

Depth Range Control in Visually Equivalent Light Field 3D

Munekazu DATE^{†,a)}, Shinya SHIMIZU[†], Nonmembers, Hideaki KIMATA[†],
Dan MIKAMI[†], Senior Members, and Yoshinori KUSACHI[†], Nonmember

SUMMARY 3D video contents depend on the shooting condition, which is camera positioning. Depth range control in the post-processing stage is not easy, but essential as the video from arbitrary camera positions must be generated. If light field information can be obtained, video from any viewpoint can be generated exactly and post-processing is possible. However, a light field has a huge amount of data, and capturing a light field is not easy. To compress data quantity, we proposed the visually equivalent light field (VELF), which uses the characteristics of human vision. Though a number of cameras are needed, VELF can be captured by a camera array. Since camera interpolation is made using linear blending, calculation is so simple that we can construct a ray distribution field of VELF by optical interpolation in the VELF3D display. It produces high image quality due to its high pixel usage efficiency. In this paper, we summarize the relationship between the characteristics of human vision, VELF and VELF3D display. We then propose a method to control the depth range for the observed image on the VELF3D display and discuss the effectiveness and limitations of displaying the processed image on the VELF3D display. Our method can be applied to other 3D displays. Since the calculation is just weighted averaging, it is suitable for real-time applications.

key words: 3D display, light field, linear blending, depth range

1. Introduction

As existing 3D displays can offer high image quality only within a limited depth range, depth range control of the displayed image is necessary for practical use. Though the depth range can be controlled by generating depth maps and calculating reconstructed 3D images, the calculations are too heavy and complex for practical use. Therefore, we aim to develop a simple depth control method.

A light field expresses the distribution of rays in 3D space. In a 4D light field [1], a ray is expressed as a line through two points on two specific planes. Though this expression is exact within geometrical optics, the quantity of resulting data is huge and capturing a light field is not so easy. Therefore, we have been studying information compression and simple methods for light field capture.

Information processing in human vision is optimized to an extreme level to increase efficiency. Information quantity of the retinal image, which is received by photoreceptors, is reduced by cells in the retina before being transferred to the brain through visual nerves and visual perception is mostly

made using lower spatial frequency components. Therefore, there is redundancy in the retinal image. Visually equivalent light field (VELF) utilizes the redundancy of the visual system to reduce the data quantity. Since one implementation of VELF uses images of a regular camera array, data can be captured easily.

If light field information can be obtained, depth range of 3D images can be controlled. However, its applicability of VELF has been unclear. In this paper, we propose a depth range compression method for VELF and confirm its effectiveness using VELF3D displays.

2. Visually Equivalent Light Field 3D

2.1 Visual Acuity of Human Vision

The visual acuity of human vision is almost 1.0 arcmin to perceive the spacing of Landolt broken ring (Landolt C) chart. However, contrast sensitivity of human vision is maximum at the spatial frequency of about 5 cycle/degree (cpd) for bright stimuli [2]. Since the half period of this frequency is 6 arcmins, this frequency is much lower than that of the visual acuity for Landolt C chart.

At high spatial frequencies, sinusoidal and rectangular stripes exhibit almost the same frequency dependence of contrast sensitivity; at low spatial frequencies, the sensitivity for rectangular stripe patterns is higher due to harmonics [3]. Therefore, it is considered that the visual system cuts off high spatial frequency components.

On the other hand, visual acuity for Vernier sight [4] is much finer, almost 0.03–0.08 arcmins [5]. Though the resolution of binocular disparity strongly depends on the subject [6], discrimination for highly sensitive subjects is almost 0.1 arcmins [6], [7]. These values are much higher than that expected from the contrast sensitivity characteristic. Since Vernier sight detects the edge position difference of two lines and stereopsis detects edge position difference for left and right eyes, we believe that human vision is extremely sensitive to edge position. In other words, the sensitivity to the phase of each spatial frequency component is 10 times or higher than the period of the stripe.

This characteristic of human vision has already been put into practical use in the autostereoscopic 3D display, the depth fused 3D (DFD) display. In this display, two or more display screen layers are stacked with fixed separation [8]. Though, retinal images of this display are double due to the

Manuscript received March 13, 2020.

Manuscript revised June 1, 2020.

Manuscript publicized August 13, 2020.

