Light Field Compression for Compressive 3D Display

Kohei Isechi, Yuto Kobayashi, Keita Takahashi, and Toshiaki Fujii

***Abstract*—Compressive 3D display that can reproduce a light field, namely dense multi-view images, with several light-attenuating LCD panels stacked in front of a backlight has been investigated. Viewers can observe different images from different viewpoints because the light rays emitted from the backlight pass through different pixels at each light-attenuating layer depending on the outgoing directions. The transmittance patterns for the layer panels (layer patterns) are calculated so that the display can reproduce a given light field as accurately as possible. We suppose a 3D video transmission system where such compressive displays are adopted as the receiving terminals. Under this scenario, either light fields or pre-calculated layer patterns should be transmitted from the sender to a receiver. It should be noted that the layer patterns by themselves are compressive representations, because the entire light field (tens to hundreds of images) are reduced into only a few layer patterns. However, it is unclear how much the quality of the reproduced light fields is degraded due to the encoding errors of the layer patterns. To clarify this point, we compared the coding efficiencies of the light fields and layer patterns under this communication scenario. Experimental results show that the layer patterns have advantages over the light fields in terms of the rate-distortion performance in a lower bit range.**

***Index Terms*—Compressive 3D display, light field compression, rate-distortion characteristics.**

I. INTRODUCTION

As one of novel glasses-free 3D displays, a compressive light field display has been investigated [1]-[5]. This display consists of a few light-attenuating layers (e.g. LCD panels) stacked on a backlight as shown in Fig. 1. The transmittance of each layer panel can be controlled pixel by pixel individually. The light rays emitted from the backlight pass through different pixels in each layer depending on the outgoing directions; thus, the display can reproduce different images according to the viewpoints. This means that the display can reproduce a light field (dense multi-view images), providing the viewers with auto-stereoscopic images and motion parallax. The transmittance patterns displayed on the layers (layer patterns) are optimized so that the display can reproduce a given light field as accurately as possible. This optimization problem for obtaining the layer patterns is solved by using non-negative tensor factorization.

In our previous work [6], the requirements for displaying a high-quality light field with a compressive display have been analyzed, which revealed that the disparity range in the given light field should be limited in order to reproduce the high-quality 3D objects with the compressive display. Based on these requirements, an end-to-end system for displaying a real 3D scene using a compressive display with three LCD panels has been developed [7]. In this system, a real 3D scene is captured by using a multi-view camera or light field camera such as Lytro Illum [8], and then layer patterns are calculated using the captured light field. In the case of using a multi-view camera, view interpolation using image-based rendering is employed to satisfy the requirement on the disparity range.

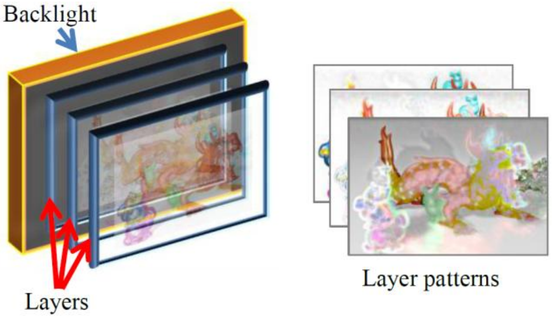


Fig. 1. Compressive 3D display and layer patterns.

In this paper, we suppose a 3D video transmission system where such compressive displays are adopted as the receiving terminals, which have not been discussed in previous works. Under this scenario, either light fields or pre-calculated layer patterns should be transmitted from the sender to the receivers, so that the compressive displays at the receiver sides can reproduce the light fields. It should be noted that the layer patterns by themselves are compressive representations, because the entire light field (tens to hundreds of images) are reduced into only a few layer patterns. However, the layer patterns are different from natural images as can be seen from Fig. 1, and thus, it is unclear how much the encoding errors of the layer patterns affect the quality of reproduced light fields. In this study, we experimentally study the coding efficiency of the light fields and layer patterns under the communication scenario with the compressive displays.

Many researchers have studied on light field image/video coding using video coding standard H.265/HEVC [9] with small modifications [10]-[18]. As a novel scheme without using H.265/HEVC, compressive representation of light field using binary images and weights has been recently studied [19]. This study proposed a method of representing a dense light field using small number of binary images and corresponding weights. These studies basically focus on the light field coding, but do not carefully consider practical scenario for light field image/video transmission. In this study, we consider a practical transmission scenario where compressive 3D displays are used as receiving terminals.

