Perceptual Assessment of Image and Depth Quality of Dynamically Depth-Compressed Scene for Automultiscopic 3D Display

Yamato Miyashita, Yasuhito Sawahata, and Kazuteru Komine

Abstract—This article discusses the depth range which automultiscopic 3D (A3D) displays should reproduce for ensuring an adequate perceptual quality of substantially deep scenes. These displays usually need sufficient depth reconstruction capabilities covering the whole scene depth, but due to the inherent hardware restriction of these displays this is often difficult, particularly for showing deep scenes. Previous studies have addressed this limitation by introducing *depth compression* that contracts the scene depth into a smaller depth range by modifying the scene geometry, assuming that the scenes were represented as CG data. The previous results showed that reconstructing only a physical depth of 1 m is needed to show scenes with much deeper depth and without large perceptual quality degradation. However, reconstructing a depth of 1 m is still challenging for actual A3D displays. In this study, focusing on a personal viewing situation, we introduce a dynamic depth compression that combines viewpoint tracking with the previous approach and examines the extent to which scene depths can be compressed while keeping the original perceptual quality. Taking into account the viewer's viewpoint movements, which were considered a cause of unnaturalness in the previous approach, we performed an experiment with an A3D display simulator and found that a depth of just 10 cm was sufficient for showing deep scenes without inducing a feeling of unnaturalness. Next, we investigated whether the simulation results were valid even on a real A3D display and found that the dynamic approach induced better perceptual quality than the static one even on the real A3D display and that it had a depth enhancing effect without any hardware updates. These results suggest that providing a physical depth of 10 cm on personalized A3D displays is general enough for showing any deeper 3D scenes with appealing subjective quality.

Index Terms—Compression technologies, depth cues, perception and psychophysics, volumetric

1 Introduction

A key feature of automultiscopic 3D (A3D) displays is that they present a 3D scene as if it naturally existed in a physical space without any need for special glasses. The displays provide depth cues of binocular disparities and motion parallax, which is induced by a viewing point movement while looking at the presented 3D images. In addition, A3D displays using light field technologies like integral photography [1] avoid the vergence-accommodation conflict (VAC) [2] that brings visual discomfort [3], which is one of the issues with stereoscopic vision using special glasses. These characteristics can also be achieved by volumetric displays [4] that directly show voxels on a physical space by projecting light into physical objects [5], [6] or by using light emission [7], [8]. In this study, we distinguish A3D displays from autostereoscopic displays [9] that project exactly two views, one for each eye. A3D displays have ideal characteristics for 3D visualization while the stereoscopic and autostereoscopic displays cannot provide them in theory.

• *The authors are with Japan Broadcasting Corporation, Setagaya City, Tokyo 157-8510, Japan. E-mail: {miyashita.y-fc, sawahata.y-jq, komine.k-cy}@nhk.or.jp.*

Manuscript received 8 June 2021; revised 19 Jan. 2022; accepted 28 Jan. 2022. Date of publication 7 Feb. 2022; date of current version 8 May 2023. (Corresponding author: Yamato Miyashita.) Recommended for acceptance by O. Bimber. Digital Object Identifier no. 10.1109/TVCG.2022.3148419

However, A3D displays have difficulty showing scenes with great depth, such as natural landscapes or field sports, although current 2D displays like smartphones, TVs, and cinema screens can often show them without artifacts such as blur. In an integral imaging display, objects reconstructed far from the display plane suffer from intolerable blur because the spatial frequency of the reconstructed image can be kept maximum only in a range near the display plane [10] (Fig. 2). In this study, we refer to the depth range where objects can be shown with the maximum spatial resolution, i.e., the depth of focus of the display, as the *depth reconstruction range*. In fact, even with state-of-the-art integral imaging displays in which pixel density and viewing angles are enhanced by parallel projection of multiple elemental images [11], the reconstruction ranges stay within an approximate depth of only a dozen centimeters. A direct approach to increase the depth reconstruction range is to develop a flat display with denser pixels, but it is still challenging to obtain the sufficient depth reconstruction range. Similarly, volumetric displays [5], [6], [7], [8] have the same limitation on their depth reconstruction ranges, as their displayable areas are physically bound by the size of their equipment or the room. Enlarging the areas requires significant hardware updates and/or construction work, which makes it difficult to expect home use with these technologies.

Depth compression has been proposed as a method to show large 3D scenes on a 3D display with a smaller depth reconstruction range. The method contracts the entire scene depth into a limited depth range by modifying 3D geometries, e.g.,

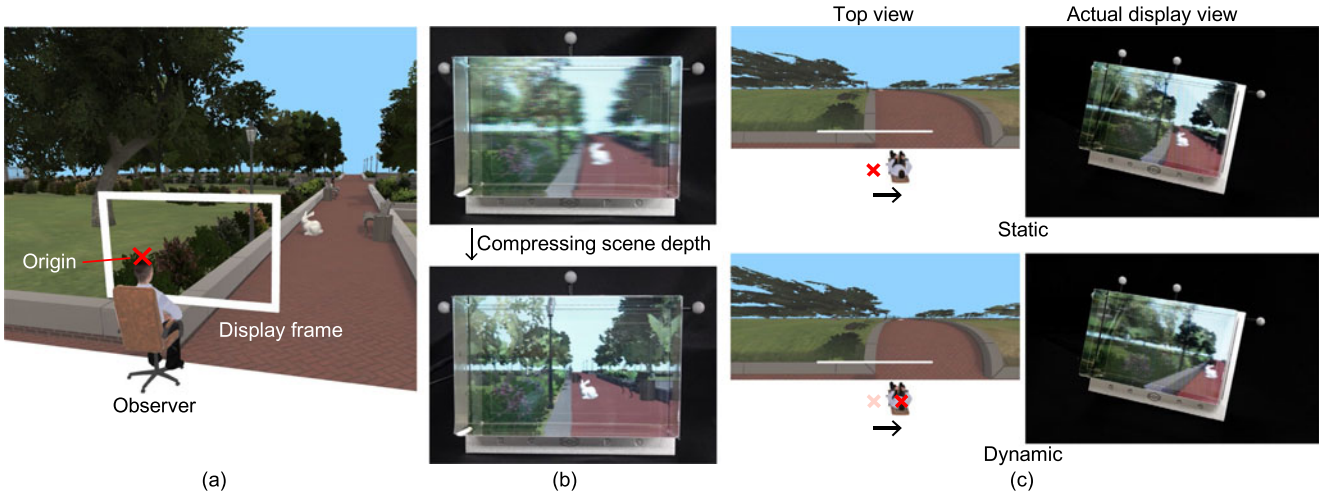

Fig. 1. Overview of extending depth reconstruction capability. (a) Relationship between positions of original scene and an observer. Watching the scene on an automultiscopic display corresponds to the observer watching the scene through the display frame while assuming that he/she is in the scene. (b) Upper image shows low-quality visualization with the original scene on the automultiscopic display viewed from the origin (at the red cross in (a)). Because of a lack of depth reconstruction capability, objects at a deep depth are observed with inevitable blur. Lower image shows this blur suppressed by depth compression. (c) Comparison of scene views generated with static and dynamic depth compression methods, assuming an oblique viewing position, i.e., shifted position from the origin. Upper-right column shows the distorted appearance of the depth-compressed scene generated by the static approach. Note that the walkway is originally straight, but the depth-compressed one looks curvy. Lower-right column shows the undistorted appearance of the depth-compressed scene generated by the dynamic approach.

translating the vertices of objects in a scene. Although the 3D geometries are distorted, they are not always perceived as they are due to the inaccuracy of human visual perception, which estimates the visual world from incomplete visual cues. The previous study [12] showed that scenes with almost infinite depth can be shown in a depth range of 1 m without inducing feelings of unnaturalness when scene depth is compressed *nonlinearly*, where the depth of far objects is contracted more than that of near objects.

However, more efficient depth compression is needed because a depth of 1 m is still difficult to reconstruct even for state-of-the-art light field technologies. Depth compression assumes a specific optimal viewing position, referred to as an *origin* of the depth compression coordinates in this study (corresponds to the red cross in Figs. 1a and 1c), where a viewer does not find any geometry manipulations applied to the scene being displayed (Fig. 1b lower). In the previous study, the origin was fixed at a typical position that corresponded to the standard viewing position, e.g., 1.5 times further away from the display center than the display height. The distortion becomes more noticeable when observing the scene further away from the origin. Because the presentation of such motion parallax is a major characteristic of A3D displays, supporting views from various viewing positions in the depth compression should unlock the potential of A3D displays.

In this study, focusing on a viewing condition that assumes just a single viewer, we investigate the required depth range for showing large 3D scenes on A3D displays with a dynamic depth compression method in which the origin in the depth compression is updated according to the viewer's viewing position in real time. Such a personal viewing condition is typical when viewing a small screen device, such as a tablet or smartphone. The real-time updates of the origin of the depth compression are computationally expensive, but we handle this issue by conducting all the computations for the geometry manipulations with a vertex shader in parallel. Thus, we can expect a more efficient use of the limited depth resources on a display.

We used an A3D display simulator to estimate the required depth range for showing 3D scenes with sufficient quality under the personal viewing condition; the position of the display screen can be handled by both hands, as when using a tablet. The use of actual A3D displays in the estimation is inappropriate because they have difficulty showing 3D scenes with substantial depth without quality degradation, as aforementioned and shown in Fig. 1b. Alternatively, we used the simulator that presented binocular disparities and motion parallax using the combination of a stereoscopic display and a motion tracker.

The required depth range was estimated by subjective evaluation experiments, following a modified version of the double stimulus impairment scale (DSIS) method [13], in which participants rated the depth-compressed scenes compared to the originals. From the results, we obtained the relationships between perceived quality (naturalness) and depth ranges in which the scenes were compressed. With these relationships, we obtained the required depth ranges that satisfy an acceptable level of quality.