[†]The authors are with NTT Media Intelligence Laboratories, Nippon Telegraph and Telephone Corporation, Yokosuka-shi, 239–0847 Japan.

a) E-mail: munekazu.date.zf@hco.ntt.co.jp

DOI: 10.1587/transele.2020DII0003

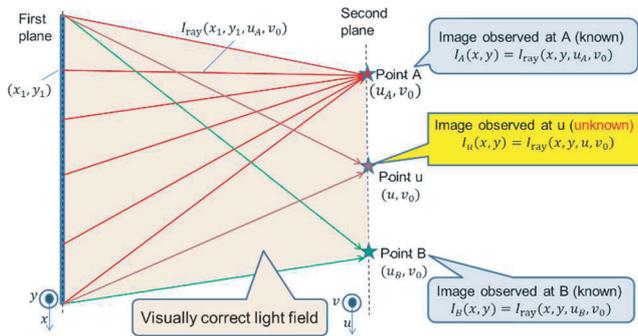


Fig. 1 Concept for visually equivalent lightfield.

parallax of the layers, when the offset of the double images is small enough, less than almost 3-5 arcmins, the double images cannot be perceived. Depth perception of the DFD display can be calculated from the stereo matching of retinal images after applying a Gaussian low pass filter [9], [10]. The cutoff of spatial frequency is 17 cpd. Since cutoff of contrast sensitivity is around 10-20 cpd [2], [3], it agrees with our theory that perception follows the application of a low pass spatial filter and that human vision has high sensitivity to phase for detecting edge position with high precision.

2.2 Expression of Light Field

In a 4D light field, a ray is expressed as a line through two points on two specific planes. Here, we assume that rays propagate from the first to the second plane. Rays to a point on the second plane compose the image observed at this point. Therefore, if observed images at all points on the second plane can be obtained, complete data of the 4D light field can be obtained. Unfortunately, the data quantity is huge.

VELF uses the visual acuity of human vision in compressing the data. If images observed at all points on the second plane can be generated from those at sparse points on the second plane, data quantity can be reduced.

Figure 1 shows the concept of the visually equivalent light field [11]. We generate image observed at point u from known images observed at points A and B on the second plane. As shown, first plane coordinates are (x, y) . Those of the second plane are (u, v) . Luminance of a ray can be expressed as $I_{ray}(x, y, u, v)$, which is a function of the two intersection points of the ray and the two planes. In other words, observed image $I(x, y)$ at viewpoint (u, v) can be expressed as $I(x, y) = I_{ray}(x, y, u, v)$. To make discussion simple, we assume that height of points, v , is v_0 . We generated the image observed at u by linear blending of those observed at points A and B .

Figure 2 shows the linear blending calculation. When the disparity between two neighboring camera images is small enough, i.e. 3-5 arcmins or less, the observer perceives a linearly blended image as a natural intermediate viewpoint image even though optically it is a double image. Image

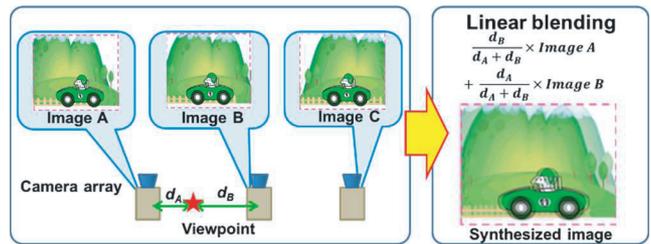


Fig. 2 Intermediate viewpoint image synthesis using linear blending.

weights for blending are in proportion to the offset of the camera position and the viewpoint. The light field is given by

$$I_{ray}(x, y, u, v_0) = (1 - \beta)I_A(x, y) + \beta I_B(x, y) \quad (1)$$

where $\beta = \frac{u_A - u}{u_A - u_B}$.

This ray is visually correct for human vision. By sweeping point u on the second plane between the two points (A and B), any ray from the first plane can be calculated and a visually correct light field can be obtained in the area highlighted in Fig. 1. That is, the visually equivalent light field can be produced by linear blending of discrete viewpoint images, if the disparity of two nearest viewpoint images is small enough for interpolation. Though this equation is for the light field that yields horizontal parallax, complete full-parallax 4D light field can be calculated by using four viewpoint images set at the vertexes of a rectangle and two-dimensional linear blending by bilinear interpolation.

We created a demonstration system for life-size video communication that used this calculation. Using a stereoscopic 3D projection display and shutter type glasses with head tracking, the system offers smooth motion parallax and a high feeling of existence [12], [13].

Interpolation via linear blending can be explained mathematically using spatial frequency analysis. When the cutoff spatial frequency can be assumed, linear blending acts as effective interpolation if the offset of the two images is less than a quarter period of the cutoff spatial frequency. Because the weighted average of two sinusoidal stripes generates an intermediate phase stripe, phase varies almost linearly depending on the weight in this condition. When the phase difference of the two waves is $\pi/2$, error in phase linearity is less than 5%. Moreover, it decreases as the phase difference decreases [14]. When the disparity is 3-5 arcmins, 5% yields errors of 0.15-0.25 arcmins, which matches the stereopsis discrimination performance of almost all subjects [6].

