3D3/VHF5-1
Invited

Perceptual Assessment of Depth-compressed Scene for Automultiscopic 3D Display

Yamato Miyashita[1], Yasuhito Sawahata[1], and Kazuteru Komine[1]

miyashita.y-fc@nhk.or.jp
[1]Japan Broadcasting Corporation, 1-10-11 Kinuta, Setagaya-ku, Tokyo 157-8510, Japan
Keywords: Depth compression, depth cues, perception and psychophysics

ABSTRACT

We present the required depth reconstruction range for future automultiscopic 3D displays. Using "depth compression" that contracts the whole scene depth into a limited depth range without inducing feelings of unnaturalness, we show experimental evidence that only a depth of 10 cm is sufficient for showing scenes with great depth.

1 Introduction

The key feature of automultiscopic 3D (A3D) displays is to present a scene as if it naturally existed in a physical space. These displays provide depth cues of binocular disparities and motion parallax without wearing special glasses. In this study, we aim to establish a design goal of future A3D displays: the depth range that should be reproduced for ensuring an adequate perceptual quality of substantially deep scenes.

A3D displays have a limitation on showing large depth scenes. For example, integral imaging displays [1] suffer from intolerable blur when showing objects far from their display plane. We need displays with denser pixels to expand a depth reconstruction range in which the 3D image is clearly shown [2]. Similarly, volumetric displays such as directly showing voxels in physical space have difficulty showing large depth scenes because their displayable space is physically limited.

In this study, we present a software-based solution called *depth compression* and the results of the perceptual assessments of depth-compressed scenes [3],[4]. In the depth compression, a large depth scene is contracted so that the whole scene is accommodated into the depth reconstruction range. The depth compression brings feelings of unnaturalness because the views of the objects and the between-objects space are significantly distorted in contrast to their originals. For designing future A3D displays, we empirically estimated the minimum requirement of depth that the depth compression needs to show a substantially deep scene without feelings of unnaturalness.

We estimated the required depth range while participants were allowed to move their heads and manipulate the display direction, assuming an actual usage scenario of A3D displays. Observing 3D scenes from various angles by moving viewing positions provides plenty of depth cues as motion parallax related to head movements have reportedly improved 3D perception in various tasks [5-8]. Therefore, head movements give viewers more opportunities to find geometrical distortion. Various depth manipulation techniques have already been proposed for the stereoscopic visualization [9-11], however, these techniques are insufficient for A3D displays because they did not account for movements of viewing positions.

2 Methods

2.1 Depth Compression

To show a large depth scene in the limited depth reconstruction range without any artifacts such as blur, we manipulate the scene geometry to fit into a given depth range. In detail, vertices of each object are transferred by

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

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

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

where (x, y, z) and (x', y', z') respectively represent the original and the transferred positions of vertices, assuming that the coordinate origin is placed at a viewing position. D and z_{front} represent the given depth reconstruction range and the nearest bound of the depth manipulated area, respectively. For better understanding, please visit our webpage: https://www.nhk.or.jp/strl/english/open2021/tenji/13/1.html.

The previous research [3] has reported that at least a depth of 1 m is still required with the conventional depth compression, which does not account for viewpoint movements. However, because the depth of 1 m is still difficult to reconstruct even by the state-of-the-art light field technologies, more efficient method is needed.

To improve the performance, we propose the dynamic approach [4] that introduced head tracking into the conventional depth compression (static approach) [3]. The origin of the depth compression is chosen optimal to avoid perceiving artifacts. In the static approach, the origin was fixed at the typical viewing position, which was at a certain distance away from the display center, e.g., 1.5 times of its height. In the dynamic approach, we successively update the origin to coincide with the center of both eyes.

With the dynamic approach, we investigated the required depth range for not to induce feelings of

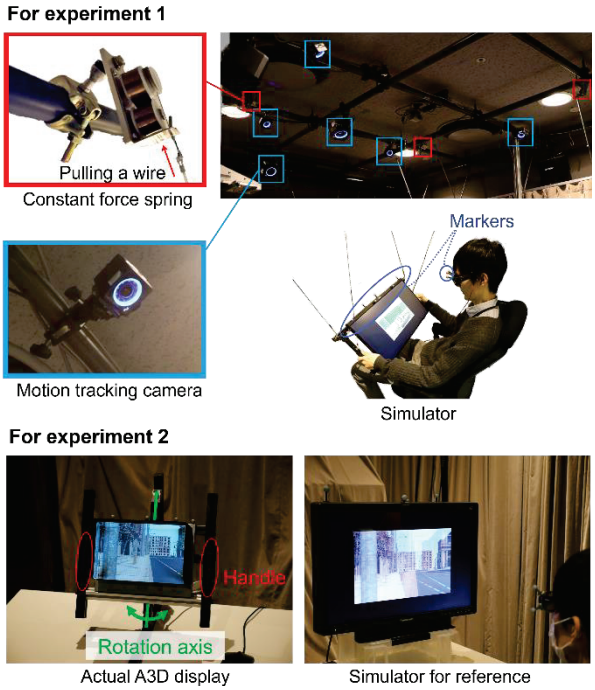

Fig. 1. Apparatuses used in the first and second experiments. (Modified from [4] / CC BY-4.0)

