https://doi.org/10.2352/ISSN.2470-1173.2017.14.HVEI-378

Depth-Compressed Expression for Providing Natural, Visual Experiences with Integral 3D Displays

Yasuhito Sawahata, and Toshiya Morita; NHK (Japan Broadcasting Corporation), Science and Technology Research Laboratories; Setayaga, Tokyo, Japan

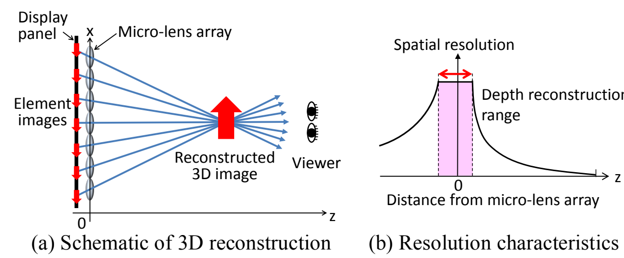
Abstract

Providing a natural 3D visualization is a major challenge in 3D display technologies. Although 3D displays with light-ray reconstruction have been demonstrated, displayable 3D scenes are selective because their depth-reconstruction range is restricted. Here, we attempt to expand the range virtually by introducing "depth-compressed expressions," in which the depth of 3D scenes are compressed or modified in the axial direction so that the appearances of depth-compressed scenes is kept natural for viewers. With a simulated system of an autostereoscopic 3D display with light-ray reconstruction, we investigated how large the depth range needed to be to show the depth-compressed scenes without inducing unnaturalness in viewers. Using a linear depth-compression method—the simplest way of depth-compression—we found that viewers did not feel unnaturalness for the depth-compressed scenes that were expressed within at most half the depth range of the originals. These results gave us a design goal in developing 3D displays for high quality 3D visualization.

Introduction

Recent progress in display technologies has achieved significant milestones in showing high quality 2D images, such as 4K/8K ultra high definition/resolution displays [1]. These advances also increase the potential for the development of glassless 3D TVs that can show objects or scenes in a realistic way because 2D and 3D displays share many basic technologies [2]. However, even with state-of-the-art, ultra high definition/resolution display technologies, realistic 3D depth information cannot be expressed properly on 2D arrays of pixels; high quality reconstruction of 3D scenes that have substantial depth is still difficult. To show various scenes with substantial depth in 3D TVs, we need to introduce an alternative approach that does not rely only on the progress of display technologies.

A type of 3D display that reconstructs light-rays in scenes usually has a limit on the depth reconstruction range (the range in the axial direction) of 3D scenes. In fact, scenes with substantial depth cannot be shown properly because the reconstructed images outside the range are blurred unintentionally. An integral 3D display, our target display, is one implementation of these types of 3D displays [3]. By implementing the principle of integral photography [4] on electronic and optical devices, integral 3D displays can produce 3D moving images. Specifically, an integral 3D display can be implemented by using a micro-lens-array placed in front of a 2D display surface with a gap width corresponding to the focal length of lenses, as shown in **Figure 1(a)**. Hence, viewers can have visual experiences characterized as being autostereoscopic and binocular as having motion parallax because pixels behind each micro lens are selectively seen depending on the eye position of each viewer. The spatial resolution of 3D reconstruction is determined based on the following parameters: lens pitch, focal length, and pixel pitch. According to the theoretical estimation of the spatial resolution of integral 3D displays [5], the spatial resolution of reconstructed images depends on the distance between the lens-array position and the reconstructed image position. The spatial resolution on the lens array plane was the highest, whereas it sharply decreased when the reconstructed image position exceeded a specific distance (**Figure 1(b)**). Therefore, the depth reconstruction range, in which a 3D reconstruction can be done with high spatial resolution, is limited. For example, even with a 13.3-inch 8K display (38.25-μm pixel pitch) and 293(H)x190(V), 1.74-mm focal length, and 1 mm-pitch micro lens array (using state-of-the-art equipment), by allowing 30 degrees of viewing angle, the depth reconstruction range is bounded to less than a dozen centimeters long.



