging a H3D image reconstructed from the LCoS-SLM with a spatially projected AR background image displayed from a 2D microdisplay. Unlike conventional display systems that stack panels of the same type the proposed hybrid NED system was experimentally implemented using a novel optical integration

approach of heterogeneous panel systems with an ultrathin HOE-based image combiner. This enabled us to demonstrate a hyper-realistic holographic scene that maximized immersion and 3D effects, even while watching a clearly combined movie without interference between the H3D objects and AR background content. We used a physical ruler-based spatial depth measurement method to evaluate the 3D image quality of the proposed hyper-reality 3D display system and validated that image appearing in the real 3D space are displayed at different depths, as evidence for accommodation cues. The benefits of our proposed display system are as follows: First, 360-degree, multi-viewed CGH content reproduced at different depths can be clearly viewed with high image quality, and immersive AR background content projected in real 3D space can be watched simultaneously without cross-talk. Second, during the CHG synthesis step, the depth of the real H3D object can be easily adjusted. The depth position and size of images in the real 3D space for AR background content can be freely controlled by the geometric arrangement of the optical path in the proposed system, that is, the direct interval setting between the microdisplay and the lens located in front of it. Third, high-quality CGH data can be generated from the FFT algorithm using the given color image and depth map estimated from the proposed deep learning method, MHDD, which is suitable for high-speed holographic content processing. Because of these unique features, we believe that the proposed hybrid display system can be used with a real-time interaction platform based on both deep-learning SW and newly emerging metaverse technologies. Finally, the accommodation effect and the free depth control characteristics demonstrated by the proposed system enable users to watch hyper-realistic 3D video content comfortably, indicating that eye fatigue issues such as copipia (eye fatigue) and dizziness can be overcome even when watching metaverse content for an extended period. Furthermore, a reflective-type HOE plate was designed and used as an alternative to the existing commercial thin beam splitter as the image combiner to improve the interference fringes that usually appear in generating hyper-reality holographic scenes. We demonstrated that this new image combiner could improve the clarity and brightness of the H3D image as well as remove the interference fringe noise. However, because the reflective-type HOE used in the current study as the image combiner was fabricated under the condition of a monochromatic laser beam (532 nm), there is a drawback in that a completely full-color-matched scene was not achieved owing to the wavelength selectivity problem, including the color mismatch issue in the background AR image projected from the monochromatic HOE plate.

In future studies, it will be necessary to solve the problem of color separation by designing and fabricating a full-color HOE plate for an RGB image combiner. In addition, an improved implementation of a wearable form factor through miniaturization and light weight of the proposed holographic NED system is required, and further investigation of real-time interaction verification using the implemented wearable hybrid display is required. As deep learning-based hyper-reality metaverse content is replayed to calculate the high-precision/high-speed

TABLE II
DEFINITION OF TERMS

No.	Abbreviation	Definition of Terms
1	CGH	Computer-Generated Holography
2	NED	Near-eye Display
3	VR	Virtual Reality
4	AR	Augmented Reality
5	HR	Holographic Reality
6	SLM	Spatial Light Modulator
7	AC	Accommodation Convergence
8	FFT	Fast Fourier Transform
9	H3D	Holographic 3D
10	HDD	Holographic Dense Depth
11	MHDD	Monochromatic-HDD
12	HOE	Holographic Optical Element
13	FWHM	Full Width at Half-Maximum
14	SSIM	Structural Similarity Index Measure
15	MSE	Mean Squared Error
16	ACC	Accuracy
17	PSNR	Peak Signal-to-Noise Ratio

CGH data, additional research on human factors is needed to minimize the user's eye fatigue problems and verify their effect on eye comfort/stability. These study findings will contribute to the development of commercially available hyper-realistic holographic NEDs that may be ideal for the metaverse and can be used by everyone.

APPENDIX

The exact meanings of the abbreviations used in this manuscript are summarized in Table II.