Manuscript received December 9, 2018; revised March 7, 2019

The authors are with Graduate School of Engineering, Nagoya University, Nagoya-shi, Aichi, Japan (Corresponding author: Kohei Isechi; e-mail: {isechi, kobayashi}@fujii.nuee.nagoya-u.ac.jp, {keita.takahashi, t.fujii}@nagoya-u.jp).

doi: 10.18178/ijmlc.2019.9.2.788

II. COMPRESSIVE 3D DISPLAY

This section describes the basic principle of the compressive 3D display. A light rays L emitted from the backlight is represented as follows:

$$L(\boldsymbol{x}, \boldsymbol{\theta}) = B_0 \prod_{k \in K} a_k(\boldsymbol{x} + k\boldsymbol{\theta}), \quad (1)$$

here $\boldsymbol{x} \in \mathbb{R}^2$ denotes the image coordinate and $\boldsymbol{\theta} \in \mathbb{R}^2$ is the outgoing direction of the light ray. Symbol K means the set of indices of layers, $a_k(\boldsymbol{x})$ indicates the transmittance patterns of the k-th layer, and B_0 is the intensity of the backlight. With a fixed direction $\boldsymbol{\theta}$ the ensemble of light rays $L(\boldsymbol{x}, \boldsymbol{\theta})$ corresponds to an image observed from the direction $\boldsymbol{\theta}$ which is denoted as $I_{\boldsymbol{\theta}}(\boldsymbol{x})$. To obtain the transmittance pattern $a_k(\boldsymbol{x})$, an optimization problem shown below is solved with a given set of $I_{\boldsymbol{\theta}}(\boldsymbol{x})$.

$$a_k(\boldsymbol{x}) = \arg\min_{a_k} \sum_{\boldsymbol{x},\boldsymbol{\theta}} ||I_{\boldsymbol{\theta}}(\boldsymbol{x}) - L(\boldsymbol{x}, \boldsymbol{\theta})||. \quad (2)$$

Equation (2) can be solved using non-negative tensor factorization (NTF) [5].

To improve the reproduction quality of light fields, time-division multiplexing (TDM) has also been proposed. In TDM framework, the layer patterns are rapidly alternated, and thus, the viewers perceive their average over time. In this case, Eq. (1) is rewritten as:

$$L(\boldsymbol{x}, \boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{T} \left\{ B_0 \prod_{k \in K} a_k^{(t)}(\boldsymbol{x} + k\boldsymbol{\theta}) \right\}, \quad (3)$$

here T is the number of time division, and $a_k^{(t)}$ indicates the transmittance pattern of the k-th layer at time t. These transmittance patterns can be obtained in the same manner as Eq. (2).

III. TRANSMISSION FRAMEWORK

We assume a 3D video transmission system where the compressive displays are adopted as the receiving terminals. The receiver terminals finally need the layer patterns with which the light fields are reproduced. Under this scenario, we can consider two possible transmission frameworks shown in Fig. 2 and Fig. 3. In the first framework (i), light fields are encoded at the sender side and transmitted through the communication channel, and then the corresponding layer patterns are obtained at the receiver sides using the decoded light fields. Meanwhile, in the second framework (ii), layer patterns are calculated in advance and encoded at the sender side, and the received and decoded layer patterns are directly displayed at the receiver sides. Note that in the second framework (ii) the data size is greatly reduced at the point where a light field (tens to hundreds of images) is transformed into a set of layer patterns (only a few transmittance images); it is expected that the coding efficiency of the framework (ii) would be better than that of the framework (i). However, we should take into consideration the effect of encoding errors on the layer patterns, because they are significantly different from natural images.

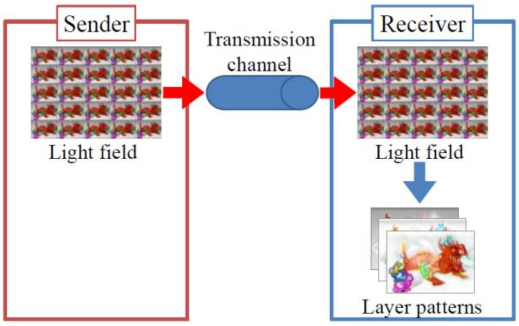


Fig. 2. Framework (i): transmitting light field.

