https://doi.org/10.2352/ISSN.2470-1173.2021.2.SDA-012

Enhancing Angular Resolution of Layered Light-Field Display by Using Monochrome Layers

Kotaro Matsuura, Keita Takahashi, Toshiaki Fujii; Nagoya University, Nagoya, Japan

Abstract

A layered light-field display is composed of several liquid crystal layers located in front of a backlight. The light rays emitted from the backlight intersect with different pixels on the layers depending on the outgoing directions. Therefore, this display can show multi-view images (a light field) in accordance with the viewing direction. This type of displays can also be used for head-mounted displays (HMDs) thanks to its dense angular resolution. The angular resolution is an important factor, because sufficiently dense angular resolution can provide accommodation cues, preventing visual discomfort caused by vergence accommodation conflict. To further enhance the angular resolution of a layered display, we propose to replace some of the layers with monochrome layers. While keeping the pixel size unchanged, our method can achieve three times higher resolution than baseline architecture in the horizontal direction. To obtain a set of color and monochrome layer patterns for a target light field, we developed two computation methods based on non-negative tensor factorization and a convolutional neural network, respectively.

Introduction

Three dimentional displays have been developed for enhancing immersive visual experiences. Among various display architectures such as those with parallax barriers [1–4] and lenticular screens [5, 6], we focus on layered light-field displays [7–12] due to their potential for presenting many views with limited hardware resources. This type of display consists of several LCD panels stacked in front of a backlight. The light rays emitted from a single point on the backlight pass through different positions on the layers depending on the outgoing directions. Therefore, these light rays can take different colors/intensities depending on the outgoing directions. As a result, this display can show multi-view images (a light field) simultaneously, presenting different images to different viewing directions.

Light-field displays have been applied for head-mounted displays (HMDs) [13–17], where two individual displays are used for left and right eyes. In this case, binocular parallax is already supported by using two different displays. However, usig a light field display for each eye has a benefit of providing accomodation cues. When the angular resolution of each display is sufficiently dense, not a single but several views can be presented for the pupil of a single eye, which provides accomodation cues and helps inducing correct focus adjustment. Providing accomodation cues prevents visual discomfort caused by vergence accommodation conflict [18,19], a common problem with ordinary HMD displays.

In this paper, we aim to improve the angular resolution of a layered light-field display in view of the application for HMDs. As shown in Fig. 1, the augular resolution is determiend by the pixel size and the interval between the layers. Several approaches can be considered to enhance the angular resolution. The first approach is to reduce the pixel size. In Kobayashi *et al.* [20], the pixels on the layers were made smaller than those of the displayed images. However, fabricating smaller pixels is not easy when the pixels are already very small, as in the case with HMDs. Moreover, using smaller pixels leads to problems related to diffractions. Another approach is to increase the distance between the layers. However, this leads to an increase of the spatial volume of the display. Meanwhile, we propose a method using monochrome layers. Our method does not change the pixel size nor the layer intervals. Instead, one or two of the layers were replaced by monochrome layers. We show that by this replacement, the angular resolution can be increased three times in the horizontal direction without sacrificing color reproduction and image quality for each view.