2.3 VELF3D Display

VELF3D not only reduces data quantity but also contributes to the efficient use of the pixels on the lightfield 3D display. Since the calculation of linear blending is so simple, only weighted averaging is used, we can realize it optically in the display.

Examples we have developed include autostereoscopic

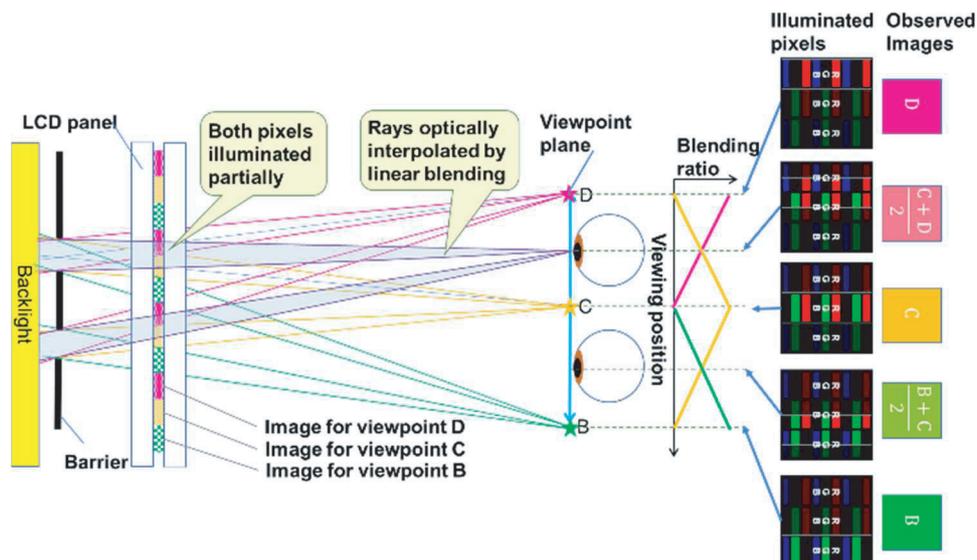


Fig. 3 Structure and mechanism of the 3-viewpoint type VELF3D display.

optical linear blending 3D displays using projection optics [15], [16] and a ball lens [17]. However, the distortion imposed by projection optics seriously degrades images and reduces the effect of linear blending. Therefore, we have proposed the flat panel 3D display, which has high resolution and no distortion. Its smooth motion parallax and high 3D image quality have been confirmed [14], [18]–[20]; we call it the visually equivalent light field 3D (VELF3D) display.

Figure 3 shows the structure of the 3-viewpoint type VELF3D display. It consists of an LCD panel and a vertical stripe parallax barrier in front of a backlight module. Though the components of the display are the same as those of parallax barrier type autostereoscopic 3D displays, we blend neighboring viewpoint images aggressively by making the aperture width of the barrier almost equal to pixel pitch, and using a horizontal RGB stripe LCD panel. As the pupil scans on the horizontal direction, illuminated area on the LCD panel by rays from the barrier slits moves and the blending ratio of neighboring pixels varies. Since the variation in blending ratio is linear, linear blending can generate intermediate viewpoint images optically. In other words, interpolated rays for viewpoints can be generated by blending rays. A full parallax display is also possible by using a parallax barrier with special aperture structure [18], [19]. Its image quality is high, because interpolation by linear blending enables the use of many pixels in each direction. Precise edge position, finer than pixel pitch, can be reproduced due to our highly precise phase reproduction. This edge reproducibility is the same as anti-aliasing [20]. Live images have been displayed using an array of 5 cameras corresponding to number of display viewpoints.

Figure 4 shows the observed images and the viewing zone of the 5-viewpoint type VELF 3D display when the width of display viewpoints A-E equals that of the display screen. When the position of the pupil is at “a”, that is pupil

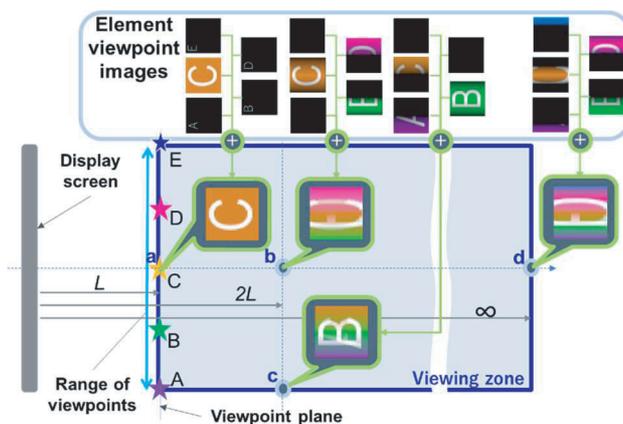


Fig. 4 Viewing zone of 5-viewpoint type VELF 3D display.

position matches viewpoint C on the viewpoint plane, observed image is element viewpoint image “C”. When the pupil is on the viewpoint plane, i.e. observation distance is L , observed image is the linear blending of the nearest two viewpoint images. From this distance, the disparity between neighboring images, δ , is designed to be less than δ_{\max} , which is usually 3-5 arcmins, to avoid double image perception.