We also investigated whether the findings obtained with the simulator were also valid on a real A3D display. Participants chose the superior expression in terms of naturalness from two presented stimuli generated by using the static and dynamic depth compression methods. The amount of depth of the presented scenes was chosen on the basis of the results obtained with the simulator experiment. In addition, to determine whether viewers actually felt a greater depth in the depth-compressed scenes than the optically presented depth, we asked them to estimate the amount of depth. On the basis of the additional experiments, we found that the results obtained using the simulator experiments were still valid even with real A3D displays and that the dynamic approach works like a virtual extender of the depth reconstruction capability.

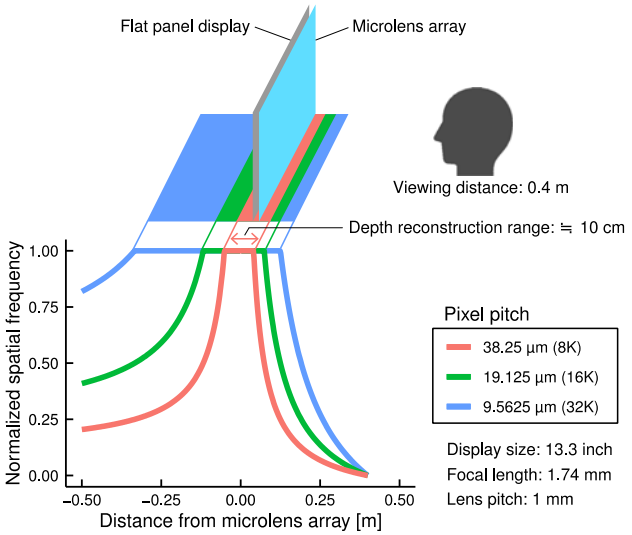

Fig. 2. Hardware-based extension of a limited depth reconstruction range in integral imaging by reducing pixel pitch of a flat panel display. Spatial frequency was divided by a maximum value for normalization.

In Section 2 of this paper, we present related works and discuss the relationship with conventional works. In Section 3, we introduce details of a dynamic depth compression approach. In Sections 4 and 5, we describe our experiments, which used simulator and real display, respectively. We discuss the results of these two experiments in Section 6 and conclude in Section 7 with a brief summary.

This study provides the following contributions.

- We found that the required depth range for showing dynamically depth-compressed 3D scenes without feelings of unnaturalness is just 10 cm.
- We obtained the finding through the assessment of the image quality of scenes with significant depth, which are infeasible to show on an actual A3D display, by using the A3D display simulator.
- The depth compression function was optimized in real time for the position of a viewer in the dynamic approach.
- We confirmed that the dynamic depth compression showed better quality in terms of naturalness and perceived depth than that provided with the static approach.

2 Related Work

To the best of our knowledge, no other studies have investigated the required depth for showing dynamically depth-compressed 3D scenes considering a viewing position without inducing a sense of unnaturalness. Here, we review the depth manipulation techniques used for improving image quality on 3D displays and discuss the differences in evaluation metrics and conditions.

2.1 Depth Manipulation Techniques

Contracting a scene depth into a shallower depth range plays important roles in 3D displays. In stereoscopic displays, adjusting the binocular disparities to within a small range is indispensable for suppressing diplopia (double vision) or the vergence-accommodation conflict that cause visual discomfort and fatigue [3], [14], [15], [16]. Since differences of the display size or viewing distance affect the amount of the disparity, scene depths optimized for a certain 3D display are often remapped for other displays, so-called content retargeting [17], e.g., a content for a large cinema screen is converted into that for a small home TV screen. For example, Lang *et al.* [18] proposed an automated nonlinear disparity remapping method using saliency estimation based on a sparse set of stereo correspondences. To enhance the viewer's depth discrimination performance compared to fixed stereoscopic parameters, Kulshreshth and Viola [19] proposed an adaptive depth manipulation method using eye tracking data. Kellnhofer *et al.* [20] proposed a gaze-driven depth manipulation that gradually adjusts depth so that it remains unnoticeable. They also proposed a depth reallocation technique in which a disparity in regions with large motion parallax was reallocated to static parts of the scene so that the perceived depth was maximized [21]. Similar to stereoscopic displays, A3D displays require depth compression due to their limited depth of field. Zwicker *et al.* [22] proposed a depth adjustment method to present deep scenes in a shallow depth of field on A3D displays by arranging multiple view inputs. Masia *et al.* [23] introduced depth remapping between light field displays with an optimization framework in which parameters of disparity manipulation were optimized so the predicted image quality and depth impression were maximized. Adhikarla *et al.* [24] proposed a real-time depth remapping for all-in-focus rendering of light field displays while preserving the 3D appearance of the salient region of a scene.

Viewpoint tracking is beneficial for improving image quality on 3D displays. Head-tracked displays (HTDs) have been proposed for adding motion parallax into stereoscopic displays [25], [26], [27]. It was reported that the addition of motion parallax improved 3D perception in a variety of tasks [28], [29], [31]. Viewpoint tracking has also been utilized on A3D displays for saving rendering resources [32] or extending a limited viewing angle [33]. In this study, we utilize measured viewpoints for updating an origin of the depth compression.

When shooting stereo videos, although modifying the interval of stereo cameras with narrower than actual eye separation affects the perception of the scene depth [34], such scene contraction cannot flexibly control a local depth (specific region) in a scene. Specifically, all objects in a scene are contracted toward a display surface with the underestimated amount of the eye separation. Generally, the original thickness of near objects should be maintained as much as possible because binocular disparities provide a dominant depth cue in the near region [35]. In this study, we compress scene depth by translating the vertices of objects with a user-defined function that allows arbitrary volume allocation in any depth as intended. This operation retains the advantages of light field displays, namely, that viewers can observe 3D shapes without the vergence-accommodation conflicts [3], [14], [15], [16], which are theoretically inevitable on stereoscopic displays like [36].

2.2 Evaluation Metrics and Conditions

While previous studies have evaluated depth-compressed 3D scenes with objective metrics [23], [24] or subjective metrics

[18], [36], they did not validate whether those scenes induced feelings of unnaturalness. For example, Adhikarla *et al.* [24] assessed image quality by SSIM and RMSE between depth-remapped and original 3D scenes. Lang *et al.* [18] asked participants to choose a presented 3D video featuring greater depth by comparing the images of depth-remapped and original 3D scenes. Oskam *et al.* [36] asked participants to choose scenes in which disparities were uncontrolled or controlled in terms of comfortable or realistic. In this study, we focused on subjective quality based on naturalness to evaluate the overall quality of 3D images, as depth perception involves complex cognitive processes that integrate various depth cues [35] (e.g., binocular and motion disparities and pictorial cues) and it is hard to quantify subjective naturalness using objective quality metrics. Chen *et al.* [37] proposed a "higher-level" subjective quality indicator, such as naturalness or sense of presence, for a proper evaluation of 3D image viewing experiments. Therefore, we also adopted "naturalness" as a subjective quality metric for the evaluation.

Motion parallax associated with viewpoint movements should be taken into account when evaluating depth-compressed 3D scenes on A3D displays, as the shape distortion appears to be significant and observers find it unnatural, particularly when the depth-compressed objects are seen from oblique viewing positions. However, the conventional depth remapping techniques for stereoscopic displays, e.g., [18], [19], [20], [21], were not designed and evaluated with any consideration of motion parallax with viewpoint movements. For example, Kellnhofer *et al.* [20], [21] used a chin rest when evaluating depth remapped 3D scenes. Other studies [22], [23], [24] addressing the blur of A3D displays did not conduct any experiments accompanied with motion parallax, although those displays inherently had the motion disparities.

3 Methods of Depth Compression

3.1 Static Approach

Depth compression is a method to contract a scene depth into a smaller depth range by manipulating scene geometry (vertex positions) so that the view at a specific position is preserved (Fig. 3a) [12]. In this paper, we refer to the view-preserved position as the origin of the depth compression. In particular, when the origin is fixed relative to a display position, we call this type of depth compression a 'static' approach. Note that depth compression assumes 3D CG scenes consisting of 3D models, point clouds, etc.; all the geometries in a scene are available.

Manipulating geometries can be formulated as follows. A vertex position $p = (x, y, z)$ is translated into $p' = (x', y', z')$ by

$$x' = xz'/z, \tag{1}$$

$$y' = yz'/z, \tag{2}$$

$$z' = f(z) = D\tanh\big((z - z_{front})/D\big) + z_{front}, \tag{3}$$

where D represents a depth range in which all the depth-compressed objects are placed. z_{front} is the nearest bound in a scene to which the depth compression is applied. $f(z)$ is a depth compression function. Typically, the origin is located at a default viewing position (e.g., a short distance in front of the screen center). The coordinate system is left-handed and the z-axis is perpendicular to the display screen, as shown in Fig. 3a. A depth position of a vertex is manipulated by $f(z)$. For $f(z)$, we chose a non-linear depth compression function in which the degree of the compression is asymptotically strengthened as the distance of the objects from the viewing position increases; far objects are strongly compressed while near ones are compressed mildly. Specifically, the gradient of $f(z)$ at z_{front} is 1, suggesting that the depth of the objects allocated there is intact. It has been shown that the non-linear function elicited more natural expressions than those elicited by a linear one through an experiment with various scenes [12]. Technically, the non-linear function is substitutable by another similar function, but we chose this function because it worked stably for various scenes in our preliminary tests.