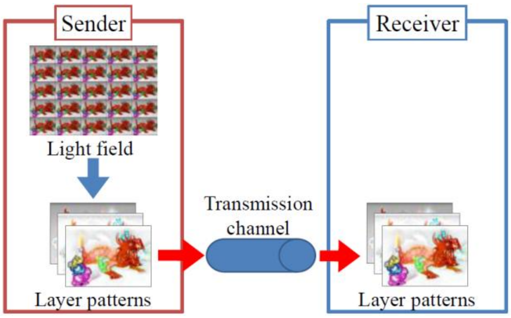


Fig. 3. Framework (ii): transmitting layer patterns.

IV. EXPERIMENT

To verify the coding efficiency of two frameworks described above, simulative experiments were conducted using two light field still-image datasets and a light field video dataset. We utilized published software for calculating layer patterns published by [20]. In experiments, a light field and layer patterns are encoded, and then reproduction of light field with the compressive display is simulated using decoded materials. The quality of displayed light field is evaluated with PSNR against the original light field. The coding efficiency is evaluated with rate-distortion characteristics describing the trade-off between the total bits of the encoded light field or layer patterns and the PSNR of displayed light field.

(a) Lego Truck (b) Amethyst (c) Animated Bunnies (first frame)

Fig. 4. Center view image of each dataset.

Light field datasets Lego Truck and Amethyst from [21] were used as the input light fields. The center view images of these dataset are shown in Fig. 4(a) and Fig. 4(b). Both datasets consist of 17×17 images, which were cropped to 640×480 and 384×512, respectively. The number of layers was set to three, and the number of time division was set to one (no time division multiplexing) and two. HEVC Test Model (HTM) 16.6 provided as H.265/HEVC reference software [22] was employed for encoding a light filed and layer patterns. For the first transmission framework, multi-view images are aligned in the raster-scan order and regarded as a single-viewpoint video. For the second framework, three layer-pattern images were also regarded as a single-viewpoint video. When the time division multiplexing was used, the first and second three patterns are individually regarded as single-viewpoint videos, and then the two videos were concatenated. The quantization parameter QP in HTM was set to 1, 10, 15, 22, 27, 32, 37, 42, and 50 to draw a curve of the rate-distortion trade-off.

Figure 8 shows the rate-distortion characteristic with Lego truck and Amethyst without time-division multiplexing. “Non-compressed” indicates the reproduction quality of the compressive display obtained with the original light field datasets; it is the upper-bound performance of the communication scenario under consideration. The center view images of the reproduced Lego Truck without TDM are shown in Fig. 5, where it is difficult to visually distinguish the difference among these images. Figure 8 demonstrates that framework (ii) outperforms framework (i) in terms of the coding efficiency; framework (ii) achieves lower bits than framework (i) with less distortion. However, the difference is not so significant as expected, given the fact that the number of images given to the HEVC encoder was 298 for framework (i) while it was only 3 for framework (ii). This means that light fields can be more efficiently compressed by using HEVC. Additionally, it seems that layer patterns are difficult to be compressed with HEVC because the layer patterns are different from natural image, for which HEVC and other video codec are optimized. Fig. 9 and Fig. 10 show the rate-distortion characteristic with time-division multiplexing. The center view images of the reproduced Lego Truck with TDM are shown in Fig. 6. It is also difficult to visually identify the difference among these images. Using TDM doubles the number of layer patterns to be encoded, and thus, more total bits were required than that without TDM in framework (ii). In this case, framework (ii) is no longer superior to framework (i). However, framework (ii) still can achieve lower total bits than framework (i) because of the small number of images to be encoded.


(a) Noncompressed (28.76dB) (b) Light field (i) (28.68dB) (c) Layer (ii) (28.64dB)

Fig. 5. Reproduction results of center view without TDM at QP = 10 (Lego Truck).


(a) Noncompressed (29.42dB) (b) Light field (i) (29.33dB) (c) Layer (ii) (29.05dB)

Fig. 6. Reproduction results of center view with TDM at QP = 10 (Lego Truck).