(a) Schematic of 3D reconstruction (b) Resolution characteristics

Figure 1 Schematic of Integral 3D display and Reconstruction characteristics

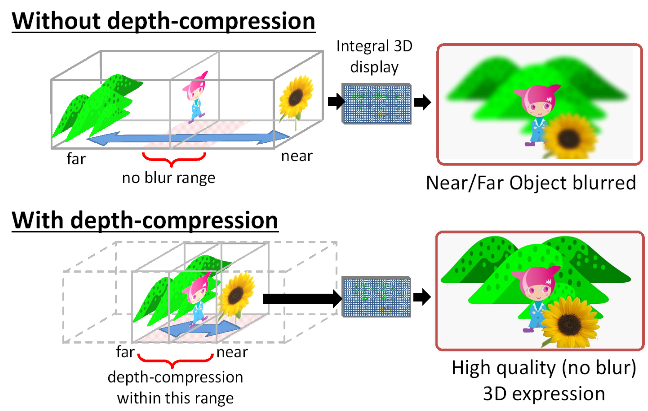


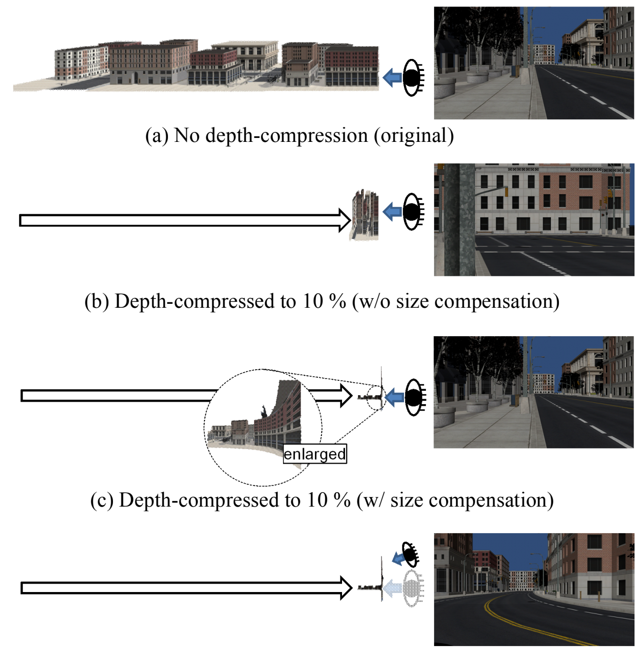
Figure 2 Depth-compressed expressions for providing natural 3D visual experiences

To solve these issues involved in the limited depth reconstruction range, a number of studies [5]–[8] have proposed disparity manipulation techniques in which the actual depth of a scene is reduced by manipulating a binocular disparity while the perceiving depth is not changed. Actually, the related techniques [9]–[11] have already been introduced in recent stereoscopic movie production [12] aiming at expressing 3D scenes in a comfort zone that prevents some kinds of visual stress, such as visual fatigue and virtual reality sickness [13]–[15]. These ideas could also be applied to integral 3D displays for providing all-in-focus 3D presentation. The required depth range in autostereoscopic displays with light-ray reconstruction is typically shallower than that of the comfort zone in stereoscopic displays; autostereoscopic displays only have about a dozen centimeters depth, whereas stereoscopic displays have several dozen centimeters depth around a screen surface at a typical viewing distance (e.g. one meter away from a screen) [13]. Therefore, the depth manipulation for such autostereoscopic displays is more challenging than that for stereoscopic ones. Furthermore, autostereoscopic displays can provide motion parallax, in which a viewer can also see the reconstructed images from different viewing positions. Even when the viewing position is moved from an ideal point, e.g., right in front of a screen, the depth perception for the presented 3D scenes should not collapse. Hence, we should take into account not only binocular disparity but also motion disparity in evaluating the performance of these manipulation techniques.

In this study, we investigated how large a depth reconstruction range was needed to enable viewers to avoid feelings of unnaturalness in depth-manipulated scenes. We introduced "depth-compressed expressions" as a manipulation method to show a 3D scene with substantial depth within a shallower depth range by modifying the shapes of 3D CG models (**Figure 2**). Depth-compressed 3D scenes were expected to be displayed with the best spatial resolution, while the original 3D scenes were expected to be unavoidably displayed with substantial unintentional blur. Although viewers would likely notice unnatural deformation of 3D scenes from depth-compressed or shape modified 3D scenes, we expected viewers to feel less difference or unnaturalness from depth-compressed scenes than the degree of deformation applied to the original scenes would suggest; that is, the relationship between unnaturalness and the degree of depth-compression should not be linear.