REFERENCES

- [1] D. Gabor, "A new microscopic principle," *Nature*, vol. 161, pp. 777–778, May 1948.
- [2] P. Blanche, "Holography, and the future of 3D display," *Light Adv. Manuf.*, vol. 2, no. 4, pp. 1–14, Dec. 2021.
- [3] S. Park and Y. Kim, "A metaverse: Taxonomy, components, applications, and open challenges," *IEEE Access*, vol. 10, pp. 4209–4251, Jan. 2022.
- [4] "Apple Vision Pro." Accessed: Jun. 5, 2023. [Online]. Available: <https://www.apple.com/apple-vision-pro/>
- [5] J. Son, H. Lee, B. Lee, and K. Lee, "Holographic and light-field imaging as future 3-D displays," *Proc. IEEE*, vol. 105, no. 5, pp. 789–804, May 2017.
- [6] J. Dionisio, W. Burns, and R. Gilbert, "3D virtual worlds and the metaverse: Current status and future possibilities," *ACM Comput. Surveys*, vol. 45, no. 3, pp. 1–38, Jun. 2013.
- [7] H. Yoo, J. Jang, H. Oh, and I. Park, "The potentials and trends of holography in education: A scoping review," *Comput. Educ.*, vol. 186, Sep. 2022, Art. no. 104533.
- [8] B. Xie et al., "A review on virtual reality skill training applications," *Front. Virtual Real.*, vol. 2, pp. 1–19, Apr. 2021.
- [9] C. Chang, K. Bang, G. Wetzstein, B. Lee, and L. Gao, "Toward the next-generation VR/AR optics: A review of holographic near-eye displays from a human-centric perspective," *Optica*, vol. 7, no. 11, pp. 1563–1578, Nov. 2020.
- [10] J. Wang, X. Xiao, H. Hua, and B. Javidi, "Augmented reality 3D displays with micro integral imaging," *IEEE Display Technol.*, vol. 11, no. 11, pp. 889–893, Nov. 2015.
- [11] B. Lee, H. Lee, W. Son, J. Son, H. Heo, and T. Venkel, "Monocular accommodation in the light field imaging," *IEEE Trans. Broadcast.*, vol. 68, no. 2, pp. 397–406, Jun. 2022.
- [12] T. Zhan, K. Yin, J. Xiong, Z. He, and S.-T. Wu, "Augmented reality and virtual reality displays: perspectives and challenges," *Iscience*, vol. 23, no. 8, Aug. 2020, Art. no. 101397.
- [13] P. W. M. Tsang, T.-C. Poon, and Y. M. Wu, "Review of fast methods for point-based computer-generated holography," *Photon. Res.*, vol. 6, no. 9, pp. 837–846, Sep. 2018.