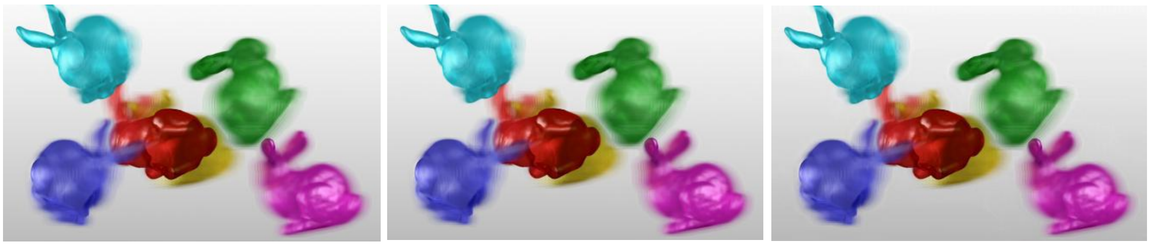
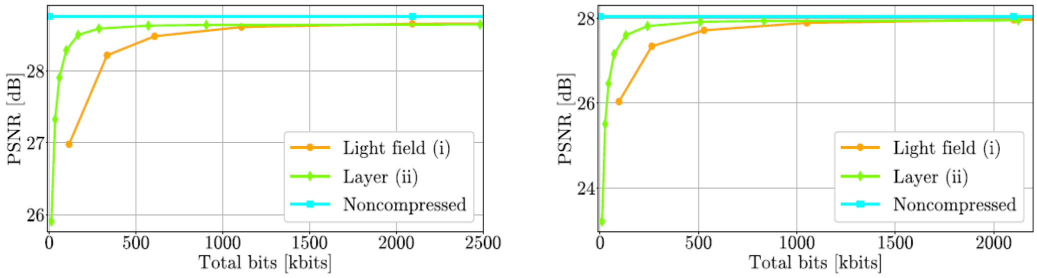

(a) Noncompressed (22.16dB) (b) Light field (i) (22.15dB) (c) Layer (ii) (22.00dB)

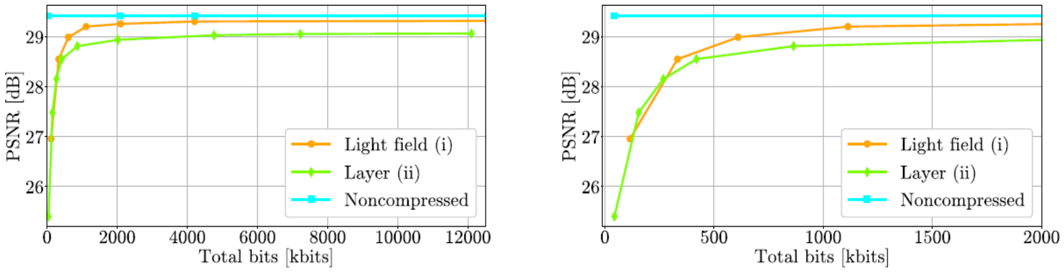
Fig. 7. Reproduction results of center view of first frame at QP = 10 (Animated Bunnies).

As a light field video dataset, Animated Bunnies from [23] was used. Fig. 4(c) shows its center view image of the first frame. Animated Bunnies was composed of 9×3 images with 89 frames and each of the images was cropped to 840×512. The number of layers was set to three, and time division multiplexing was not employed. For the first transmission framework, the light field video was treated as a multi-view videos and each video (corresponding to each viewpoint) was individually encoded using HTM. In this case, we have 27 videos, each of which consists of 89 frames. In the second framework, the patterns for each layer pattern over time were treated as a video, and individually encoded using HTM. In this case, we have 3 videos, each of which consists of 89 frames. QP was set to the same values as the experiments with the still images. The value of PSNR was obtained from the Mean Square Error (MSE) calculated over all frames, instead of taking the average of the PSNR values obtained from individual frames. Figure 11 shows the rate-distortion characteristic of Animated Bunnies. The center view images of the reproduced first frame are shown in Fig. 7. Again, frameworks (ii) surpasses framework (i) in the lower bit range.



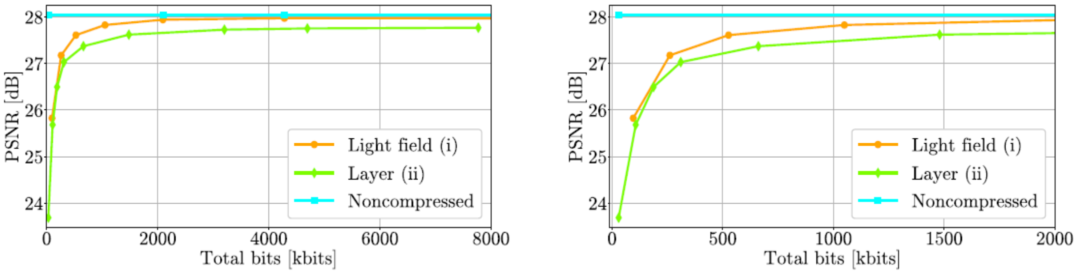
(a) Lego Truck (b)Amethyst

Fig. 8. Rate-distortion characteristics without TDM.