Our method, depth-compressed expressions, can be seen as a manipulation of depth cues. To perceive the 3D world, the human visual system integrates a variety of visual cues, such as binocular parallax, motion parallax, accommodation, convergence, object size, occlusion, shading, and textures [15]–[17]. The importance of each visual cue reportedly depends on the viewing distance [18]. This suggests that the result of depth perception might not be affected very much when unimportant depth cues are modified. In our method, by assuming a Cyclopean eye [16]—a virtual eye located at the center of two eyes—we adjusted the vertical and horizontal sizes of depth-compressed objects so that the retinal image size in the Cyclopean eye was kept the same as the original one. Therefore, under a special situation with monocular vision and no motion parallax, pictorial depth cues can remain unchanged. However, when the observation of depth-compressed objects or scenes is done under the more realistic situations with binocular vision and motion parallax, there could be some significant cues could detect some distorted expressions and induce unnaturalness regarding scene expression. In this study, we investigated the relationship between the levels of depth compression and the acceptability of naturalness of 3D scene expressions.

Figure 3 shows an example of the depth-compression method and some expected effects. **Figure 3(a) and (b)** show the resulting images that could be viewed when only the depth information of the 3D scene was compressed. As a result of the depth-only compression, various depth cues are modified, and the perceived scene expression should be largely different from the original one. Also, this modification of scenes could induce so-called cardboard effects [19]. After applying depth-compression, the distance to



(a) No depth-compression (original)
(b) Depth-compressed to 10 % (w/o size compensation)
(c) Depth-compressed to 10 % (w/ size compensation)
(d) Depth-compressed to 10 % (w/ size compensation) + Motion Parallax

Figure 3 Examples of depth compression: compressed scenes and the comparisons between their viewed images

each object was perceived as if they were actually there on the basis of size consistency[16] because the position of far and near objects moved closer to the 3D display screen . By compensating the vertical and horizontal sizes of objects for the movement of object positions, the retinal image size in the Cyclopean eye was kept the same as that in the original ones. Therefore, under a special situation with monocular vision and no motion parallax, the pictorial depth cues were able to remain unchanged, as shown in **Figure 3(c)**. However, when a viewer moved far away from the origin of the depth compression, he/she could notice the unnaturally distorted expressions of the depth-compressed scenes, as shown in **Figure 3(d)**. Such unnatural distortion could be expected to depend on the levels of depth-compressions.

To obtain a required depth reconstruction range for natural 3D viewing, we should take into account practical viewing conditions, including motion parallax and views from various viewing positions. Although previous works have reported that all-in-focus visualization in autostereoscopic displays could be provided by depth manipulation techniques [6]–[8], [20], [21], subjective qualities dependent on motion parallax and viewing positions were poorly considered in their performance evaluations. For example, Kellnhofer et al. [21] used a chin rest during evaluation experiments to view depth-manipulated images. This was due to the limited angular resolution of a currently existing autostereoscopic display. In our experiment, we used a simulation system of an autostereoscopic display in which binocular and motion disparities were simulated by combining a stereoscopic display and viewpoint tracking. Therefore, we could investigate a required depth reconstruction range for providing a natural 3D visualization without relying on the restricted specifications of current autostereoscopic displays.

In this paper, we present the concept of depth-compressed expressions and show that viewers did not feel unnaturalness regarding the depth-compressed scenes even when the scene depths were compressed to about 50% of the original ones. Also, we show that the method for depth-compressed expressions can be a key technology in enabling 3D TV systems to show a large variety of scenes that we can usually experience on contemporary broadcasts.

Methods

Participants

Thirty-six healthy adults (twelve subjects each in the 20s, 30s, and 40s age groups with male-female ratio of 1 in each age group) with normal or corrected-to-normal visual acuity and normal stereo acuity participated in this study. No post-screening of participants was conducted. All the participants gave written informed consent, and the study was approved by the Ethics Committee of NHK Science and Technology Research Laboratories.