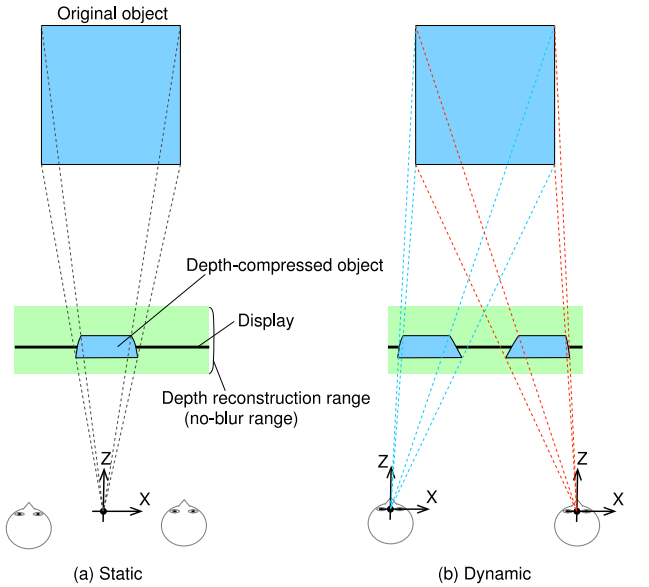

Fig. 3. Comparison of depth compression methods (top view). The original object (blue square) is contracted into a depth reconstruction range of the display by translating all vertices of the object. Even with the manipulation, the view of the square from the origin is completely preserved. (a) Statically depth-compressed object. The origin in the static depth compression is fixed relative to the display position regardless of the observer's viewing position. (b) Dynamically depth-compressed object. The origin in the dynamic depth compression is updated so that it corresponds to the viewing position.

To preserve perspective depth cues [35], we adjusted x and y values with the ratio of z'/z (in Equations (1) and (2)). This adjustment keeps the size of objects before and after applying the depth compression; more specifically, the retinal image size of each object is completely preserved when the viewing position is placed at the origin. If this adjustment is not applied, depth-compressed objects appear to be larger than the original uncompressed objects because their positions are closer to the observer's position.

Views of a depth-compressed object are strongly affected by the observer's viewing position in the static approach. Views of a depth-compressed object are completely preserved when viewing it from the origin because each vertex is transferred along with lines connecting the origin and each vertex (dashed lines in Fig. 3a). However, if the observer moves his/her viewing position away from the origin of the static depth compression, he/she would have more opportunity to see

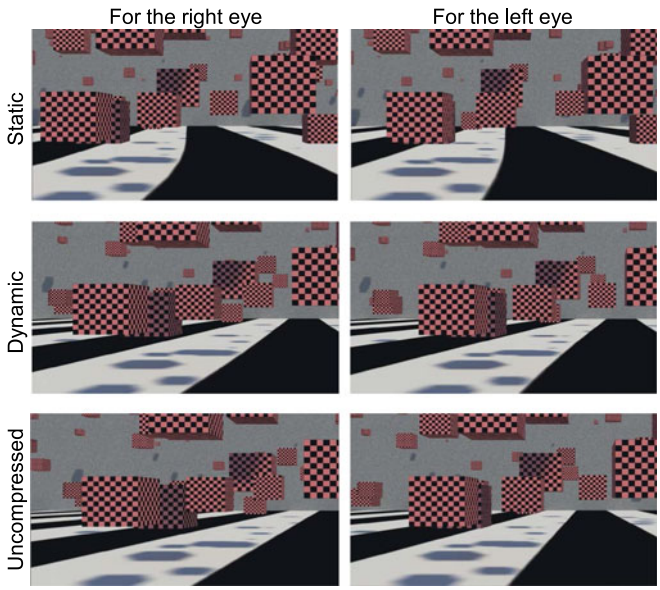

Fig. 4. Comparisons of the scene appearances with static or dynamic depth compression shown in stereoscopy. The left and right columns depict the views for the right and left eyes, respectively, so these images can be seen as stereo by cross-eyed viewing. These images are observable when a viewer shifted his/her head position 10 cm to the right from the depth compression origin. In the top and middle rows, scene depths are compressed into 40 cm. The bottom row shows the uncompressed (original) scene.

distorted views of the depth-compressed objects, e.g., a walkway and floor that are originally straight looks curvy (see Fig. 1c upper-right and Fig. 4 top row).

3.2 Dynamic Approach

We introduce a dynamic approach for depth compression in which the origin of the depth compression is updated depending on the viewing position in real time (see Fig. 3b), whereas the origin is fixed relative to the display screen in the static approach. By allocating the origin to the midpoint of the eyes, observable distortion by the depth compression can be suppressed to a certain extent. Given that the interpupil distance (IPD) is 6 cm (a typical value in adults), the difference between each eye and the origin is at most 3 cm. Thus, even with the dynamic approach, viewers still observe distortions, but the magnitude of them should be much smaller than those in the static approach (see Fig. 4 middle row and the supplementary movie for a better understanding). The viewing position can be accurately measured by a motion capture system or face-tracking techniques.

Updating the origin of depth compression in real time is computationally expensive. If the number of vertices is too large, it would cause intolerable delays in rendering. In this study, we minimize the processing time by manipulating vertices in parallel with a vertex shader of the GPU and achieve the processing in real time.

Although binocular disparities are attenuated to a certain extent, the dynamic depth compression is expected to compensate depth perception thanks to the presence of motion parallax. Not only the shape perception for the depth-compressed objects but also the arrangements of objects are well preserved in the dynamic approach even with the viewpoint movements. Thus, the real-time updates of the origin of the depth compression with the dynamic approach also contribute to maintaining the motion parallax that would be observed in the original scene.

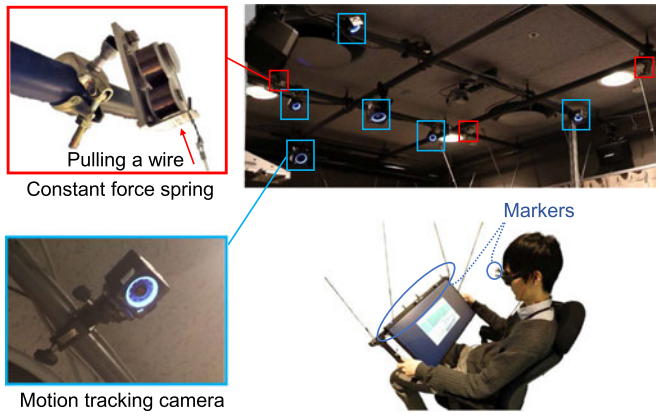

Fig. 5. Custom-made A3D display simulator that can easily be manipulated by both hands. Constant force springs pull wires toward the ceiling to ease the burden of holding the display. Motion-tracking cameras are placed on the ceiling and track the motions of glasses and displays for simulating motion parallax. Binocular disparities are simulated by the stereoscopic display and shutter glasses.

4 Estimation of Acceptable Depth Range

First, we conducted an experiment to assess how much we can compress scene depth without inducing feelings of unnaturalness in viewers by using an A3D display simulator. The static and dynamic depth compression methods were used for comparison. We utilized a simulated A3D display that has an ideal depth reconstruction capability with no blur at any depth, as currently available A3D displays still do not have sufficient capability to reconstruct substantial depth. At this time, with a real A3D display, it would be impossible to conduct evaluation experiments to elucidate the depth range required for showing depth-compressed scenes with appealing quality.

4.1 Participants

We recruited 40 participants (14 in their twenties and thirties and 12 in their forties) who had normal or corrected-to-normal vision and normal stereoacuity. Male-to-female ratios were one-to-one for each age group. No post-screening was conducted. All participants gave written informed consent, and the study was in compliance with the Declaration of Helsinki.

4.2 Apparatus

We implemented an A3D display simulator that produces binocular disparities and motion parallax by combining a stereoscopic display with a motion tracker, as shown in Fig. 5. The stereoscopic display can show 3D scenes without blur, which typically appears in a real A3D display, and this allows participants to focus only on the naturalness of the presented 3D images without the interference of blur. A regular 24-inch display (Acer KG 241 YU) was used as a simulated tablet display, as we were not able to find a suitable tablet display that featured both high resolution and stereo functionality. The display did not originally have stereo functionality, but we drove it at 120 Hz and produced stereoscopic images by presenting images for the left and

right eyes frame by frame that could be viewed through shutter glasses (NVIDIA 3D VISION 2). A 12-inch area at the display center (1280×800 pixels) was used for image presentation to simulate a typical tablet display. Motion parallax was produced by a motion-tracking system (OptiTrack, USA). Markers were attached to the display and glasses and measured at 120 Hz with sub-millimeter accuracies in their 3D positions. We obtained their 3D positions and directions for producing motion parallax and analyzing the relationship between the mean opinion score (MOS) of unnaturalness and the magnitude of viewpoint change. Tracking delay is officially reported as 4.2 milliseconds. The effect of the delay can be minimized because it is applied equally to both conditions (original and depth-compressed scenes).

We designed the simulator display so that participants were able to manipulate it easily by both hands as if they were using a tablet. Orienting the display direction makes quality ratings harsh enough because relative viewing positions to the display screen tend to be largely changed by moving or rotating the display. In our setup, because the weight of the display was too heavy to handle like a tablet, we hung the simulator display from the ceiling with four wires connected to four constant force springs. Therefore, participants could handle the display like a tablet and easily manipulate its direction with little effort.

We used custom-made software constructed by Unity (Unity Technologies, USA) for the pseudo-3D stimulus presentation. Left and right eye images were taken by two virtual cameras in a scene for providing stereovision. The interval of the stereo cameras was configured to match the actual IPD of each participant. The stereo cameras were moved according to the output of the motion tracker. We assumed that an ideal A3D display acted as a virtual window that connected between a virtual 3D scene and the real world. To achieve this, each camera was given oblique perspective projection matrices, with the near/far planes always parallel to the window frame, for which the four lines of each view frustum always passed through the corresponding four corners of the window frame, regardless of the viewing positions. We assumed the virtual window was placed in a fixed position of a 3D scene. Thus, when the display was rotated, the virtual scene was also rotated in the same direction as the window. This software also controlled open and close timings of the shutter glasses via a custom-made controlling device.