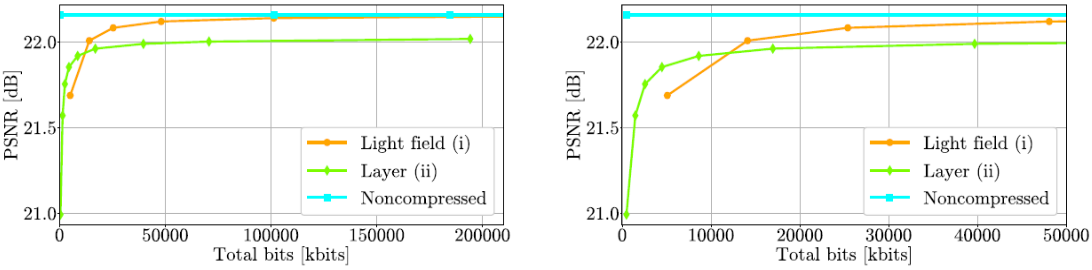
(a)Entire bit range (b)Lower bit range only

Fig. 9. Rate-distortion characteristics of Lego Truck with TDM.



(a)Entire bit range (b)Lower bit range only

Fig. 10. Rate-distortion characteristics of Amethyst with TDM.



(a)Entire bit range (b)Lower bit range only

Fig. 11. Rate-distortion characteristics of Animated Bunnies.

V. CONCLUSION

We studied the coding efficiency of light fields for compressive 3D displays. We supposed a transmission system where such compressive displays were adopted as the receiving terminals and compared two transmission frameworks where the light fields or the layer patterns were encoded and transmitted. Through simulative experiments, we demonstrated that the latter framework achieves better rate-distortion performance in lower bit ranges. However, this superiority was lost in the case with time-division multiplexing.

Our experimental validation was still limited because we only applied HEVC directly to the light fields or layer patterns. In future work, we will investigate better coding methods for the layer patterns, which will improve the rate-distortion performance of the latter transmission framework. Additionally, we will investigate a transmission scenario using a focal stack. The focal stack is composed of only a few images each of which have different focus distance, and can be used for calculating layer patterns [20]. The focal stack can be obtained by shift-and-add operations on a light field or by an ordinary camera. We can consider the focal stack as one of transmitted data instead of layer patterns.

REFERENCES

[1] S. Suyama, H. Takada, and S. Ohtsuka, "A direct-vision 3-D display using a new depth-fusing perceptual phenomenon in 2-D displays with different depths," *IEICE Trans. Electron.*, vol. E85-C, no. 11, pp. 1911–1915, 2002.

[2] D. Lanman, M. Hirsch, Y. Kim, and R. Raskar, "Content-adaptive parallax barriers: Optimizing dual-layer 3D displays using low-rank light field factorization," *ACM Trans. Graph*., vol. 29, no. 6, 2010.

[3] D. Lanman, G. Wetzstein, M. Hirsch, W. Heidrich, and R. Raskar, "Polarization fields: Dynamic light field display using multi-layer LCDs," *ACM Trans. Graph.*, vol. 30, no. 6, 2011.

[4] G. Wetzstein, D. Lanman, W. Heidrich, and R. Raskar, "Layered 3D: Tomographic image synthesis for attenuation-based light field and high dynamic range displays," *ACM Transactions on Graphics*, vol. 30, no. 4, pp. 95:1–95:12, 2011.

[5] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor displays: Compressive light field synthesis using multilayer displays with directional backlighting," *ACM Transactions on Graphics*, vol. 31, issue 4, no. 80, pp. 80:1-80:11, July 2012.

[6] T. Saito, Y. Kobayashi, K. Takahashi, and T. Fujii, "Displaying real-world light-fields with stacked multiplicative layers: Requirement and data conversion for input multi-view images," *IEEE/OSA Journal of Display Technology*, vol. 12, issue 11, pp. 1290-1300, November 2016.

[7] Y. Kobayashi, S. Kondo, K. Takahashi, and T. Fujii, "A 3-D display pipeline: capture, factorize, and display the light field of a real 3-D scene," *ITE Transactions on Media Technology and Applications*, vol. 5, no. 3, pp. 88–95, July 2017.

[8] R. Ng *et al*., "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, pp. 111, 2005.

[9] G. J. Sullivan, J. Ohm, W. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard, " *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[10] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, "Efficient intra prediction scheme for light field image compression," in *Proc. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp.539–543.