Apparatus

An ideal 3D display that produced binocular and motion parallax features was simulated by combining a stereo display and viewpoint tracking. A 55-inch stereo display (55UF9500; LG, Korea) with 4K resolution was utilized to provide dichoptic viewing by using polarized glasses and by viewing line-by-line polarized images interlaced in each pixel-line for both eyes. Hence, the image resolution in the vertical direction was half of the 4K. Viewpoint tracking was provided using a three-dimensional position sensor (Fastrak; Polhemus, USA). The sensor was attached to the temple of the polarized glasses. The sampling frequency was 120 Hz. Position data were used not only to provide motion parallax but also to analyze the relationship between unnaturalness and the magnitude of motion parallax.

The standard viewing position was set 1.035 m away orthogonally from the center of the display screen. The participants observed stimuli sitting on a height-adjustable chair placed at the standard viewing position. Note that the distance corresponded to 1.5 times longer than the height of the display screen, providing an approximately 60-degree field of view (the standard viewing condition for 4K display). This viewing position was used only initially and the participants were allowed to move their viewing position during the experiments.

Custom made software built by Unity (Unity Technologies, USA) was used for the pseudo 3D stimulus presentation. Two images taken using two virtual cameras placed with an inter-axial interval in the 3D scenes were used to provide stereovision. The cameras were moved according to the output of the viewpoint tracking. To simulate the ideal 3D display, we needed to render images of 3D scenes as if the display frame were a window frame that connected the real and the virtual world. Therefore, each virtual camera was given oblique perspective projection matrices in which near/far planes were always parallel to the window frame in which the four edge lines of each view frustum always passed through the corresponding four corners of the window frame, regardless of the viewing positions.

Stimuli

Nine stimuli (scenes) defined using 3D CG models were presented (**Figure 4**). Each stimulus was categorized into three groups—near, middle, and far scenes—on the basis of the size of the scenes. Specifically, scenes within 1, 20 and 250 m were defined as near, middle and far scene groups, respectively. Three stimuli were used in each scene group, consisting of two natural scenes and one artificial scene. The artificial scenes were generated so that the object density in the space was constant among the near, middle, and far scenes. These scenes had a number of checker-pattern textured cubes and a striped floor in a space. The only difference between the scenes for the "cube scenes" in each scene group was the position of a wall that masked the far space behind it.

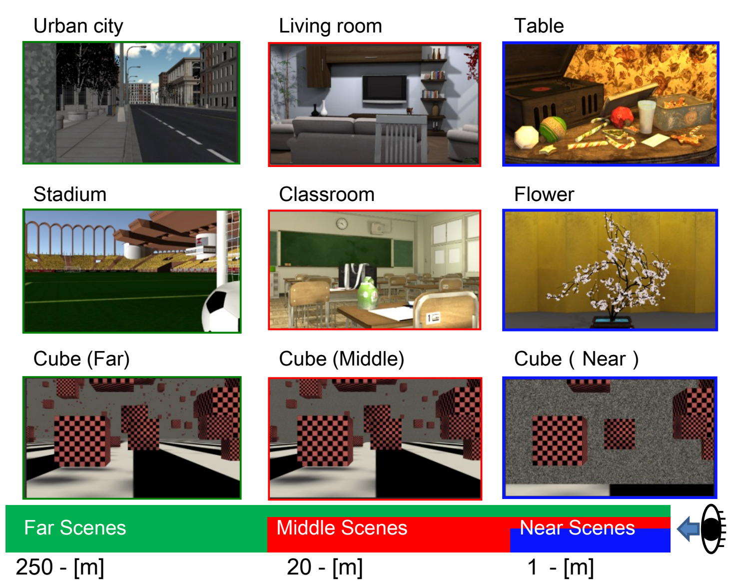
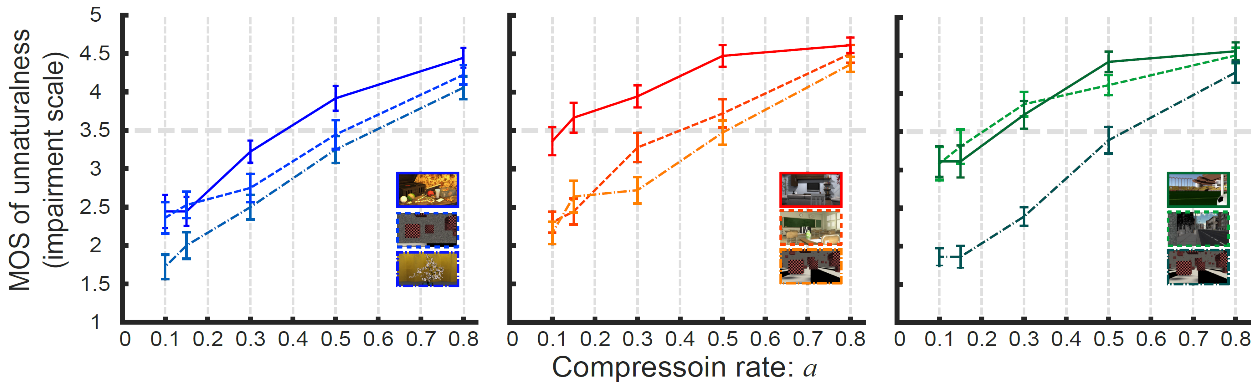