4.3 Stimuli

We presented six stimuli (scenes) that were defined using 3D CG models, as shown in Fig. 6, which were the same as those used in [12]. Shadows were baked into lightmap textures for avoiding appearance changes of shadows before and after compressing depth. Each stimulus was categorized into three groups—near, middle, or far scenes—based on the maximum depth of the scenes. These scenes were originally designed to be shown as the same scale as the objects in the real world when we displayed them on a 55-inch 3D display. However, assuming the use of the small display, we needed to convert all scenes into 12 / 55 sized miniature scenes so that the appearance on a 12-inch 3D display was as similar to that on a 55-inch display as possible

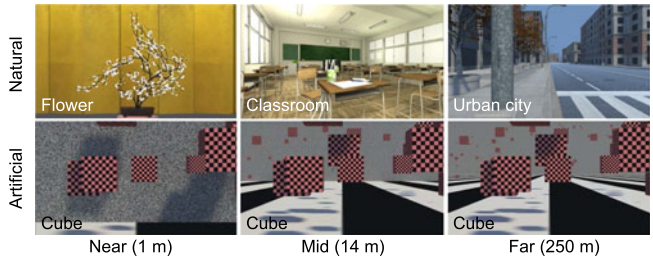

Fig. 6. Six stimuli for evaluating the naturalness of depth compression. These consist of two categories: natural and artificial (upper and lower rows). They are categorized into three depth groups (near, middle, and far; left to right columns) based on the scene depth that is the distance from the nearest object to the farthest object. These scenes produce the actual scale of a real scene (depths of 1, 14, and 250 m, respectively) when shown on a 55-inch 3D display that is assumed as a standard display. In the experiment, these scenes were downscaled according to the size of the displays.

because the views and picture compositions of scenes with the original sizes were apparently inappropriate on the small display and the display seemed to be too small to show scenes with the original sizes. In detail, near, middle, and far scene groups had depths of 0.218, 3.05, and 54.5 m, respectively. Two stimuli were used for each scene group, consisting of natural and artificial scenes. The natural scenes featured familiar objects from the real world that could be easily recognized by their shape or size, e.g., flowers, chairs, and streetlamps. The artificial scenes were generated to maintain constant density of cubes in the space through near, middle, and far scenes. These scenes had numerous checker-pattern textured cubes, a striped floor, and a wall. We placed the wall in the "cube scenes" at three levels of depth to generate the near, middle, and far scenes. The number of vertices in a scene ranged from 628 (the near cube scene) to 477156 (the flower scene). Tessellation was enabled to provide smooth shape deformation in the depth compression processes. All the scenes were rendered in 8.3 msec, achieving a refresh rate of 120 Hz.

To investigate the relationships between depth ranges and subjective quality for the scenes that were shown in the depth ranges, we compressed the depth of these scenes into six levels: 0.025, 0.05, 0.075 0.10, 0.125, and 0.15 m for near scenes, 0.1, 0.25, 0.5, 1.0, 1.5, and 2.0 m for middle scenes, and 0.1, 0.25, 0.5, 1.0, 1.5, and 5.0 m for far scenes. We chose these levels based on our preliminary experiments, assuming that the MOSs of unnaturalness were appropriately distributed from 1 to 5 so we could determine the required depth ranges that gave the acceptable level of depth compression (MOS 3.5). The total number of stimuli was $72 = 2$ (natural or artificial) $\times$ 3 (depth of the scene; near, middle, and far) $\times$ 6 (depth-compression levels) $\times$ 2 (depth-compression approaches (static or dynamic)). The order of stimulus presentation was randomized for each participant.

4.4 Procedure

The participants alternately compared the original and the depth-compressed scenes twice and scored the unnaturalness after the comparison on the basis of a modified version of the standard procedure used with the double stimulus impairment scale (DSIS) from ITU-R BT.500-13. Participants were asked to rate the levels of the unnaturalness. They were explained the concept of unnaturalness while comparing

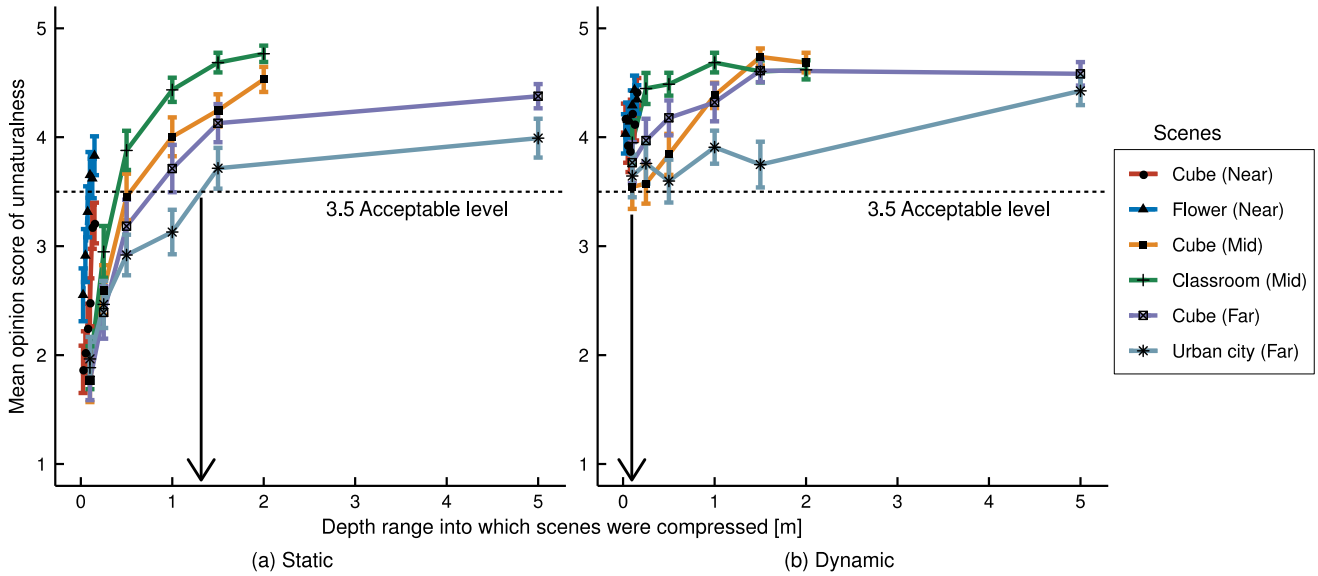

Fig. 7. Comparison of reported unnaturalness (mean opinion score (MOS)) for depth-compressed scenes. MOSs of the scenes that were depth-compressed with (a) static and (b) dynamic approach are shown. MOSs of unnaturalness tended to decrease (i.e., naturalness decreased) as the depth range decreased. The degradation of the dynamic depth compression (in (b)) was smaller than that of the static depth compression (in (a)). The arrows in (a) and (b) indicate the depth ranges that were required to show the depth-compressed scenes with acceptable quality level (MOS 3.5). Note that all observed data were converted into interval scale from category scale by using the method of successive categories. The MOSs are means across 40 participants, and error bars are standard error to mean (s.e.m.).

example images of the depth-compressed and original scenes; namely, it was not just the amount of shape or spatial distortion but the level of unnatural feelings. Thus, if they noticed shape distortion in the presented scenes but did not feel they were unnatural, they could rate the stimuli as natural. One trial consisted of four stimulus-viewing periods that were 5 seconds long and an evaluation period. The inter-stimulus interval was 2 seconds, during which time a character A or B was shown at the center on a solid gray background. The character indicated the label of the next scene: A for original and B for depth-compressed scene. After the comparison, participants rated the unnaturalness for stimulus B using a five-point impairment scale: 5 – imperceptible, 4 – perceptible but natural, 3 – slightly unnatural, 2 – unnatural, 1 – very unnatural.

Before the experiment, we conducted test trials using stimuli that were different from those in the main experiment to help participants learn the appearance of depth-compressed scenes and how to evaluate them.

4.5 Analysis

All the observed scores were sorted according to each stimulus and converted into the interval-scale from the category-scale by using the method of successive categories. As suggested with the DSIS method, we specified that a threshold of acceptable unnaturalness was a mean opinion score (MOS) of 3.5 across participants.

We analyzed the obtained viewing positions recorded at 120 Hz during the periods showing depth-compressed scenes. We obtained viewing angles by calculating the angles between a normal vector of the display screen and a viewing position vector represented relative to the display center. We then investigated the relationship between the reported unnaturalness and the maximum viewing angles in each trial by calculating Pearson's correlation coefficients. We adopted the maximum viewing angles because they are the most severe condition for depth compression. We also investigated the difference of the maximum viewing angles observed with the static and dynamic approaches using a Wilcoxon signed-rank test. We adopted the significance level of $p = 0.05$ that was corrected for the false discovery rate (FDR) in the multiple comparisons using the Benjamini-Hochberg procedure [38].

4.6 Results

Fig. 7 shows a comparison of the reported unnaturalness (MOS) for the depth-compressed scenes generated by the static and dynamic depth compression methods. With both methods, the unnaturalness became stronger (the MOSs of unnaturalness decreased) as the depth ranges became shallower. However, the static approach decreased the MOSs more significantly than the dynamic one, as suggested by the comparison between Figs. 7a and 7b. We found that a depth of at least 1.3 m was required for providing the acceptable level of MOS (3.5) in the static approach while a depth of 10 cm was sufficient in the dynamic approach. These results demonstrate that the dynamic approach was able to suppress the unnaturalness efficiently and provided a significantly high rate of depth compression, about 1 / 500, as the far scenes with a depth of 54.5 m were compressed to 10 cm.

Next, we investigated the relationship between MOS and the amount of viewpoint change. The percentages of viewing angles were 86% within 30 degrees, 95.6% within 40 degrees, and 97.7% within 45 degrees. We calculated Pearson's correlation coefficients between evaluation scores and the maximum horizontal angles in test stimuli (depth-compressed scenes) of each trial, but no significant correlations ($p > 0.05$; FDR corrected) were observed in any conditions. We also found no significant difference of the maximum horizontal angles observed with the static and dynamic approaches in the test stimuli of each trial (Wilcoxon signed-rank test, $p > 0.05$; FDR corrected).