[11] F. Dai, J. Zhang, Y. Ma, and Y. Zhang, "Lenselet image compression scheme based on subaperture images streaming," in *Proc. 2015 IEEE International Conference on Image Processing (ICIP)*, Sep. 2015, pp. 4733–4737.

[12] A. Vieira, H. Duarte, C. Perra, L. Tavora, and P. Assuncao, "Data formats for high efficiency coding of Lytro-Illum light fields," in *Proc. 2015 International Conference on Image Processing Theory*, Tools and Applications (IPTA), Nov. 2015, pp.494–497.

[13] G. Tech, Y. Chen, K. Müller, J. R. Ohm, A. Vetro, and Y. K. Wang, "Overview of the multiview and 3D extensions of high efficiency video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 35–49, Jan. 2016.

[14] D. Liu, L. Wang, L. Li, Z. Xiong, F. Wu, and W. Zeng, "Pseudo-sequence-based light field image compression," in *Proc. 2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016, pp.1–4.

[15] S. Zhao, Z. Chen, K. Yang, and H. Huang, "Light field image coding with hybrid scan order," in *Proc. 2016 Visual Communications and Image Processing (VCIP)*, Nov. 2016, pp. 1–4.

[16] F. Hawary, C. Guillemot, D. Thoreau, and G. Boisson, "Scalable light field compression scheme using sparse reconstruction and restoration," in *Proc. 2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3250–3254.

[17] X. Jiang, M.L. Pendu, R.A. Farrugia, and C. Guillemot, "Light field compression with homography-based low-rank approximation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, Oct. 2017, pp. 1132–1145.

[18] R. Conceição, M. Porto, B. Zatt, and L. Agostini, "LF-CAE: Context-adaptive encoding for lenslet light fields using HEVC," in *Proc. 2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 3174–3178.

[19] K. Komatsu, K. Takahashi, and T. Fujii, "Scalable light field coding using weighted binary images," in *Proc. 2018 25th IEEE International Conference on Image Processing (ICIP)*, Oct. 2018, pp. 903–907.

[20] K. Takahashi, Y. Kobayashi, and T. Fujii, "From focal stack to tensor light-field display," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4571–4584, 2018.

[21] The (New) Stanford Light Field Archive. [Online]. Available: http://lightfield.stanford.edu

[22] Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute. High Efficiency Video Coding (HEVC) JCT-VC. [Online]. Available: https://hevc.hhi.fraunhofer.de/

[23] Synthetic Light Field Archive. [Online]. Available: https://www.media.mit.edu/~gordonw/SyntheticLightFields/

Kohei Isechi received his B.E. degree from National Institution for Academic Degrees and Quality Enhancement of Higher Education, Japan, in 2015, and his M.E. degree in electrical engineering and computer science from Nagoya University, Japan, in 2017. He is working on 3D video compression.

He is currently a Ph.D. student in information and communication at Nagoya University.

Mr. Isechi is a student member of Institute of Electronics, In-formation and Communication Engineers.

Yuto Kobayashi received his B.E. and M.E. degrees in electrical engineering from Nagoya University, Japan in 2016 and 2018, respectively.

His research topics were light field acquisition and rendering for 3-D displays when he was a student.

Keita Takahashi received his B.E., M.S., and Ph.D. degrees in information and communication engineering from the University of Tokyo, in 2001, 2003, and 2006, respectively. His research interests include computational photography, image-based rendering, and 3-D display.

He was a project assistant professor at the University of Tokyo from 2006-2011 and was an assistant professor at the University of Electro-Communications from 2011-2013. He is currently an associate professor at the Graduate School of Engineering, Nagoya University, Japan.

Assoc. Prof. Takahashi is a member of IEEE Signal Processing Society and Computer Society, Institute of Electronics, Information and Communication Engineers, Information Processing Society of Japan, and Institute of Image Information and Television Engineers.

Toshiaki Fujii received his B.E., M.E., and Dr. E. degrees in electrical engineering from the University of Tokyo in 1990, 1992, and 1995, respectively. His current research interests include multi-dimensional signal processing, multi-camera systems, multi-view video coding and transmission, free-viewpoint television, and their applications.

From 1995, he has been with the Graduate School of Engineering, Nagoya University. From 2008 to 2010, he was with the Graduate School of Science and Engineering, Tokyo Institute of Technology. He is currently a professor at the Graduate School of Engineering, Nagoya University.

Prof. Fujii is a member of IEEE Signal Processing Society, Institute of Electronics, Information and Communication Engineers, Information Processing Society of Japan, and Institute of Image Information and Television Engineers.