Figure 4 Scenes used as stimuli and scene groups



(a) Near (b) Middle (c) Far

Figure 5 Subjective evaluation for depth-compressed scenes

Depth-compressed scenes were created by modifying the three-dimensional vertex positions of each object in the original scenes. The procedure of depth compression consisted of two steps: 1) enabling transformation from world coordinates to the viewing position coordinates whose origin corresponded to the standard viewing position described in Apparatus, 2) modifying the coordinate values of each vertex. In the viewing position coordinates, the axial direction (orthogonal to the display screen) was defined as the depth direction and represented by the z-coordinate. Horizontal and vertical directions were represented by the x- and y-coordinates, respectively. We modified the vertex position in the viewing position coordinates by using the following equations,

$$\begin{aligned} z' &= a(z-d)+d \\ x' &= xz'/z \\ y' &= yz'/z \end{aligned} \quad , \qquad (1)$$

where x, y, and z were the original vertex position, x', y', and z' were the depth-compressed vertex position, a was the compression parameter (compression rate), and d was the position of the display screen. a was chosen at five levels: a = 0.8, 0.5, 0.3, 0.15, and 0.1. By using Eq.(1), 3D scenes were compressed toward the display surface on the basis of the chosen compression parameter a. Note that all the light effects, such as shadows and reflections, were baked into textures (light maps) before applying depth compression, so the light effects were not changed even when the shapes of objects or scenes were modified using the operations of depth-compression.

Experimental design

Participants alternately viewed the original and the depth-compressed scenes, and evaluated the unnaturalness of these scenes compared to the original ones, following the modified version of the procedure of the double stimulus impairment scale (DSIS) method [22]. They were asked about the levels of unnaturalness in our experiment, whereas evaluators were asked about the levels of image distortion caused by image processing (e.g., image/video coding) in the standard DSIS. One stimulus evaluation trial consisted of four stimulus-viewing epochs that were 5 sec long. In these epochs, original and depth-compressed scenes were alternately presented twice, followed by an evaluation epoch. In the evaluation epoch, viewers scored the presented stimulus using a five-grade impairment scale of depth compression; 5 - imperceptible, 4 - perceptible but natural, 3 - slightly unnatural, 2 - unnatural and 1- very unnatural. As

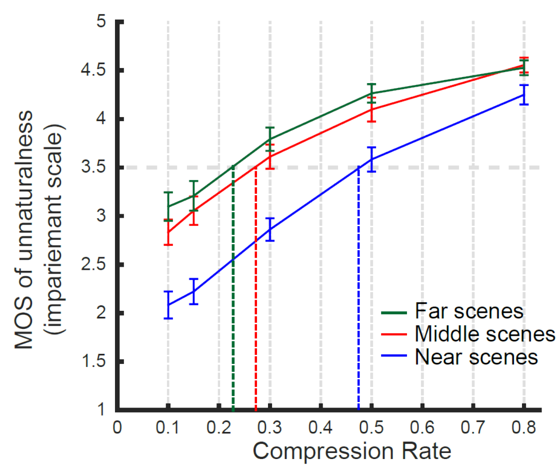


Figure 6 A Relationship between acceptable compression-rates and scene groups

suggested by the DSIS method, we defined the mean opinion score (MOS) across subjects = 3.5 as the threshold of acceptable unnaturalness. A total of 45 trials (9 stimuli x 5 depth-compression levels) were conducted in the evaluation experiment. The order of the stimulus conditions was randomized in each experiment.