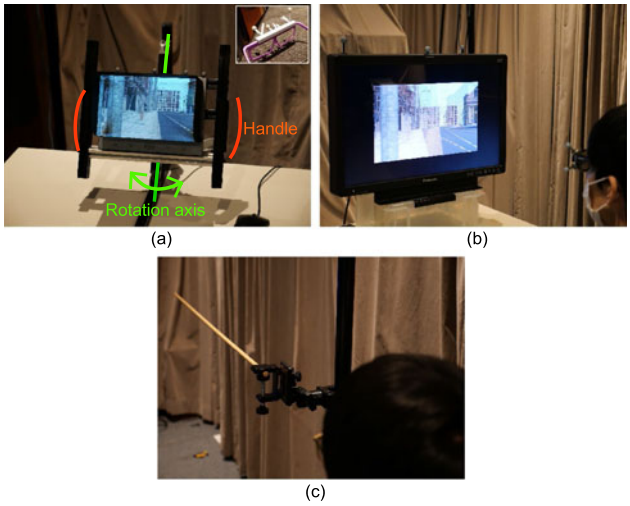

Fig. 8. Experimental setup for evaluating naturalness and depth perception of depth-compressed scenes on a real A3D display. (a) A real A3D display with a holding jig. Depth-compressed scenes were shown on an A3D display mounted on a holding jig that allowed participants to rotate the display around one axis by both hands. (b) A stereoscopic display used for presenting reference scenes. Reference scenes were shown on the stereoscopic display. Participants learned the original appearances of test scenes. The images were given with motion parallax provided using viewpoint tracking. (c) A test object for assessing participants' depth estimation ability. Participants observed and estimated the depth of real objects (wooden sticks) that were horizontally held in front of them.

5 Application on a Real A3D Display

We additionally tested whether the acceptable depth range estimated in the simulator environment was also valid on a real A3D display. Since one of the currently available A3D displays has enough depth reconstruction capability to cover the estimated acceptable depth range (10 cm), we implemented the static and dynamic depth compression into the real A3D display. We then investigated the subjective quality in terms of naturalness and perceived depth on it.

We performed two subjective evaluation experiments for the validation of our findings obtained with the simulator. First, we conducted paired comparisons between expressions generated with the static and dynamic depth compression methods, asking participants to choose a preferable expression in terms of naturalness from two candidates. We presented all the original uncompressed 3D scenes to participants before the comparisons so that they could obtain prior knowledge about the scenes and understand their original appearances. The reference scenes were shown on a stereoscopic display with motion parallax (Fig. 8b).

Next, we investigated the strength of the depth feelings the scenes with significantly contracted physical depth induced in participants. Participants engaged in a depth estimation task in which they estimated the depth of presented scenes whose physical depths were all 10 cm given by the static or dynamic depth compression. The depth estimation ability of all participants was evaluated in pre-tests in which they estimated the depths of real objects with various depths (lengths) made by wooden sticks.

5.1 Participants

We recruited 24 participants (eight each in their twenties, thirties, and forties) who had normal or corrected-to-normal vision and normal stereoacuity. None had participated in the previous experiment with the simulator. Male-to-female ratios were one-to-one for each age group. No post-screening was conducted. All participants gave written informed consent, and the study was in compliance with the Declaration of Helsinki.

5.2 Apparatus

The dynamic and static depth compressions were implemented in Looking Glass, a commercially available A3D display (Looking Glass Factory, USA). Its specifications were 8.9-inch, 2560×1600 pixels input resolution, 60-Hz refresh rate, and 50 degrees field of view. For saving computation time and reducing rendering latency, the number of views was reduced so that the field of view of the display covered only the area around the viewing position, which was the same approach as introduced by [32]. In detail, the number of views presented on the display was 45 in default, but in reality, the number of visible views from an observer was much smaller than the presented views. We omitted the rendering for the invisible views and rendered only the visible ones by considering the observer's viewing positions updated at 120 Hz according to inputs from a motion tracker. In the same way as in the simulator experiment, viewing positions and display directions were used for updating the coordinate origin of the dynamic depth compression. Markers were attached to plain glasses (without lenses) and the display edges, as shown in Fig. 8a. Note that because the A3D display did not originally require 3D glasses, the glasses were just used for measuring the viewing position of a viewer. We restricted display directions by mounting the display on a display stand such that it could be moved (rotated) only in the horizontal direction, as the display produces only disparities in a horizontal direction.

Fig. 8b shows the stereoscopic display used to give participants prior knowledge of the original scenes with binocular disparities and simulated motion parallax. The stereoscopic display (BT-3DL2550, Panasonic, Japan) presented stereoscopy by a circular polarized sheet showing left and right eye images with a line-by-line basis. Participants wore a pair of circular polarized glasses with markers for motion tracking. Motion parallax was simulated by the motion tracker, the same as in the A3D display. We showed stimuli at the central 17-inch (1280×800 pixels) area of the stereoscopic display. Participants viewed the stimulus images while seated on a chair at a viewing distance of approximately 76 cm, i.e., 3.3 times the height of the displayed area. We chose this viewing distance to provide a field of view equivalent in size to that on the A3D display (around 30 degrees).

Fig. 8c shows a real object (a wooden stick) used for assessing participants' depth estimation ability. A stick with a length ranging from 10 to 140 cm was held horizontally with a clip in front of an observer. The viewing distance was set to approximately 40 cm so that it was almost the same as that of the A3D display. Participants sat on a chair and were allowed to move their viewing position horizontally when observing the stick.

5.3 Stimuli

On the stereoscopic display, we presented the six scenes (shown in Fig. 6) without depth compression to give

participants prior knowledge about the scenes. To present scenes that were as similar to those in the simulator experiment as possible, we miniaturized the full-scale (original) scenes with a ratio of 17/55 isometrically on the x, y, and z axes. Thus, the depths of the scenes were also changed; the near, middle, and far scenes, which originally had depths of 1, 14, and 250 m, were downscaled to the depths of 0.3, 4.3, and 77 m, respectively.

On the A3D display, we presented the same six scenes. We miniaturized the full-scale scenes with a ratio of 8.9/55 in the same manner as above. Depths of the near, middle, and far scenes were downscaled to 0.16, 2.3, and 40 m, respectively. Then, all the downscaled scenes were compressed into a depth range of 10 cm with the static or dynamic approaches. Finally, twelve stimuli (6 scenes $\times$ 2 depth compression methods) were defined in the experiments in which the naturalness and the amount of perceived depth for the depth-compressed scenes were investigated.

For assessing the ability of depth estimation, we used wooden sticks with six different lengths: 10, 25, 70, 100, 125, and 140 cm. The diameter of the sticks was 1 cm. Sticks with lengths of 10 and 100 cm were presented as references; participants were informed of their actual length in advance. The rest of the sticks were used for the depth estimation task.

5.4 Procedure

The procedure consisted of four steps. The first and second steps were for examining the preference for the depth-compressed expressions with the static and dynamic approaches to validate whether the findings obtained in the simulator experiment were also valid on a real A3D display. The third and fourth steps were for determining the extent to which participants felt the extended depth with the dynamic depth compression.

As the first step, the six uncompressed scenes were presented on the stereoscopic display with randomized order. Each scene was presented for 10 seconds. Participants were instructed to look at the presented scenes and allowed to move their viewing position horizontally (while sitting on a chair) when the scenes were displayed. They were informed that those scenes were original and would be shown on the A3D display with some geometrical conversions.

Next, on the A3D display, participants alternately viewed depth-compressed scenes generated with the static and dynamic approaches and chose the preferable one in terms of naturalness. A trial consisted of four stimulus-viewing periods that were 5 seconds long. In these periods, the static or dynamic stimuli were alternately presented twice, followed by an evaluation period. To report preferences, participants moved a cursor for A or B shown on the display by using the wheel on a computer mouse. The inter-stimulus interval was 2 seconds. During the interval, a character A or B was shown on a solid gray background with no disparity indicating the label of the next stimulus. The order of stimulus conditions (static or dynamic) was randomized in each trial. We instructed participants to rotate (swing) the display slowly by using both hands.

Each participant underwent 12 trials to give their preference for the expressions. Participants observed the same stimulus in different trials twice because we presented stimuli with the static and dynamic methods in the forward and reverse orders to ensure a counterbalance of the presenting order.

After that, we tested participants' depth estimation ability by using wooden sticks. First, they observed 100-cm and 10-cm wooden sticks as references in this order and were informed of the actual depth (length). Then, they observed a series of wooden sticks with the lengths of 140, 25, 125, and 70 cm in this order. We interleaved the long and short sticks in the presentation order because consecutive presentation of sticks with similar lengths could allow a direct comparison with the previous stick and affect the depth estimation. Participants wrote down their perceived length of a presented stick on a form in each trial.

Finally, the participants estimated the depth of depth-compressed scenes generated with the static or dynamic approaches, shown on the A3D display. They observed stimuli over 10 seconds while rotating (swinging) the display slowly and reported the perceived depth of the scene. Before doing the depth estimation tasks, we informed them that the scene depth was the depth range between the nearest and farthest objects, e.g., the desk to the wall in the classroom scene in Fig. 6.

Notably, participants conducted all the tasks in this procedure while sitting on a chair. They were allowed to move their head position but not to leave the chair. Thus, the viewing areas they could choose relative to the displays or wooden stick stimuli were common throughout the procedure.

Before the experiments, we explained that the presented scenes on the A3D display were geometrically transformed ones of the original scenes. The participants conducted practice trials on both displays to learn the evaluation process.

5.5 Analysis

All the preference data given by participants was pooled and sorted in each scene. We tested the preference in terms of naturalness for the depth compression methods (static or dynamic) by using a binomial test and obtained p-values of the preference in each scene. These obtained p-values were corrected for the false discovery rate (FDR) in the multiple comparisons using the Benjamini-Hochberg procedure [38]. In addition, to examine the effects of scene categories (near, middle, and far) on the preference, we re-sorted the data set according to the scene categories and applied a chi-square test for them. The significance level was $p = 0.05$ in both analyses.