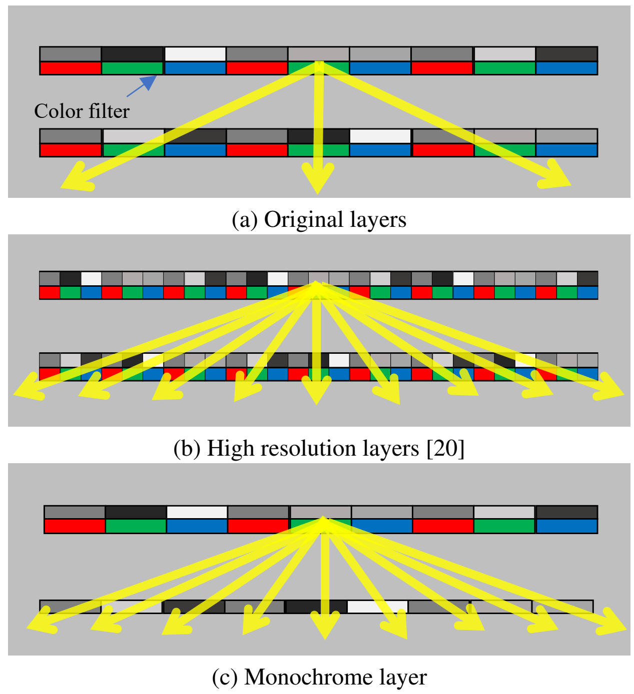

Figure 1. *Comparison of layer configurations*

Proposed Method

Layer Configurations

We assume that the display has two layers, but the following discussion can be extended to different numbers of layers. Shown in Fig. 1(a) is the structure of the original layered display (the baseline structure). Both layers are implementd with color dis-

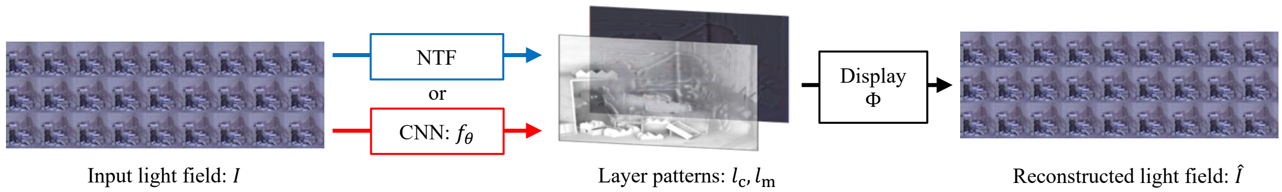

Figure 2. *Two process flows for obtaining layer patterns for display*

play panels, and thus, color filters are attached to both the display panels. A single pixel on the display panel is represented by a set of three subpixels corresponding to its red, green, and blue components aligned in the horizontal direction. The augular resolution of the display depends on the pixel size; the horizontal unit angle (i.e., the minimal distinguishable angle) is determined by the width of three subpixels. Figure 1(b) shows a case with Kobayashi *et al.* [20], which has three-times smaller pixels than the baseline structure. In this case, the angular resolution becomes three times denser than the baseline structure. However, fabricating smaller pixels involves several issues as mentioned before.

Meanwhile, we propose a method using monochrome layers as shown in Fig. 1(c). We do not change the pixel size, but use a monochorme display panel without color filters as the front layer. Eliminating the color filters allows all light-rays to pass through the layer regardless of the wavelength. Therefore, the horizontal unit angle is determined by the width of a single subpixel. As a result, the angular resolution can theoretically be increased up to three times in the horizontal direction compared to the baseline structure.

Layer Pattern Optimization

We need to obtain the transmitance patterns for the layers (layer patterns) to display a target light field, i.e., a set of multi-view images of a target 3-D content. To this end, the layer patterns should be optimized so as to minimize the error between the multi-view images given as the input and those emitted from (or recosntructed by) the display. To achieve this, we developed two methods based on non-negative tensor factorization (NTF) and convolutional neural network (CNN), as shown in Fig. 2.

NTF-based method

In the first method, we directly optimize the layers patterns using the multiplicative update rule of non-negative tensor factorization (NTF) [21, 22]. Similar appproaches were adopted in preivous works [7, 12, 20, 23], but we tailored our method so as to support our layer configuration having monochrome layers. The optimization is described as

$$\arg\min_{l_m, l_c} ||I - \Phi(l_m, l_c)||^2 \qquad (1)$$

where I denote the set of input multi-view images, and l_m and l_c are the patterns for the monochrome (front) and color (back) layers. I is represented as a 5-D tensor , $I(s,t,x,y,ch)$, where (s,t) denotes the viewpoint, (x,y) denotes the pixel position, and ch denotes the color channel. We assume that the displayed images have H (height) $\times$ W (width) pixels and three color channels. Φ denotes the display's physical process of reconstructing multi-view images from the layers, and is written as