unnaturalness. The dynamic approach is expected to significantly reduce the senses of scene distortion. However, subtle gaps from the original image remain because the eyes are approximately 3 cm shifted from the origin. Therefore, there still exists a cause of feelings of unnatural distortion, and how much we can compress the scene depth is unknown. Here, we estimated and validated the depth range through the following two experiments; we first estimate the required depth on a simulator of A3D displays and then validate the estimated depth on the actual A3D display.

2.2 Experiment 1

As shown in Fig. 1 upper, we used a simulator of A3D display to assess the required depth reconstruction range without inducing the feelings of unnaturalness. Current A3D displays, such as integral imaging displays, are difficult to show scenes with deeper depths than a few dozen centimeters without quality degradation, suggesting that such actual displays cannot be used in the estimation of the required depth. Thus, we used a simulator of A3D displays that show scenes with any extent of depth without any artifacts such as blur. The screen size of the simulator was chosen assuming the typical size of tablet devices. The small size allows viewers to update relative viewing position easily and offers harsher evaluations of depth-compressed scenes.

We recruited 40 participants who had normal or corrected-to-normal vision and normal stereoacuity. Male-to-female ratios were equal. We conducted no post-screening.

We developed an A3D simulator that consists of stereoscopic display with motion tracker (see Fig. 1. upper). We used a 24-inch display (Acer KG 241 YU) that was drove at 120 Hz. It produced stereoscopic vision with shutter glasses (NVIDIA 3D VISION 2). Assuming a tablet-sized display, we used 12-inch area at the display center (1280×800 pixels). To ease the burden of holding the display, constant force springs pulled four wires attached to the display. The positions and posture of the display and the glasses were captured by motion tracker (OptiTrack, USA) for producing motion parallax.

We presented six scenes as stimuli that were generated by Computer Graphics (CG) software Unity (see Fig. 2.). The scenes consisted of two categories: natural and artificial. The natural scenes were ones likely to exist in the real world and were named flower, classroom, and urban city. The artificial scenes consisted of uniformly distributed cubes with a checker pattern. The six scenes were also categorized into three depth groups: near, middle, and far, each depth was 0.218, 3.05, and 54.5 m, respectively. The natural scenes were miniaturized for tablet display from the real size assuming the 55-inch display was the reference display, e.g., the natural scenes were 12/55 of the real size.

We compressed the scenes into six levels. The compression levels were 0.025, 0.05, 0.075, 0.1, 0.125 and 0.15 m for the near scenes, 0.1, 0.25, 0.5, 1.0, 1.5, and 2.0 m for the middle scenes, 0.1, 0.25, 0.5, 1.0, 1.5, and 5.0 m for the far scenes. Those levels were chosen based on our preliminary experiments, assuming the mean opinion scores (MOSs) were distributed largely.

The stimuli were presented based on the modified version of the standard evaluation procedure, double stimulus impairment scale (DSIS) from ITU-R BT.500-13 [12]. The participants alternately observed the original and the depth-compressed scenes twice and scored the naturalness. Each stimulus was presented for 5 seconds, and their inter-stimulus intervals were 2 seconds. The participants scored the naturalness of the depth-compressed scenes from the five-level choices based on the impaired scale: 5 – imperceptible, 4 – perceptible but natural, 3 – slightly unnatural, 2 – unnatural, 1 – very unnatural.

2.3 Experiment 2

As shown in Fig. 1. bottom-left, we used a real A3D display to investigate the validity of the required depth reconstruction range estimated on the first experiment. Since one of the currently available A3D displays can show the estimated depth of 10 cm (see detail in Section 3.1), we implemented the static and dynamic approach into the display. We hypothesized that the dynamic approach was the preferable expression in terms of naturalness over the static one according to the result of the first experiment.