Pupil position “b” is further than the viewpoint plane. If we set the first plane of Fig. 1 onto the plane of display screen and the second plane onto the viewpoint plane, any ray from screen to pupil passes through regions between viewpoints A-E, and so can be calculated as a linear blending of those of the nearest two viewpoints according to Eq. (1). This calculation of linear blending is equivalent to the sum of element viewpoint images, which are viewpoint images with spatially varying weights. Since observation distance is $2L$, observed image is a blend of viewpoint images B, C and D. Since the disparity of neighboring two

images is $\delta/2$ at this distance, it is less than δ_{\max} and no double image is observed. When pupil position is “c”, observed image is a blend of viewpoint images A, B and C.

When the distance of pupil position is infinity, as indicated by “d”, all five element viewpoint images are blended linearly. The width of the viewing zone limited by vignetting is the same. Since observation distance is quite far, observed disparity between element viewpoint images is almost zero.

Since the periods of the LCD panel and the barrier are similar, moiré is produced. However, in the VELF3D display, luminance distribution of element viewpoint image corresponds to a single fringe of moiré, so moiré is essential and doesn't degrade image quality.

We state that images observed when the pupil is in the viewing zone do not suffer from vignetting or double images. When pupil position moves outward past the extension of the display screen width, some rays pass through the outside of the range of viewpoints A-E and they cannot be generated by linear blending; as a result vignetting occurs. Therefore, point “b” is the left end of the viewing zone. Since this discussion is independent of the observation distance from the screen, the width is also independent of observation distance. Since $\delta \leq \delta_{\max}$ at viewpoint plane, no double images are observed. That is, the viewing zone is a rectangle from the viewpoint plane to infinity and its width equals screen width.

3. Proposed Method and Results

In the VELF3D display, only interleaved viewpoint images are sent as the images to the LCD screen. As shown in Fig. 5, in the normal shooting condition of live images, the relationships between viewpoints of the display and those of the screen and to the camera positions and objects are the same or geometrically similar. When the layouts are the same, displayed image size matches object size.

For reducing display depth range, we considered two methods. The first method (Method A), calculates small disparity viewpoint images to display from camera images by reducing the spacing between viewpoints. Figure 6 shows

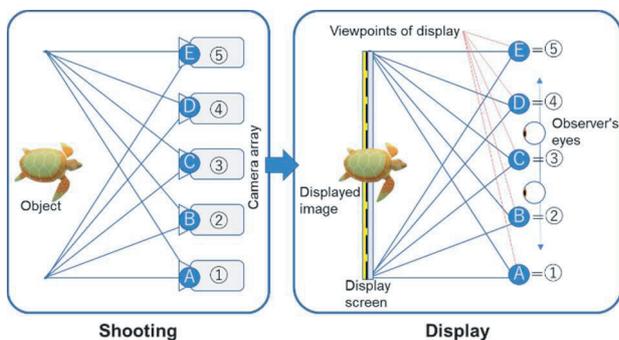


Fig. 5 Normal shooting condition. Relationships between objects and camera positions are the same as those between displayed images and viewpoints of the display.

an example of using Method A to produce a half depth range. VELF generates viewpoint images A-E from camera images ②-④. Viewpoint images A, C and E are camera images ②, ③ and ④, respectively. Viewpoint images B and D can be generated by linear blending of the two nearest camera images ②, ③ and ③, ④, respectively.

The second method (Method B), calculates the observed image from far viewpoints. Figure 7 shows an example of using Method B to attain half depth range. It uses linear blending to generate viewpoint images A-E from camera images ①-⑤. In this method, the blending ratio of camera images ①-⑤ varies depending on the horizontal position on the display screen. The graphs correspond to the weights of elemental viewpoint images shown in Fig. 4.

Though these two methods provide half depth range, blending conditions are different. We compared these two methods in terms of displayed image distortion in 3D space. As shown in Fig. 8, these two methods reduce the depth range by half. However, Method A expands object width at the front while that at the rear is compressed. That is, objects are distorted in 3D space. On the other hand, Method B yields natural depth compression without unexpected distortion. Therefore, we choose Method B.