For testing participants' depth estimation ability, we plotted the estimated depth against the actual depth. When the relationship between them was proportionate (i.e., the length of a stick was recognized as it was with small errors), we assumed that participants had appropriate depth estimation ability.

For the depth estimation on the real A3D display, the estimated depths were pooled and sorted for each scene and depth compression method. To test whether there was a difference between the estimated depths given by the static and dynamic depth compression methods, we applied a Wilcoxon signed-rank test in each scene. Note that as we cannot assume that estimated depths follow the normal distribution, we adopted the Wilcoxon signed-rank test as a method of non-parametric statistical test. The same as in the

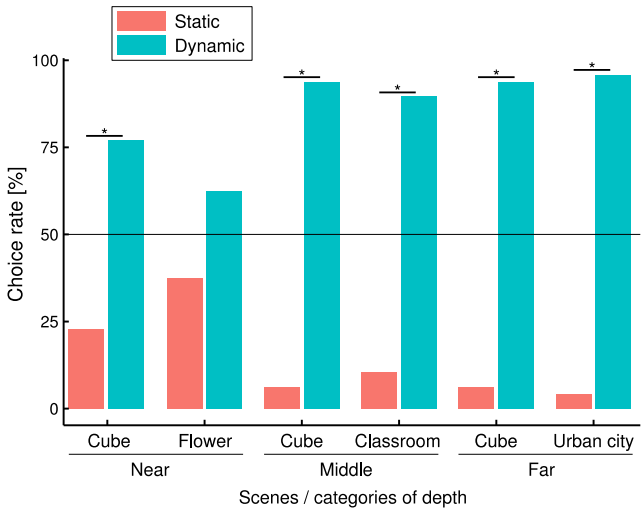

Fig. 9. Choice rate of preferable expression in terms of naturalness comparing static and dynamic approaches. Black horizontal line indicates the chance level (50%). Significance in binomial test is indicated with * ($p < 0.05$), where FDR was corrected for multiple comparisons.

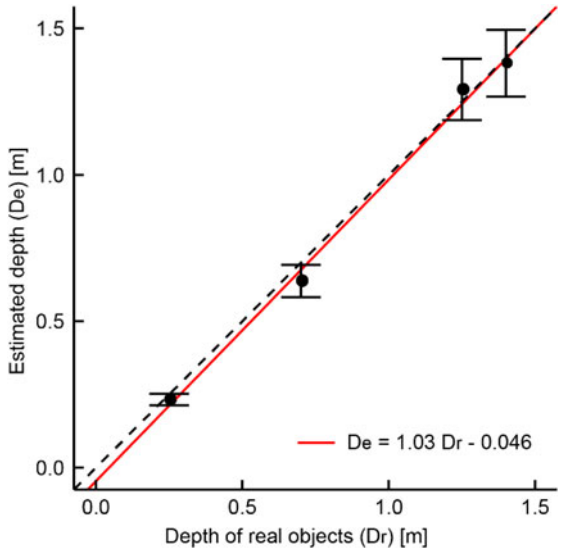

Fig. 10. Relationship between depth of real objects (wooden sticks) and estimated depth. Plotted values are mean of estimated depth. Error bars indicate 95% confidence intervals. Solid red line and dashed line represent the linear regression line and ground truth, respectively.

preference analysis, p-values were corrected for multiple comparisons. The significance level was $p = 0.05$.

5.6 Results

In the preference analysis, the dynamic approach was preferred over the static approach in terms of naturalness (see Fig. 9). We observed a significant preference for the dynamic approach in all scenes except the flower scene (binomial test, $p < 0.05$, FDR corrected). As predicted by the results in the simulator experiment, the dynamic depth compression provided the preferred expression in terms of naturalness, as evidenced in the simulator experiment (Fig. 7) showing that the MOSs of unnaturalness with the static approach dropped to a less than acceptable level of MOS (3.5) when the scene depth was strongly compressed to 10 cm. The viewing angles of 99.8% were within the 25 degrees that were sufficiently covered in the experiment using the simulator, where the range of 45 degrees contained 97.7% of the viewing positions.

As a result of testing the scene category effect using the chi-square test, we found significant effects on the "near" category, suggesting that the preference for the dynamic depth compression in near scenes was significantly weakened compared to those in the middle and far scenes (chi-square test, $p < 0.05$; FDR corrected). These results were also consistent with the results of the simulator experiment (Fig. 7) showing that MOSs at the depth of 10 cm for the near scenes were comparable in the static and dynamic depth compressions while those for the middle and far scenes were significantly different; none of the scenes with the static method performed above the acceptable level but ones with the dynamic method performed above the acceptable level.

The results of the participants' depth estimation ability are shown in Fig. 10. We found that the estimated depth (length) of real objects had a nearly proportionate relationship between actual and estimated values. The slope and intercept of the fitted line was 1.03 and –0.046, respectively. The variances of the estimated depth became smaller as the wooden sticks became shorter, where confidence intervals decreased from 0.23 to 0.039 with the stick lengths of 1.4 and 0.25 m, respectively. These results indicate that the participants had sufficient ability to precisely estimate the depth. Thus, we can expect rigorous evaluation with the participants on the display.

The results of the depth estimation on the A3D display showed that the depth of the depth-compressed scenes was perceived as deeper than the actual depth in all the scenes with both approaches (see Fig. 11). We found significant differences between the depth compression approaches in the middle cube, classroom, and far urban city, but not in the near scenes and far cube (Wilcoxon signed-rank test, $p < 0.05$; FDR corrected). These results suggest that the depth compression enhanced the perceived depth and that the dynamic approach had a stronger effect than the static one, particularly with scenes that had some extent of depth, as the middle and far scenes did.

6 Discussion

In this study, we have shown that, focusing on the condition of just a single observer, a shallow depth range was sufficient to assure enough perceptual quality in terms of naturalness and could induce the perception of enhanced depth to a greater extent than the actual depth without modifying the original hardware structure. In the simulator experiment, we found that a depth of just 10 cm was sufficient with the dynamic approach while 1.3 m was required with the static approach. In addition, we obtained consistent findings even on the real A3D display; the dynamic approach actually provided a preferable expression in terms of naturalness compared to the static one. We also found that the perceived depth was much deeper than the optically reconstructed depth (10 cm). In particular, the perceived depth in the middle and far scenes was strongly boosted by the dynamic depth compression approach. These results suggest that the dynamic depth compression is a practical way for presenting large depth scenes on A3D displays with a limited depth

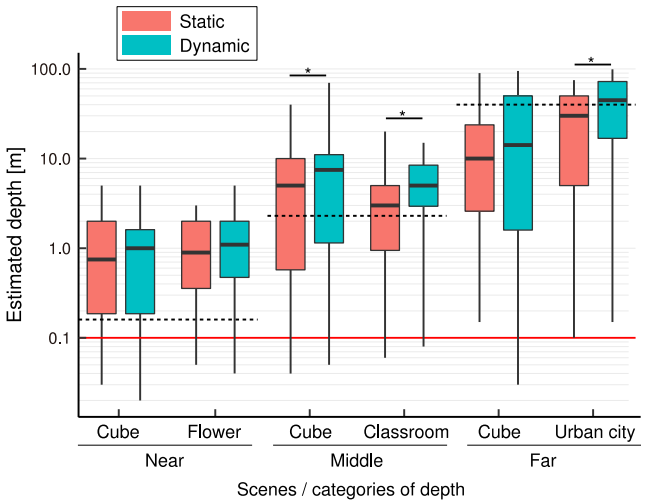

Fig. 11. Estimated depth of statically or dynamically depth-compressed 3D scenes that had a depth of 10 cm (red solid lines) on the A3D display. Black bold lines of each box indicate medians and dashed lines of each column indicate the depth of original uncompressed scenes. The scale of the vertical axis is log with a base of 10. Significance in Wilcoxon signed-rank test is indicated by * ($p < 0.05$), where FDR was corrected for multiple comparisons. Effect size of each scene was 0.73, 0.80, and 0.71, for mid cube, classroom, and far urban city, respectively.

reconstruction capability without dropping the subjective quality in terms of naturalness of the presented images.

6.1 Estimation of Acceptable Depth Range

While we used only six scenes (presented in Fig. 6) for estimating the depth range sufficient for showing scenes with the acceptable quality of unnaturalness, enrichment of a set of test stimuli (scenes) might affect the depth range to some degree. In the simulator experiment, we had expected the MOSs for the far cube scene would be strongly affected by the depth compression because it had plenty of visual features that could potentially lead to unnaturalness on the distorted shapes given by the depth compression, such as straight stripes on a floor or checker textures on the cubes. However, the MOSs for the scene "Urban city," a familiar scene with deep depth, was strongly affected when the depth was compressed into a depth of about 1 m, as shown in Figs. 7a and 7b. This might be because prior knowledge about the scenes affected the quality of unnaturalness. For example, participants knew the road was hard and rarely twisted due to actual experience in daily life, while they did not know the characteristics of objects in the artificial cube scenes. Thus, given the fact that prior knowledge affects the perceived quality, the use of other scenes could update the estimated depth range. However, at the acceptable quality level (MOS = 3.5), the required depth range for each scene varied in a small range with the dynamic depth compression. Therefore, even if new test scenes were introduced, our findings that the dynamic depth compression provides a much deeper perceived depth than the depth capability the display can originally have will be kept valid. We believe that the estimated depth range would not be varied so largely.

The required depth range obtained by the static approach was approximately 1.3 m, which was larger than that (1 m) obtained in the previous study [12]. This could be related to the difference of display sizes—12 and 55 inches in the present and previous studies, respectively—and the scale of the presented scenes. We presented the same scenes as those in the previous study, but they were downscaled so that the appearances relative to the screen size were consistent between the displays. The downscaled scenes gave viewers more opportunities to see the sides of objects. When looking at objects from a slanted viewing position, the distortion became more apparent for viewers. Therefore, the viewers felt a greater sense of unnaturalness in this study with a smaller display.