We recruited 24 participants who had normal or

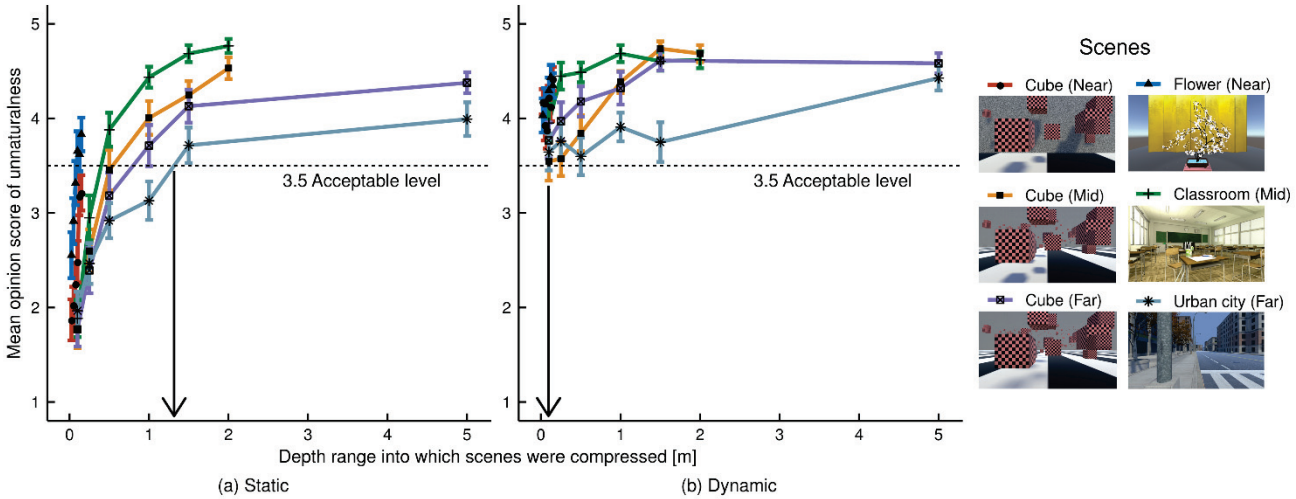

Fig. 2. Mean opinion scores (MOSs) of depth-compressed scenes with the (a) static or (b) dynamic methods. The arrows indicate the depth range that satisfies the acceptable level of unnaturalness with MOS 3.5. Error bars indicate standard error to mean (s.e.m.). (Modified from [4] / CC BY-4.0)

corrected-to-normal vision and normal stereoacuity. None had participated in the first experiment. Male-to-female ratios were equal. No post-screening was conducted.

The static and dynamic approaches were implemented in 8.9-inch (2560×1600 pixels) Looking Glass (Looking Glass Factory, USA), a commercially available A3D display. We captured the positions and posture of the display and plane glasses (without lenses or filters) just for implementing the dynamic approach on the display. The display was mounted on a stand with handles that allowed only horizontal rotation because the display produced only horizontal binocular disparities or motion parallax (see Fig. 1. bottom-left).

We also used a 17-inch (1280×800 pixels) stereoscopic display (BT-3DL2550, Panasonic, Japan) for presenting uncompressed scenes as the reference scenes (see Fig. 1. bottom-right). We presented motion parallax by using the motion tracker and circular polarized glasses with the same manner as in the first experiment.

We used the same six scenes as those used in the first experiment. Each scale was adjusted according to the size of displays, e.g., with the ratio of 8.9/55 and 17/55. On the real A3D display, depths of all scenes were compressed into 10 cm with the static or dynamic approaches.

Initially, participants observed all the uncompressed scenes on the simulator display. Each scene was presented for 10 seconds in a randomized order. The participants were instructed to observe scenes while moving their viewing position horizontally and sitting on a chair. We informed them in advance that those scenes were original; no geometrical modification was applied.

Then, participants observed depth-compressed scenes on the real A3D display. They directly compared views given by the static and dynamic approaches in the randomized order. The presentation time of each stimulus was 5 seconds and inter-stimulus intervals were 2 seconds. The static or dynamic stimuli were alternately presented twice. After the comparison, they chose a scene that they felt preferable expression in terms of naturalness. Participants were instructed to rotate the display horizontally with both hands during the presentation of stimuli. Participants observed the same stimulus in different trials twice in reversed order to ensure the counterbalance.

3 Results

3.1 Experiment 1

Fig. 2. shows the MOSs of depth-compressed scenes with the static and dynamic approaches. With both approaches, unnaturalness became stronger when the scene depth was more compressed. The MOSs of the static approach more sharply decreased compared to the dynamic one. The required depth range for the acceptable level, MOS of 3.5, was approximately 1.3 m for the static approach. With the dynamic approach, a depth of just 10 cm was sufficient and still slightly above the acceptable level.

3.2 Experiment 2

Fig. 3. represents the ratios of preference in terms of naturalness comparing the static and dynamic approaches. The participants significantly preferred the dynamic approach to the static one in all the scenes except for the flower scene (binomial test, $p < 0.05$, false discovery rate was corrected using Benjamini-Hochberg procedure [13]).