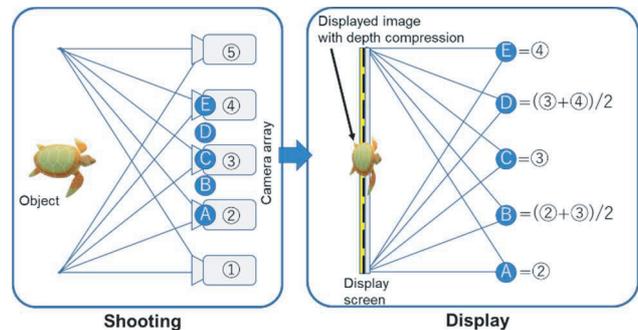


Fig. 6 Depth range compression by reducing disparity of viewpoint image using VELF (Method A). Viewpoint images of display A-E are generated using linear blending of camera images.

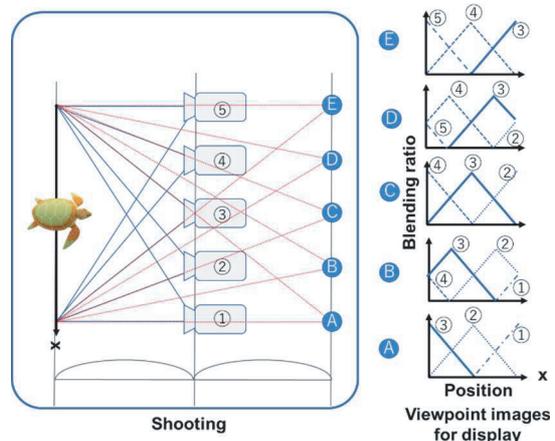


Fig. 7 Depth range compression (1/2) by displaying far viewpoint images using VELF (Method B).

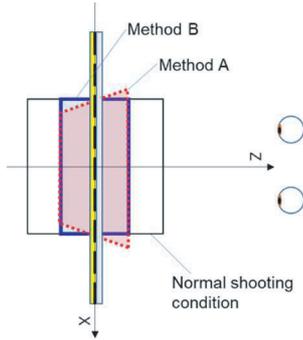


Fig. 8 Image distortions in 3D space induced by depth compression.

Figure 9 shows images (a) under normal shooting condition at viewpoints ㉔ (left) and ㉔ (right) and (b) and (c) compressed depth images using Method B at viewpoints ㉔ (left) and ㉔ (right). The compression ratios in thickness direction were 1/2 and 1/6, respectively. Though the size of the turtle as a 2D image is the same, its disparity is reduced, and the turtle becomes thinner. Natural depth compressed images can be achieved. The VELF3D display was confirmed to offer natural 3D images and depth control.

4. Discussion

4.1 Limitation of Depth Compression

From the viewpoint of VELF3D application, generated images should not be corrupted by depth compression. We assume that there was no corruption in the normal shooting condition and discuss the limitation of display viewpoint image generation.

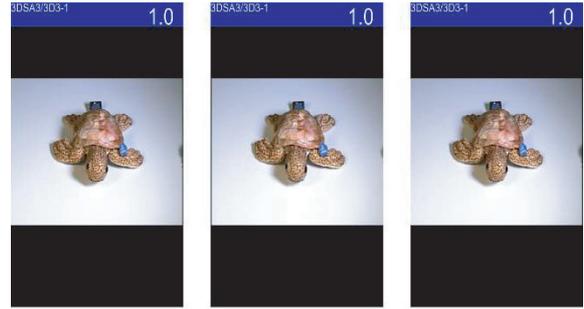
We consider two issues with our method. First is the vignetting induced by a lack of camera images and the second one is the limitation of disparity between neighboring viewpoint images.

In the case of Method A, since viewpoints are on the line of the camera array and viewpoint spacing is smaller than that of cameras, no vignetting occurs.

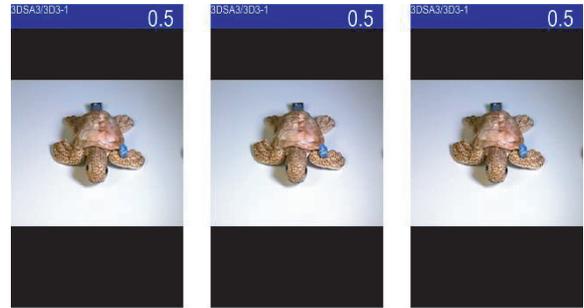
In the case of Method B, viewpoints are not on the line of the camera array. Figure 10 shows image generation for higher depth compression (1/3) than that of Fig. 7 (1/2). Viewpoint of image generation is 1.5 times distant.