In the simulator experiment with the dynamic depth compression (Fig. 7b), while there was a mild impact on the MOSs regardless of the relatively high degree of the depth compression when the depth ranges were more than about 1 m, there was a rapid drop of the MOSs at the highest degree of the depth compression with the depth ranges of less than 1 m. The mild impact indicates a strong contribution of the dynamic depth compression with real-time updates of the depth compression origin because the tendency was weaker with the static depth compression (Fig. 7a). One source of the mild and rapid drops of MOSs could be the small gap between the depth compression origin and each eye. Thus, even with the real-time updates of the viewing positions, viewers still had a chance to find some distortion on the depth-compressed scene views, as mentioned in Section 3.2. The rapid drop could be explained by the relationship between the gap and the amount of the manipulated depth of objects. A comparison of the results between the static and dynamic depth compressions, shown in Fig. 7, suggested that the gap between the origin and viewing points affects the depth-compression performance, as the gap in the static approach was much larger than that in the dynamic one. Therefore, the rapid drop was caused because too strong a depth compression against the given gap performed under the threshold of the subjective quality.

In Fig. 7b, the MOSs for all scenes were better than or equivalent to the acceptable level (MOS = 3.5). We specified the required depth range of 10 cm because the MOS for the middle cube scenes first reached the acceptable level as the depth range decreased. We anticipate that the MOSs would be decreased to a lower than acceptable level if we added depth ranges shallower than 10 cm in the experimental condition. We did not put the shallower depth range in the experimental condition because a depth compression into a much shallower depth range generates an almost 2D image and loses the major characteristics of 3D displays. In fact, reconstructing a depth of 10 cm is already feasible with the current display technology, so there is no need to compress scene depths into such a shallow depth range. Moreover, we found that depth-compressed scenes within a depth of 10 cm sufficiently kept the depth perception that is one of the major characteristics of 3D displays via the experiments on a real A3D display (Fig. 11). Therefore, these results reliably suggest that our method can provide preferable 3D visual experiences on A3D displays with a depth reconstruction capability of 10 cm. Although these results might suggest that a head-tracked 2D display provides a good depth expression by motion parallax driven by head or display movements, they do not directly suggest the unnecessity of binocular disparities in 3D visualization because it has advantages over the simple 2D visualization in the performance of various tasks, such as spatial understanding,

memory, and recall [39], [40], [41] as well as presence [42]. We are further investigating how these advantages are affected by the depth compression.

Although our A3D display simulator used a stereoscopic display to provide a binocular disparity, it is unlikely that our results (Fig. 7) were significantly affected by the vergence-accommodation conflict (VAC), for the following three reasons. First, because the evaluation experiment was conducted in the same display condition potentially with VAC, the only difference between the images in the comparisons was given by the ways of geometry manipulation, not by VAC. Second, the obtained depth ranges were small enough to avoid exerting a sense of discomfort by VAC. Finally, the results obtained on the simulator were validated on the real 3D display. Therefore, our findings are general enough and will contribute to future 3D display developments.

Our findings on the required depth for A3D displays could potentially be updated by using other types of non-linear functions in the dynamic depth compression approach. In this study, we revealed that a depth of just 10 cm was sufficient for presenting originally much deeper scenes without inducing feelings of unnaturalness by using the asymptotic non-linear function, as shown in Equation (3). By combining other types of non-linear depth compression such as those based on saliency [18], [43], subjective quality for the 3D visualization may be improved even with the same amount of depth. Although improving the image quality would be valuable, there is no particularly strong need to reduce the required depth range because the current technology for light field displays has already achieved the depth of 10 cm. Because the image quality on light field displays has a trade-off relation between the sizes of the depth reconstruction and viewable zones, our findings provide evidence that the next light field display development should prioritize the enlargement of viewable zones.

6.2 Application on a Real A3D Display

When evaluating the subjective quality in terms of naturalness on the A3D display, we had to use a smaller display than the simulator (see Figs. 4 and 7a), which may have biased the selection of the preferable expression towards the dynamic approach in a specific situation. We used an 8.9-inch A3D display, which was slightly smaller than the 12-inch one used in the simulator, due to the absence of the exact size. On a smaller display with the static approach, depth-compressed scenes sometimes induced more unnatural feelings because we scaled down the size of the scenes according to the display size before applying the depth compression so that the composition shown on the display screen was kept even across different screen sizes (mentioned in the Section 5.2), and that might have brought more visibility to the distorted sides when viewers observed the scenes from oblique viewing positions. For example, if subjects shifted their viewing position horizontally in front of the same objects but with different scales, they might have had more chance to observe the sides of the ones with the smaller scale and feel unnaturalness on them. Thus, a smaller display could bring more unnaturalness depending on the amount of viewpoint change as the stronger scaling was applied on a smaller display. However, the range of the display manipulations for a smaller display was narrower than that for a slightly larger simulator display, as the range of horizontal angles of the A3D display was mechanically limited: approximately 25 degrees on the smaller A3D display and 45 degrees on the simulator display. Because the amount of viewpoint change on the A3D display was much smaller than on the simulator, it was unlikely that the subjects had more chance to find the unnatural distortions. Nonetheless, the subjects showed preferences for the dynamic approach. Therefore, we believe our results regarding the preference to the dynamic approach are valid even with the use of the different sizes of displays in the evaluation experiment.

We observed that a much deeper depth was perceived than the actual depth (10 cm) in the depth-contracted 3D scenes shown on the A3D display, as shown in Fig. 11. This suggests a significance of the pictorial cues in the mechanism of depth perception. Depth perception has been considered an outcome of the integration of various depth cues consisting of not only pictorial cues but also binocular disparities, motion parallax, accommodation, vergence, etc. [35]. In depth compression, the visual cues of binocular disparities, accommodation, and vergence signal the physical depth, but the pictorial cues, such as the relative size of objects, perspectives, occlusions, and shadings, signal a deeper depth than the physical depth. In the integration, because the weight for the pictorial cues was large [35], participants felt a much deeper depth, e.g., more than 10 m for the far scenes, in the physical depth of only 10 cm. In addition, preserved pictorial cues also preserve motion parallax. Thus, the perceived depth with the dynamic approach was larger than the static one. The effects of the depth compression hinge on the complex processing of human perception and cognition. Manipulating other depth cues may improve the quality of depth-compressed scenes.

6.3 Limitations and Future Work

In this study, we used only still 3D stimuli that did not include animation. If we used stimuli with animation, it might affect the size of the required depth range. Object movements along with the depth direction produce a variability in the object thickness, and that might cause an unnatural feeling for observers. To avoid this, it would be effective to dynamically change $f(z)$ in Equation (3). For example, observers might feel better about a depth-compressed scene in which the depth of a focused object, which could be determined with a saliency map or gaze tracking, is retained as much as possible by adjusting the depth compression function $f(z)$, as a similar approach was used in [20]. On the other hand, because the inclusion of animation in stimuli provides more depth cues for observers to estimate a scene depth, it might help improve the naturalness of the depth-compressed scene. Further investigation is needed to elucidate the effect of the inclusion of animation.

The dynamic approach has a limitation in that only a single observer benefits from it, as depth compression is designed to optimize for a single viewing position. This limitation is not problematic when a personal viewing style, as seen with a tablet or smartphone, is assumed. However, to support multiple observers, we would need additional hardware updates on displays, e.g., displays should have high enough directionality and provide 3D images selectively for observers depending on their location.

The dynamic approach requires significant computational resources, and this could be problematic when we need to show more complex scenes. Rendering should be done with a low latency on the updates of viewer positions. If the latency reaches a certain level, it could induce feelings of unnaturalness because the depth compression origin becomes distant from the proper position. Remote rendering with a rich computational power would not be a solution here because the interactive element is vital. As future work, we will investigate how to provide a high-quality rendering with reasonable computational requirements.

Even with the above limitations, the dynamic approach has a high degree of availability not only for A3D displays but also for other displays. For example, near-eye displays (NEDs) share a problem with A3D displays that requires hardware costs for showing deep depth scenes. In fact, NEDs induce inevitable artifacts such as VAC for showing such scenes [44], [45]. Because NEDs usually have a head tracking function, the dynamic approach that preserves pictorial depth cues compensating user's head motion can easily be introduced to NEDs without requiring any additional hardware. Such high availability of the dynamic approach will contribute to serving a variety of applications on other types of displays.

7 Conclusion

While A3D displays have a limitation when it comes to displaying arbitrary 3D scenes with substantial depth, our findings in this study suggest that dynamically depth-compressed scenes have sufficient perceptual quality even when compressing scenes that we see on a current television service without changing the hardware structure. We believe our findings will contribute to designing future 3D displays and providing attractive 3D content with full-parallax vision.

References

[1] G. Lippmann, "La photographic integrale," *C. R. Acad. Sci.*, vol. 146, pp. 446–451, 1908.

[2] Y. Takaki, "High-density directional display for natural three-dimensional images," in *Proc. Int. Meeting Inf. Display*, 2006, pp. 211–216.

[3] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, "The zone of comfort: Predicting visual discomfort with stereo displays," *J. Vis.*, vol. 11, no. 8, pp. 1–29, 2011.

[4] G. E. Favalora, "Volumetric 3D displays and application infrastructure," *Computer*, vol. 38, no. 8, pp. 37–44, 2005.

[5] Y. Matoba, T. Tokui, R. Sato, T. Sato, and H. Koike, "SplashDisplay: Volumetric projection using projectile beads," in *Proc. ACM SIGGRAPH 2012 Emerg. Technol. SIGGRAPH Conf.*, 2012, Art. no. 4503.

[6] P. C. Barnum, N. G. Srinarasimhan, and T. Kanade, "A multi-layered display with water drops," in *Proc. ACM SIGGRAPH 2010 Papers Conf.*, 2010, Art. no. 76.

[7] M. Gately, Y. Zhai, M. Yeary, E. Petrich, and L. Sawalha, "A three-dimensional swept volume display based on LED arrays," *IEEE/OSA J. Disp. Technol.*, vol. 7, no. 9, pp. 503–514, Sep. 2011.