- [14] M. H. Eybposh, N. W. Caira, M. Atisa, P. Chakravarthula, and N. C. Pegard, "DeepCGH: 3D computer-generated holography using deep learning," *Opt. Exp.*, vol. 28, no. 18, pp. 26636–26650, Aug. 2020.
- [15] H. Lee, W. Son, M. Kim, Y. Yoon, and M. Yoon, "Near-eye holographic 3D display and advanced amplitude-modulation encoding scheme for extended reality," *Appl. Sci.*, vol. 13, no. 6, pp. 1–16, Mar. 2023.
- [16] Z. Huang, D. L. Marks, and D. R. Smith, "Out-of-plane computer-generated multicolor waveguide holography," *Optica*, vol. 6, no. 2, pp. 119–124, Jan. 2019.
- [17] J. Park and B. Lee, "Holographic techniques for augmented reality and virtual reality near-eye displays," *Light Adv. Manuf.*, vol. 3, no. 9, pp. 1–14, Feb. 2022.
- [18] J. Dong et al., "High-speed real 3D scene acquisition and 3D holographic reconstruction system based on ultrafast optical axial scanning," *Opt. Exp.*, vol. 31, no. 13, pp. 21721–21730, Jun. 2023.
- [19] T. Kozacki, "On resolution and viewing of holographic image generated by 3D holographic display," *Opt. Exp.*, vol. 18, no. 26, pp. 27118–27129, Dec. 2010.
- [20] G. D. Love, D. M. Hoffman, P. J. W. Hands, J. Gao, A. K. Kirby, and M. S. Banks, "High-speed switchable lens enables the development of a volumetric stereoscopic display," *Opt. Exp.*, vol. 17, no. 18, pp. 15716–15725, Aug. 2009.
- [21] T. Zhan, Y.-H. Lee, and S.-T. Wu, "High-resolution additive light field near-eye display by switchable Pancharatnam–Berry phase lenses," *Opt. Exp.*, vol. 26, no. 4, pp. 4863–4872, Feb. 2018.
- [22] K. Yin et al., "Advanced liquid crystal devices for augmented reality and virtual reality displays: principles and applications," *Light Sci. Appl.*, vol. 11, no. 1, pp. 1–22, May 2022.
- [23] X. Xia et al., "Towards a switchable AR/VR near-eye display with accommodation-vergence and eyeglass prescription support," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 11, pp. 3114–3124, Nov. 2019.
- [24] Z. Wang, X. Zhang, G. Lv, Q. Feng, H. Ming, and A. Wang, "Hybrid holographic Maxwellian near-eye display based on spherical wave and plane wave reconstruction for augmented reality display," *Opt. Exp.*, vol. 29, no. 4, pp. 4927–4935, Feb. 2021.
- [25] W.-K. Lin, W.-T. Liu, B.-S. Lin, and W.-C. Su, "A hybrid display with 2D/3D image based on static hologram and LCD panel," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE-TW)*, 2017, pp. 47–48.
- [26] J. Zhao, Q. Ma, J. Xia, J. Wu, B. Du, and H. Zhang, "Hybrid computational near-eye light field display," *IEEE Photon. J.*, vol. 11, no. 1, pp. 1–10, Feb. 2019.
- [27] D. Kim, S. Lee, S. Moon, J. Cho, Y. Jo, and B. Lee, "Hybrid multi-layer displays providing accommodation cues," *Opt. Exp.*, vol. 26, no. 13, pp. 17170–17184, Jun. 2018.
- [28] H.-H. Lee, P.-J. Huang, J.-Y. Wu, P.-Y. Hsieh, and Y.-P. Huang, "A 2D/3D hybrid integral imaging display by using fast switchable hexagonal liquid crystal lens array," in *Proc. 3D Imag. Vis. Display*, vol. 10219, 2017, pp. 1–7.
- [29] A. Maimone, A. Georgiou, and J. Kollin, "Holographic near-eye displays for virtual and augmented reality," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–16, Jul. 2017.
- [30] J. Xiong, E.-L. Hsiang, Z. He, T. Zhan, and S.-T. Wu, "Augmented reality and virtual reality displays: Emerging technologies and future perspectives," *Light Sci. Appl.*, vol. 10, no. 1, pp. 1–30, Oct. 2021.
- [31] Y. Arima, K. Matsushima, and S. Nakahara, "Hybrid CGH by digitized holography: CGH for mixed 3D scene of virtual and real objects," in *Proc. Digit. Holography 3D Imag. Opt. Soc. America (DWOC)*, 2011, pp. 1–3.
- [32] J. Roh et al., "Full-color holographic projection display system featuring an achromatic Fourier filter," *Opt. Exp.*, vol. 25, no. 13, pp. 14774–14782, Jun. 2017.
- [33] N. Padmanaban, Y. Peng, and G. Wetzstein, "Holographic near-eye displays based on overlap-add stereograms," *ACM Trans. Graph.*, vol. 38, no. 6, pp. 1–13, Nov. 2019.
- [34] J. Li, Q. Smithwick, and D. Chu, "HoloBricks: Modular coarse integral holographic displays," *Light Sci. Appl.*, vol. 11, no. 57, pp. 1–15, Mar. 2022.
- [35] Y. Peng, S. Choi, J. Kim, and G. Wetzstein, "Speckle-free holography with partially coherent light sources and camera-in-the-loop calibration," *Sci. Adv.*, vol. 7, no. 46, pp. 1–6, Nov. 2021.
- [36] S. Choi, J. Kim, Y. Peng, and G. Wetzstein, "Optimizing image quality for holography near-eye displays with Michelson holography," *Optica*, vol. 8, no. 2, pp. 143–146, Jan. 2021.
- [37] B. Lee, D. Kim, S. Lee, C. Chen, and B. Lee, "High-contrast, speckle-free, true 3D holography via binary CGH optimization," *Sci. Rep.*, vol. 12, no. 2811, pp. 1–12, Feb. 2022.
- [38] L. Shi, B. Li, C. Kim, P. Kellnhofer, and W. Matusik, "Towards real-time photorealistic 3D holography with deep neural networks," *Nature*, vol. 591, pp. 234–239, Apr. 2021.
- [39] D. Yang et al., "Diffraction-engineered holography: Beyond the depth representation limit of holographic displays," *Nat. Commun.*, vol. 13, no. 6012, pp. 1–11, Oct. 2022.
- [40] H. Kim et al., "High-precision depth map estimation from missing viewpoints for 360-degree digital holography," *Appl. Sci.*, vol. 12, no. 19, pp. 1–20, Sep. 2022.
- [41] M. Chen, Y. Jin, T. Goodall, X. Yu, and A. C. Bovik, "Study of 3D virtual reality picture quality," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 1, pp. 89–102, Jan. 2020.
- [42] M. Kozhevnikov and R. P. Dhond, "Understanding immersivity: Image generation and transformation processes in 3D immersive environments," *Front. Psychol.*, vol. 3, no. 284, pp. 1–10, Aug. 2012.
- [43] Y. Ji et al., "Virtual experiment of optical spatial filtering in MATLAB environment," in *Proc. 14th Conf. Educ. Training Opt. Photon. (ETOP)*, vol. 10452, 2017, pp. 1–7.
- [44] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.
- [45] S. Anstis, "The purkinje rod-cone shift as a function of luminance and retinal eccentricity," *Vis. Res.*, vol. 42, no. 22, pp. 2485–2491, Oct. 2002.
- [46] "Maya autodesk." 2018. [Online]. Available: <https://www.autodesk.com/products/maya>
- [47] H. Kim, B. Yang, and B. Lee, "Iterative Fourier transform algorithm with regularization for the optimal design of diffractive optical elements," *J. Opt. Soc. Amer. A*, vol. 21, no. 12, pp. 2353–2365, Dec. 2004.
- [48] W. H. Lee, "Sampled fourier transform hologram generated by computer," *Appl. Opt.*, vol. 9, no. 3, pp. 639–643, Mar. 1970.
- [49] M. Talo, "Automated classification of histopathology images using transfer learning," *Artif. Intell. Med.*, vol. 101, no. 101743, pp. 1–8, Oct. 2019.
- [50] M. Born and E. Wolf, *Principles of Optics*, 7th ed. Cambridge, U.K.: Cambridge Univ. Press, 1999, pp. 286–732.
- [51] L. A. Aleman-Castaneda, B. Piccirillo, E. Santamato, L. Marrucci, and M. A. Alonso, "Shearing interferometry via geometric phase," *Optica*, vol. 6, no. 4, pp. 396–399, Mar. 2019.
- [52] C. Yu, Y. Peng, Q. Zhao, H. Li, and X. Liu, "Highly efficient waveguide display with space-variant volume holographic gratings," *Appl. Opt.*, vol. 56, no. 34, pp. 9390–9397, Nov. 2017.
- [53] Y. Yang, L. Deng, L. Zhu, C. Gu, and L. Xu, "Resolution-preserving passive 2D/3D convertible display based on holographic optical elements," *Opt. Exp.*, vol. 30, no. 8, pp. 13356–13371, Apr. 2022.