For the center viewpoint ㉔, though images from three camera are used in Fig. 4, those of five cameras are used in Fig. 10. No vignetting occurs in these two cases. Even if the display viewpoint distance becomes infinite, rays became parallel and no vignetting occurs, because in this configuration, the display has smaller width than the camera array. For the viewpoint at the right end, viewpoint ㉔, images from three cameras are used in Fig. 7 and those of four cameras are used in Fig. 10. Also in these cases, no vignetting occurs. If the display viewpoint distance becomes infinite, rays became parallel and no vignetting occurs, as the display has smaller width than the camera array.

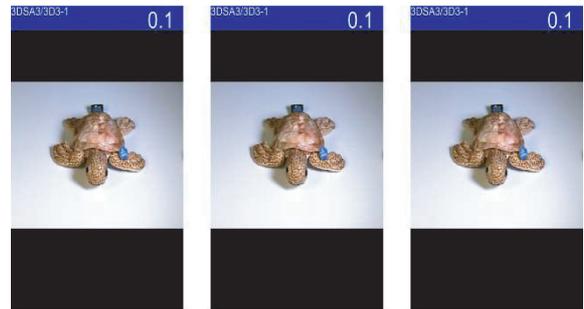
Next we discuss disparity. VELF sets a limit on dis-



(a) Normal condition shooting images (depth: 100%)



(b) Image generated by Method B (depth: 50%)



(c) Image generated by Method B (depth: 17%)

Fig. 9 Original and generated images. Images are arranged in the order of left, right and left for free-fuse.

parity due to the interpolation imposed by linear blending. When the disparity between blended images is large, double images are observed. Therefore, disparities between neighboring two viewpoint images must be less than 3-5 arcmins in visual angle.

In the case of Method A, disparities between neighboring viewpoint images are smaller than those of camera images. When the condition of linear blending is satisfied in the normal shooting condition, no double images are observed with Method A. In the case of Method B, though blending ratios depend on the position on the screen, disparities of blended images are less than the limit of linear blending. That is, if the disparity is small enough for the normal shooting condition, depth of displayed image can be

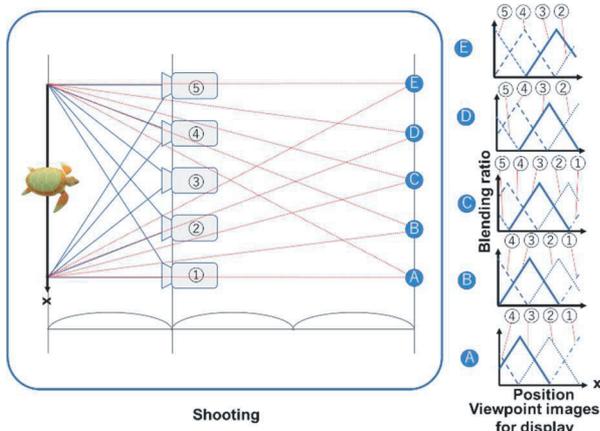


Fig. 10 Depth range compression (1/3) by increasing the disparity of farther viewpoint images (Method B).

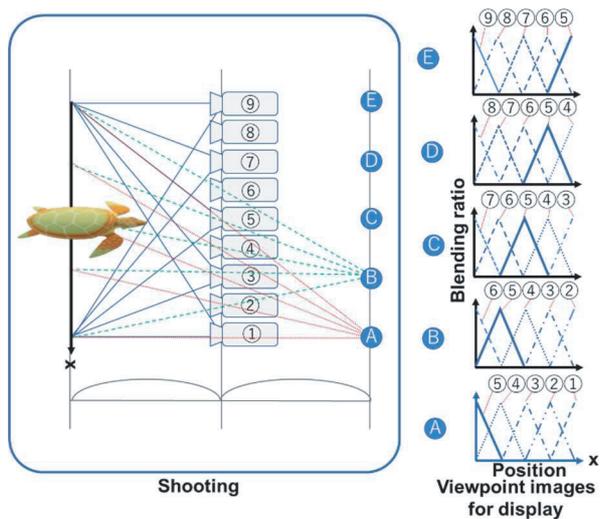


Fig. 11 Depth range compression (1/2) by displaying far viewpoint images using VELF3D for large depth object (Method B).

compressed to any degree. However, if the object has large depth, camera image disparity exceeds the limit of the linear blending condition and the viewpoint image generation fails.

4.2 Shooting Large Depth Objects

To shoot large depth objects, there are two approaches. The first one improves viewpoint image generation. For example, if coarse depth information can be obtained, viewpoint images can be generated without failure [21]. However, this incurs increased calculation cost.