[8] Y. Ochiai, K. Kumagai, T. Hoshi, J. Rekimoto, S. Hasegawa, and Y. Hayasaki, "Fairy lights in femtoseconds: Aerial and volumetric graphics rendered by focused femtosecond laser combined with computational holographic fields," *ACM Trans. Graph.*, vol. 35, no. 2, 2016, Art. no. 17.

[9] N. A. Dodgson, "Autostereoscopic 3D displays," *Computer*, vol. 38, no. 8, pp. 31–36, 2005.

[10] H. Hoshino, F. Okano, H. Isono, and I. Yuyama, "Analysis of resolution limitation of integral photography," *Opt. Soc. Amer.*, vol. 15, no. 8, pp. 2059–2065, 1998.

[11] H. Watanabe, N. Okaichi, H. Sasaki, and M. Kawakita, "Pixel-density and viewing-angle enhanced integral 3D display with parallel projection of multiple UHD elemental images," *Opt. Exp.*, vol. 28, no. 17, 2020, Art. no. 24731.

[12] Y. Sawahata and T. Morita, "Estimating depth range required for 3D displays to show depth-compressed scenes without inducing sense of unnaturalness," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 488–497, Jun. 2018.

[13] ITU-Recommendation BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," *ITU*, Geneva, Switzerland, 2012.

[14] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *J. Imag. Sci. Technol.*, vol. 53, no. 3, 2009, Art. no. 030201.

[15] K. Terzić and M. Hansard, "Methods for reducing visual discomfort in stereoscopic 3D: A review," *Signal Process. Image Commun.*, vol. 47, pp. 402–416, 2016.

[16] Y. Y. Yeh and L. D. Silverstein, "Limits of fusion and depth judgment in stereoscopic color displays," *Hum. Factors*, vol. 32, no. 1, pp. 45–60, 1990.

[17] A. Artusi *et al.*, "Siggarph ASIA 2011 course: Multidimensional image retargeting," in *Proc. SIGGRAPH Asia, 2011 Courses Conf.*, 2011, Art. no. 15.

[18] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, "Nonlinear disparity mapping for stereoscopic 3D," *ACM Trans. Graph.*, vol. 49, 2010, Art. no. 75.

[19] A. Kulshreshth and J. J. La Viola, "Dynamic stereoscopic 3D parameter adjustments for enhanced depth discrimination," in *Proc. Conf. Hum. Factors Comput. Syst.*, 2016, pp. 177–187.

[20] P. Kellnhofer, P. Didyk, K. Myszkowski, M. M. Hefeeda, H. P. Seidel, and W. Matusik, "GazeStereo3D: Seamless disparity manipulations," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–13, 2016.

[21] P. Kellnhofer, P. Didyk, T. Ritschel, B. Masia, K. Myszkowski, and H. P. Seidel, "Motion parallax in stereo 3D: Model and applications," *ACM Trans. Graph.*, vol. 35, no. 6, pp. 1–12, 2016.

[22] M. Zwicker, W. Matusik, F. Durand, H. Pfister, and C. Forlines, "Antialiasing for automultiscopic 3D displays," in *Proc. 17th Eurographics Conf. Rendering Techn.*, 2006, pp. 73–82.

[23] B. Masia, G. Wetzstein, C. Aliaga, R. Raskar, and D. Gutierrez, "Display adaptive 3D content remapping," *Comput. Graph.*, vol. 37, no. 8, pp. 983–996, 2013.

[24] V. K. Adhikarla, F. Marton, T. Balogh, and E. Gobbetti, "Real-time adaptive content retargeting for live multi-view capture and light field display," *Vis. Comput.*, vol. 31, no. 6–8, pp. 1023–1032, 2015.

[25] C. Ware, K. Arthur, and K. S. Booth, "Fish tank virtual reality," in *Proc. Conf. Hum. Factors Comput. Syst.*, 1993, pp. 37–42.

[26] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, "Surround-screen projection-based virtual reality: The design and implementation of the CAVE," in *Proc. 20th Annu. Conf. Comput. Graph. Interact. Tech. SIGGRAPH*, 1993, pp. 135–142.

[27] Q. Zhou *et al.*, "CoGlobe - a co-located multi-person FTVR experience," in *Proc. ACM SIGGRAPH 2018 Emerg. Technol. Conf.*, 2018, Art. no. 5.

[28] K. W. Arthur, K. S. Booth, and C. Ware, "Evaluating 3D task performance for fish tank virtual worlds," *ACM Trans. Inf. Syst.*, vol. 11, no. 3, pp. 239–265, 1993.

[29] C. Ware and G. Franck, "Evaluating stereo and motion cues for visualizing information nets in three dimensions," *ACM Trans. Graph.*, vol. 15, no. 2, pp. 121–140, 1996.

[30] I. Cho, W. Dou, Z. Wartell, W. Ribarsky, and X. Wang, "Evaluating depth perception of volumetric data in semi-immersive VR," in *Proc. Work. Adv. Vis. Interfaces*, 2012, pp. 266–269.

[31] A. Kulshreshth and J. J. Laviola, "Exploring 3D user interface technologies for improving the gaming experience," in *Proc. Conf. Hum. Factors Comput. Syst.*, 2015, pp. 125–134.

[32] X. Pan, M. Zheng, and A. Campbell, "Exploration of using face tracking to reduce GPU rendering on current and future auto-stereoscopic displays," in *Proc. ACM SIGGRAPH 2019 Posters*, 2019, pp. 1–2.

[33] X. Shen, M. Martinez-Corral, and B. Javidi, "An overview of head tracking integral imaging three-dimensional display using smart pseudoscopic-to-orthoscopic conversion," *Three-Dimensional Imag. Vis. Display*, vol. 10219, 2017, Art. no. 102190Y.

[34] Z. Wartell, L. F. Hodges, and W. Ribarsky, "Balancing fusion, image depth and distortion in stereoscopic head-tracked displays," in *Proc. 26th Annu. Conf. Comput. Graph. Interact. Tech. SIGGRAPH*, 1999, pp. 351–358.

[35] J. E. Cutting and P. M. Vishton, "Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth," *Perception Space Motion*, vol. 5, pp. 69–117, 1995.
[36] T. Oskam, A. Hornung, H. Bowles, K. Mitchell, and M. Gross, "OSCAM - Optimized stereoscopic camera control for interactive 3D," *ACM Trans. Graph.*, vol. 30, no. 6, pp. 1–8, 2011.
[37] W. Chen, J. Fournier, M. Barkowsky, and P. L. Callet, "Quality of experience model for 3DTV," in *Stereoscopic Displays and Applications XXIII*, Burlingame, CA, USA: SPIE, 2013, pp. 613–621.
[38] Y. Benjamini and Y. Hockberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *R. Statist. Soc.*, vol. 57, no. 1, pp. 289–300, 1995.
[39] M. H. P. H. van Beurden, W. A. Ijsselsteijn, and Y. A. W. de Kort, "Evaluating stereoscopic displays: Both efficiency measures and perceived workload sensitive to manipulations in binocular disparity," in *Stereoscopic Displays Appl. XXII*, Burlingame, CA, USA: SPIE, 2011.
[40] A. C. Neubauer, S. Bergner, and M. Schatz, "Two- vs. three-dimensional presentation of mental rotation tasks: Sex differences and effects of training on performance and brain activation," *Intell.*, vol. 38, no. 5, pp. 529–539, 2010.
[41] R. L. Sollenberger and P. Milgram, "Effects of stereoscopic and rotational displays in a three-dimensional path-tracing task," *Hum. Factors*, vol. 35, no. 3, pp. 483–499, 1993.
[42] W. Ijsselsteijn, H. De Ridder, J. Freeman, S. E. Avons, and D. Bouwhuis, "Effects of stereoscopic presentation, image motion, and screen size on subjective and objective corroborative measures of presence," *Presence*, vol. 10, no. 3, pp. 298–311, 2001.
[43] A. Chapiro *et al.*, "Optimizing stereo-to-multiview conversion for autostereoscopic displays," *Comput. Graph. Forum*, vol. 33, no. 2, pp. 63–72, 2014.
[44] G. A. Koulieris, K. Akşit, M. Stengel, R. K. Mantiuk, K. Mania, and C. Richardt, "Near-eye display and tracking technologies for virtual and augmented reality," *Comput. Graph. Forum*, vol. 38, no. 2, pp. 493–519, 2019.
[45] K. Akşit *et al.*, "Manufacturing application-driven foveated near-eye displays," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 5, pp. 1928–1939, May 2019.

Yamato Miyashita received the BS and MSc degrees in information and computer science from the Keio University in 2013 and 2015. He worked in News Technical Center News Production & Network Engineering Department, Japan Broadcasting Corporation in Tokyo, Japan from 2015 to 2017. Since 2017, he has worked in Science and Technology Research Laboratories, Japan Broadcasting Corporation. His research interests include cognitive science, computer graphics, and virtual reality.

Yasuhito Sawahata received the BS, MSc, and PhD degrees in information science and technology from the University of Tokyo in 2001, 2003, and 2015. Since 2003, he has worked in Science and Technology Research Laboratories, Japan Broadcasting Corporation in Tokyo, Japan. He was a research scientist with the National Institute of Information and Communications Technology and Computational Neuroscience Laboratories, Advanced Telecommunication Research Institute in Kyoto, Japan, from 2006 to 2010. His research interests include focusing on cognitive science, neuroscience, psychophysics, computer graphics, and virtual reality.

Kazuteru Komine received the BS and MS degrees from Tohoku University, Miyagi, Japan, in 1990 and 1992, and the PhD degree from the Tokyo Institute of Technology, Tokyo, Japan, in 2008. He joined NHK (Japan Broadcasting Corporation) in 1992, and successively worked for NHK Science and Technology Research Laboratories since 1994. His major research interests include cognitive science of human vision, human factors of information display and three-dimensional imaging systems.