Hyoung Lee received the M.S. degree in signal processing and multimedia engineering from Daegu University, South Korea, in 2010, the Doctoral degree in electrical and electronic engineering from Yonsei University in 2013, and the Ph.D. degree in science and engineering from Shimane University, Japan, in 2019. He is currently a Senior Researcher with Electronics and Telecommunications Research Institute, Daejeon, South Korea. His interests are 3-D image displays, human factor in SMV, plenoptic image, and laser-based optical instrumentations and

measurements. In particular, he is concentrating on research on holograms and AR glass related to metaverse.



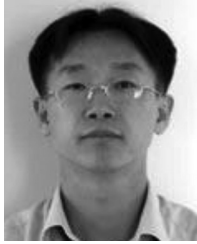
Hakdong Kim received the B.S. degree in computer science from Kyungsoong University, South Korea, in 2016. He is currently pursuing the integrated M.S./Ph.D. degree with Sejong University, Seoul, South Korea. From 2016 to 2017, he worked as a Quality Assurance Engineer with Telecommunications Technology Association. His interests are artificial intelligence (machine learning and deep learning), natural language processing, and pattern recognition (holography and image processing).



Taeheul Jun received the B.S. degree in digital contents from Sejong University, Seoul, South Korea, in 2021, where he is currently pursuing the M.S. degree. His interests are artificial intelligence (machine learning and deep learning), and pattern recognition (holography, image processing, and audio processing).

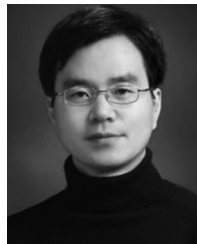


Cheongwon Kim received the B.S. and M.S. degrees in computer science from Southern Illinois University, USA, in 1987 and 1990, respectively, and the Ph.D. degree in computer informatics and science from Syracuse University, USA, in 2000. From 2001 to 2002, he worked as a CEO with Artificial Intelligence Tech, USA. From 2002 to 2003, he worked as a Professor with Ajou University, Suwon, South Korea. Since 2003, he has been a Professor with Sejong University, Seoul, South Korea. His interests are artificial intelligence (machine learning and deep learning), bioinformatics (drug repurposing and protein engineering), pattern recognition (holography, image processing, and audio processing), and forecasting models (stock forecasting and environment forecasting).



Wookho Son received the B.S. degree in computer science from Yonsei University, Seoul, South Korea, in 1987, and the M.S. and Ph.D. degrees in computer science from Texas A&M University at College Station, College Station, TX, USA, in 1996 and 2001, respectively. He joined as a Senior Member of Engineering Staff with Electronics and Telecommunications Research Institute, Daejeon, South Korea, in 2001, where he is currently a Principal Member of Engineering Staffs of Communication and Media Research Laboratories.

His research interests include virtual and augmented reality, robotics, digital holography, physically based dynamic simulation, and haptic interaction.



MinSung Yoon received the D.Phil. degree in physics from the University of Oxford. He worked as a Project Leader of Holographic 3-D Display with LG Display R&D Center from 2009 to 2014. He is currently a Principal Researcher with Electronics and Telecommunications Research Institute for developing holographic optical devices and 3-D imaging. His scientific research concerns the precise expression of RGB color images using deep-learning-based image process and digital holography.