The second method increases the number of cameras as shown in Fig. 11. Four cameras are added to the spaces between the original cameras, which are shown in Fig. 5. This approach cuts the disparity in half. Therefore, even if object depth is doubled, no image failure occurs in the normal shooting condition, and any degree of depth compression is possible.

5. Conclusions

We have developed a depth range compression method and confirmed its feasibility using a VELF3D display. It incurs only the light image processing cost of parallelizable addition and multiplication operations, as linear blending is based on weighted averaging. When contents are shot without image failure in the normal shooting condition, display depth range can be compressed to any extent without failure.

Since our method doesn't require camera position changes or complex calculations such as depth map generation, it is fast, exact, and practical enough that it can support live action. Our method is not limited to VELF3D displays. It is applicable to viewpoint image generation in other 3D displays, because any ray can be reproduced if the ray passes through the VELF viewing zone.

References

- [1] M. Levoy and P. Hanrahan, "Light Field Rendering," Proc. ACM SIGGRAPH, ACM Press, pp.31–42, 1996. 10.1145/237170.237199
- [2] R.L.D. Valois, H. Morgan, and D.M. Snodderly, "Psychophysical studies of monkey vision III. Spatial luminance contrast sensitivity tests of macaque and human observers," Vision Research, vol.14, no.1, pp.75–81, 1974. 10.1016/0042-6989(74)90118-7
- [3] F.W. Campbell and J.G. Robson, "Application of Fourier analysis to the visibility of gratings," Journal of Physiology, vol.197, no.3, pp.551–566, 1968. 10.1113/jphysiol.1968.sp008574
- [4] E.A. Wulffing, "Über den kleinsten Gesichtswinkel," Zeitschrift für Biologie, vol.29, pp.199–202, 1892.
- [5] G. Westheimer and S.P. McKee, "Spatial configuration for visual hyperacuity," Vision Research, vol.17, no.8, pp.941–947, 1977. 10.1016/0042-6989(77)90069-4
- [6] H. Oishi, H. Takemura, S.C. Aoki, I. Fujita, and K. Amano, "Microstructural properties of the vertical occipital fasciculus explain the variability in human stereoacuity," Proceedings of the National Academy of Sciences of the United States of America, vol.115, no.48, pp.12289–12294, 2018. 10.1073/pnas.1804741115
- [7] C.W. Tyler, "Spatial organization of binocular disparity sensitivity," Vision Research, vol.15, no.5, pp.583–590, 1975. 10.1016/0042-6989(75)90306-5
- [8] S. Suyama, S. Ohtsuka, H. Takada, K. Uehira, and S. Sakai, "Apparent 3-D Image Perceived from Luminance-Modulated Two 2-D Images Displayed at Different Depths," Vision Research, vol.44, no.8, pp.785–793, 2004. 10.1016/j.visres.2003.10.023
- [9] M. Date, H. Takada, Y. Gotoh, and S. Suyama, "Depth reproducibility for inclined view in DFD (depth fused 3-D display)," Proceedings of the International Display Workshops, vol.19, pp.1793–1794, 2005.
- [10] M. Date, Y. Andoh, H. Takada, Y. Ohtani, N. Matsuura, "Depth Reproducibility of Multiview Depth-Fused 3-D Display," Journal of the Society for Information Display, vol.18, no.7, pp.470–475, 2010. 10.1889/JSID18.7.470
- [11] M. Date, H. Fujii, H. Kimata, T. Kojima, K. Iwata, R. Kimura, A. Sugiura, and M. Miyao, "Accommodation response for visually equivalent light field 3D display," IEEE Industry Applications Society Annual Meeting, 2017-ILDC-0863, 2017. 10.1109/IAS.2017.8101803
- [12] M. Date, H. Takada, Y. Honda, S. Ozawa, S. Mieda, and A. Kojima, "Highly Realistic 3D Display System for Space Composition Telecommunication," Journal of Display Technology, vol.11, no.2, pp.121–128, 2015. 10.1109/JDT.2014.2338858