4 Discussion

In this study, we perceptually assessed the required depth range without inducing feelings of unnaturalness. Assuming a single observer condition with a handheld display, we utilized the head tracker to optimize the depth compression in real-time. In the first experiment

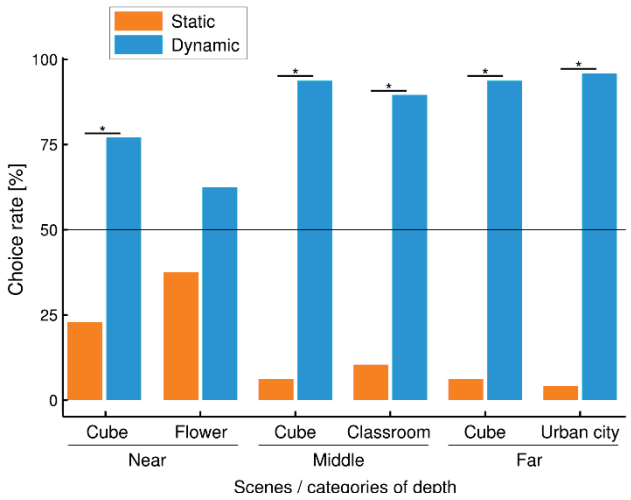

Fig. 3. Choice rates for each approach as the preferable expression in terms of naturalness. The black horizontal line indicates the chance level (50%) of the two-alternative forced choice task (2AFC). The significance of the binomial test was represented by * ($p < 0.05$). (Modified from [4] / CC BY-4.0)

with the A3D display simulator, we confirmed that at least a depth of 10 cm was sufficient for reproducing a substantially deep scene by the dynamic approach. The compression ratio was about 1/500 in the far scene with a depth of 54.5 m. In the second experiment, we confirmed the consistent result even on a real A3D display; the preference in terms of the naturalness of the dynamic approach was significantly higher than the static one.

In this study, we used only still 3D stimuli without any animation. In the animated scene, the depth positions of objects are dynamically changing that may cause an artifact of successively altered thickness. Our previous study found that the perceiving quality was further improved by dynamically allocating limited physical depth resources on selected objects [14].

The dynamic approach only works in the environment of a single observer because the depth compression is optimized for his/her viewing position. This limitation is not problematic particularly when a scene was observed by a single user. To support multiple observers, e.g., additional functions that can present light rays with a highly directional and user-selective manner will be required.

5 Conclusion

We empirically confirmed that the required depth range was just 10 cm by using the depth compression with the head tracking (the dynamic approach). We believe that our findings will contribute to developing future A3D displays and presenting various attractive 3D contents with full-parallax vision.

References

[1] G. Lippmann, “La photographic integrale,” C. R. Acad.Sci., no. 146, pp. 446–451, 1908.

[2] H. Hoshino, F. Okano, H. Isono, and I. Yuyama,“Analysis of resolution limitation of integral photography,” Opt. Soc. Am., vol. 15, no. 8, pp. 2059–2065, 1998.

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

[4] Y. Miyashita, Y. Sawahata and K. Komine, "Perceptual Assessment of Image and Depth Quality of Dynamically Depth-compressed Scene for Automultiscopic 3D Display," in IEEE Trans. Vis Comput. Graph., doi: 10.1109/TVCG.2022.3148419.

[5] 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.

[6] 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.

[7] I. Cho, W. Dou, Z. Wartell, W. Ribarsky, and X. Wang, “Evaluating depth perception of volumetric data in semiimmersive VR,” Proc. Work. Adv. Vis. Interfaces AVI, pp. 266–269, 2012.

[8] A. Kulshreshth and J. J. Laviola, “Exploring 3D user interface technologies for improving the gaming experience,” Conf. Hum. Factors Comput. Syst. - Proc., vol. 2015-April, pp. 125–134, 2015.

[9] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, “Nonlinear disparity mapping for stereoscopic 3D,” ACM SIGGRAPH 2010 Pap. - SIGGRAPH ’10, p. 1, 2010.

[10] A. Kulshreshth and J. J. La Viola, “Dynamic stereoscopic 3D parameter adjustments for enhanced depth discrimination,” Conf. Hum. Factors Comput. Syst. - Proc., pp. 177–187, 2016.

[11] 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.

[12] ITU-R BT.500-13, “Methodology for the Subjective Assessment of the Quality of Television Pictures,” 2012.

[13] Y. Benjamini and Y. Hockberg, “Controlling the false discovery rate: a practical and powerful approach to multiple testing,” R. Stat. Soc., vol. 57, no. 1, pp. 289–300, 1995.

[14] Y. Sawahata, Y. Miyashita, K. Komine, “Intended 3D Content Expressions on Light-field Displays using Adaptive Depth Compression”, ITE Trans. Media Technol. Appl., vol. 10, no. 2, pp. 75-88, 2022.