Before the experiments, we carefully explained the concept of depth compression and the possibility that scenes or object shapes of presenting stimuli could be distorted. Also, participants conducted some test trials to learn how depth-compressed scenes were viewed and how they were evaluated.

Results

In all scenes, naturalness in the depth-compressed scenes decreased, coinciding with the level of the depth compression (**Figure 5**). Acceptable levels of depth compression (MOS > 3.5) depended on the scenes. We found that at least approximately 50% (a=0.5) of depth compression provided "acceptable unnaturalness" in all the scenes. Thus, these results suggest that depth-compressed scenes whose depth ranges were only half of the original could be viewed without inducing unnaturalness.

After redrawing the same data shown in **Figure 5** with respect to each scene group, higher levels of depth compression (smaller values of a in Eq.(1)) tended to be accepted in scenes of the far scene group than the middle or near scene groups (**Figure 6**).

Viewers moved their viewing position within ±0.71 m horizontally, ±0.45 m vertically and ±0.14 m axially. We calculated Pearson's correlation coefficient between viewers' evaluation scores and viewpoint ranges for each condition (45 conditions: 9 stimuli x 5 depth-compression levels), but no significant correlations ($P>0.05$, false discovery rate was corrected for multiple comparisons) were found in any of these conditions.

Discussion

We have found that 3D scenes could be expressed within a shallower depth range than the original one, without inducing unnaturalness in viewers. We found that only a half depth of the originals was needed to display them, even with the simple transformation of the shapes of scenes (Eq.(1)). These results suggest that our approach has the capability to produce 3D displays with a promising depth reconstruction range and to show a variety of scenes that originally require a deeper depth reconstruction range. Also, these results give us a design goal in developing 3D displays for a high quality, natural 3D visualization. For example, to show a 3D scene with 1-meter depth naturally for viewers, our results suggest that at least 50 centimeters depth reconstruction should be achieved in the future development of autostereoscopic 3D displays. Furthermore, we found that acceptable levels of depth compression depended on the size of scenes. These results suggest that more efficient depth compression may be possible by applying depth compression that is dependent on the position in original scenes.

Although larger scenes tended to accept higher levels of depth compression, the cube (far) and cube (middle) scenes, depicted by a dot-and-dash line in **Figure 5(b) and (c)**, showed comparable scores with the near scenes, shown in **Figure 5(a)**. These cube scenes were intentionally designed to have the density of depth cues per depth position. Hence, if viewers attended only to the near depth area, the depth cues they received would be almost the same. Although the stadium and the urban city scenes, which were the only other scenes than the cube (far) in the far scene group, had objects in the near area (a soccer ball or a pole of a street lamp), the density of depth cues could be less than the cues in the cube (far) scene. Therefore, optimal levels of depth compression could be chosen based not only on the size of the scenes but also on what the main subject in a scene was.

We found no significant correlations between viewers' evaluation scores and viewpoint ranges. These results suggest that viewers evaluated the stimuli focusing not only on motion parallax, but also on other depth cues. However, these results might be related to the limited range of viewpoint movements because it was not easy to move a larger amount while sitting on a chair. If viewers could move around more and move more easily, the results might be changed because large movements are known to lead to large spatial distortion (as shown in **Figure 3(d)**). Further investigation on control of viewing ranges is needed.

As mentioned in the introduction, the goal of this study was to display a variety of scenes, even with a substantial depth, in 3D displays whose depth reconstruction range is limited. We showed that scenes with substantial depth can be naturally expressed within half the range of the original. Actually, we believe that such depth compression could be helpful. In addition, depth-manipulation methods previously proposed for stereoscopic [9]–[11] and autostereoscopic displays [6]–[8], [20], [21] could be integrated. Further approaches are applicable to express a large variety of scenes within this limited range of depth. Therefore, the required depth reconstruction range can be dramatically reduced.