- [13] M. Date, H. Takada, and A. Kojima, "Real-time viewpoint image synthesis using strips of multi-camera images," *Proc. SPIE*, vol.9391, pp.939109–1-939109-7, 2015. 10.1117/12.2083011
- [14] M. Date, D. Ochi, and H. Kimata, "Flat panel visually-equivalent light field 3D display," *The Journal of the Institute of Image Electronics Engineers of Japan*, vol.48, no.2, pp.264–272, 2019. (in Japanese)
- [15] M. Date, T. Kawakami, M. Sasai, and H. Takada, "Smooth Motion Parallax Autostereoscopic 3D Display Using Linear Blending of Viewing Zones," *SID Symposium Digest of Technical Papers*, vol.46, no.1, pp.983–986, 2015. 10.1002/sdtp.10300
- [16] M. Date, T. Kawakami, M. Sasai, and H. Takada, "Luminance profile control method using gradation iris for autostereoscopic 3D displays," *11th Conference on Lasers and Electro-Optics Pacific Rim*, 26B3_6, 2015.
- [17] T. Kawakami, M. Date, M. Sasai, and H. Takada, "360-degree screen-free floating 3D image in a crystal ball using a specially imaged iris and rotational multiview DFD technologies," *Applied Optics*, vol.56, no.22, pp.6156–6167, 2017. 10.1364/AO.56.006156
- [18] M. Date, M. Isogai, and H. Kimata, "Full parallax visually equivalent light field 3D display using RGB stripe liquid crystal display panel," *3DSA DIGEST, LFII-4*, 2018.
- [19] M. Date, M. Isogai, and H. Kimata, "Full Parallax Table Top 3D Display Using Visually Equivalent Light Field," *IEEE Conference on Virtual Reality and 3D User Interfaces (VR) 2019*, pp.1297–1298, 2019. 10.1109/VR.2019.8797796
- [20] M. Date et al., "Table Top Visually Equivalent Light Field 3D Display Using 15.6-inch 4K LCD Panel," *SID Display Week 2019*, 56-5, pp.791–794, 2019. 10.1002/sdtp.13040
- [21] M. Date, H. Takada, and A. Kojima, "Viewpoint image generation for head tracking 3D display using multi-camera and approximate depth information," *Journal of the Society for Information Display*, vol.23, no.8, pp.339–346, 2015. 10.1002/jsid.315



Munekazu Date received the B.S. degree in physics from Gakushuin University, Tokyo, Japan, in 1990, an M.E. degree in applied electronics from Tokyo Institute of Technology, Yokohama, Japan, in 1992, and the Ph.D. degree in chemistry from Tokyo University of Science in 2003. He was with Nippon Telegraph and Telephone Corporation (NTT), Tokyo, Japan, in 1992–2010. There, he has been engaged in research on holographic optical devices using polymer/LC composites and 3D displays. He

joined NTT COMWARE Corporation, Chiba, Japan, in 2010–2012, and engaged in research on 3D digital signage and ad-hoc communication. He rejoined NTT, Kanagawa, Japan in 2012 and engaged in research on high-reality 3D display system and its human factor. He was a recipient of IEEE IAS Technical Committee Prize Paper Award in 2014, Presidential Citations from SID in 2017, Best Paper Award from 3DSA in 2017, IEEE VR Best VRSJ Demo Award in 2019 and IDW Best Paper Award in 2019.



Universe Data Handling Project of NTT Media Intelligence Laboratories, Kanagawa, Japan.

Shinya Shimizu received the B.Eng. and M.Info. degrees in social informatics from Kyoto University, Japan and the Ph.D. degree in electrical engineering from Nagoya University, Japan, in 2002, 2004, and 2012, respectively. He joined Nippon Telegraph and Telephone Corporation (NTT) in 2004 and has engaged in R&D of image and signal processing. His research interests also include computer vision, machine learning and neuroscience. He is currently a Senior Research Engineer with the



Senior Research Engineer and also the Supervisor at NTT Media Intelligence Laboratories. He is a Chair of Technical Committee on Image Engineering of the Institute of Electronics, Information and Communication Engineers of Japan.

Hideaki Kimata received the B.E. and M.E. degrees in applied physics, and the Ph.D. degree in electrical engineering respectively from Nagoya University, Nagoya, Japan, in 1993, 1995, and 2006. He joined Nippon Telegraph and Telephone Corporation (NTT) in 1995, and has been involved in the research and development of video coding, realistic communication, computer vision, and video recognition based on machine learning (deep learning). He is currently a



IEICE Best Paper Award 2010, the IEICE KIYASU-Zen'iti Award 2010, and the IPSJ SIG-CDS Excellent Paper Award 2013.

Dan Mikami received the B.E. and M.E. degrees from Keio University, Kanagawa, Japan, in 2000 and 2002, respectively, and the Ph.D. degree from Tsukuba University in 2012. Since 2002, he has been with Nippon Telegraph and Telephone Corporation. His current research activities are mainly focused on computer vision and information technologies for enhancing sport performance. He was a recipient of the Meeting on Image Recognition and Understanding 2009, the Excellent Paper Award 2009, the



Yoshinori Kusachi is a senior research engineer in the Universe Data Handling Laboratory of NTT Media Intelligence Laboratories. He received the masters degree from Nara Institute of Science and Technology and joined NTT in 1997. He has been engaged in research and practical application development in the fields of image processing, computer vision, and pattern recognition. He received the Ph.D. degree in information engineering from Nara Institute of Science and Technology in 2007.