In future work, the discovery of the appropriate depth range to display arbitrary scenes naturally will facilitate designing the specifications of future 3D displays, such as pixel pitches, lens pitches, and focal lengths.

References

[1] F. Okano, W. L. Aylsworth, R. Schafer, and P. J. Hearty, "Beyond HDTV," *Proc. IEEE*, vol. 101, no. 1, pp. 5–7, Jan. 2013.

[2] J.-Y. Son, W.-H. Son, S.-K. Kim, K.-H. Lee, and B. Javidi, "Three-Dimensional Imaging for Creating Real-World-Like Environments," *Proc. IEEE*, vol. 101, no. 1, pp. 190–205, Jan. 2013.

[3] F. Okano, J. Arai, K. Mitani, and M. Okui, "Real-time integral imaging based on extremely high resolution video system," *Proc. IEEE*, vol. 94, no. 3, pp. 122–125, 2006.

[4] G. Lippmann, "La Photographic Integrale," *C. R. Acad. Sci.*, vol. 146, pp. 446–451, 1908.

[5] H. Hoshino, F. Okano, H. Isono, and I. Yuyama, "Analysis of resolution limitation of integral photography," *J. Opt. Soc. Am. A*, vol. 15, no. 8, p. 2059, Aug. 1998.

[6] 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, Dec. 2013.

[7] S. Xie, P. Wang, X. Sang, C. Li, W. Dou, and L. Xiao, "Depth-tunable three-dimensional display with interactive light field control," *Opt. Commun.*, vol. 371, pp. 166–172, 2016.

[8] 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.

[9] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, "Nonlinear disparity mapping for stereoscopic 3D," *ACM Trans. Graph.*, vol. 29, no. 4, p. 1, 2010.

[10] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H.-P. Seidel, "Apparent stereo: the Cornsweet illusion can enhance perceived depth," in *Human Vision and Electronic Imaging 2012*, 2012, vol. 8291, p. 82910N–82910N–12.

[11] 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, p. 1, 2011.

[12] B. Mendiburu, *3D movie making : stereoscopic digital cinema from script to screen*. Focal Press/Elsevier, 2009.

[13] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review," *J. Imaging Sci. Technol.*, vol. 53, no. 3, p. 30201, 2009.

[14] 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.

[15] I. P. Howard, *Perceiving in Depth*. Oxford University Press, 2012.

[16] J. M. Wolfe, K. R. Kluender, and D. M. Levi, *Sensation & Perception*. Sinauer Associates, 2011.

[17] B. Masia, G. Wetzstein, P. Didyk, and D. Gutierrez, “A survey on computational displays: Pushing the boundaries of optics, computation, and perception,” *Comput. Graph.*, vol. 37, no. 8, pp. 1012–1038, 2013.

[18] J. E. Cutting and P. M. Vishton, “Perceiving Layout and Knowing Distances,” in *Perception of Space and Motion*, vol. 22, no. 5, Elsevier, 1995, pp. 69–117.

[19] H. Yamanoue, M. Okui, and I. Yuyama, “A study on the relationship between shooting conditions and cardboard effect of stereoscopic images,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 3, pp. 411–416, 2000.

[20] A. Chapiro, S. Heinzle, T. O. Aydin, S. Poulakos, M. Zwicker, A. Smolic, and M. Gross, “Optimizing stereo-to-multiview conversion for autostereoscopic displays,” *Comput. Graph. Forum*, vol. 33, no. 2, pp. 63–72, 2014.

[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, Nov. 2016.

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

Author Biography

Yasuhito Sawahata received his BS, MSc and PhD in information science and technology from the University of Tokyo (2001), (2003) and (2015), respectively. Since 2003, he has worked in Science and Technology Research Laboratories, Japan Broadcasting Corporation in Tokyo, Japan. His work has focused on human information processing, neuro-imaging, neural-decoding and psychophysics.

Toshiya Morita received his BS in information science from University of Tsukuba (1984). He joined Japan Broadcasting Corporation at same year, and has been with Science and Technology Research Laboratories since 1988. His work has focused on human information processing, 3D display and evaluation method of images.