$$\begin{aligned} \Phi(l_m, l_c) &= \hat{I}(s,t,x,y,ch) \\ &= l_m(y+t, 3x+ch+s) l_c(y,x,ch) B \end{aligned} \qquad (2)$$

where B is the luminance of the backlight, and l_m and l_c are represented as $H \times 3W$ and $H \times W \times 3$ tensors, respectively. At the begining of optimization, l_m and l_c are initialized to random values. The update rule is applied to l_m and l_c alternatively. As the number of updates increases, the reconstruction quality improves, but it also increases the computation time.

We implemented the NTF-based method using CuPy ver 7.7.0, based on the software provided by Maruyama *et al.* [23].

CNN-based method

The second method is implemented as a convolutional neural network (CNN), similarly to Maruyama *et al.* [23]. The network, denoted by f_θ, learns the mapping between the input multi-view images (I) and the corresponding layer patterns (l_m and l_c). The parameters of the network (θ) are optimized through the training process on a large amount of data ($\mathscr{D}$).

$$\arg\min_{\theta} \sum_{I \in \mathscr{D}} ||I - \Phi(f_\theta(I))||^2 \qquad (3)$$

Table 1. Top: architecture of entire network (in: input to layer, chns: number of input/output channels, act: activation function, n_l: number of output layers (2 or 3).

layer	in	chns	act
input	I		
conv2D-1a	input	27/64	ReLU
conv2D-1b	conv2D-1a	64/64	ReLU
conv2D-1c	conv2D-1b	64/64	
Add-1	conv2D-1a + conv2D-1c		ReLU
conv2D-2a	Add-1	64/64	ReLU
conv2D-2b	conv2D-2a	64/64	ReLU
conv2D-2c	conv2D-2b	64/64	
Add-2	conv2D-2a + conv2D-2c		ReLU
⋮	⋮		
conv2D-24a	Add-24	64/64	ReLU
conv2D-24b	conv2D-24a	64/64	ReLU
conv2D-24c	conv2D-24b	64/64	
Add-24	conv2D-24a + conv2D-24c		ReLU
conv2D-L	Add-24	64/n_l	Hard Sigmoid
output	conv2D-L		

(a) Original layers (b) High resolution layers (c) Monochrome layer

Figure 3. *Comparison on layer configurations: displayable multi-view images (top) and reconstructed images (bottom).*

The network architecture is summarized in Table 1. We adopted a similar architecture as that in [23], but increased the number of convolutional layers to improve the accuracy. Our network is composed of a stack of 2-D convolutinal layers, for all of which, the kernel size and strides are set to 3×3 and $(1,1)$, respectively. We adopt boundary paddings to keep the image size unchanged before and after the convolution. The inputs to the network is a set of multi-view images from 3×9 viewpoints and the outputs from the network is a set of layer patterns. The input is reshaped to a tensor of $H\times3W\times27$, where three color components are expanded in the width dimension, and 3×9 images are stacked along the channel dimension:

$$I(s,t,x,y,ch)\rightarrow I(y,3x+ch,9t+s) \qquad (4)$$

The output consists of the color and monochrome layers (l_m and l_c), each of which is represented as a $H\times3W$ tensor. The number of channels in the output (n_l) corresponds to the number of layers, which is 2 or 3 in this paper. For the color layer l_c, the color components are included in the width demension, and thus, it should be reshaped as

$$l_\mathrm{c}(y,3x+ch)\rightarrow l_\mathrm{c}(y,x,ch) \qquad (5)$$

Finally, from these layers, multi-view images were reconstructed using Eq. (2).

We collected the traing data from two light field datasets [24, 25]. Each image in a light field was cropped into a 64×64 region, and a collection of these resions from 27 viewpoints constitute a training sample. We applied data augmentation to the luminance values of each sample, and finally collected 55K samples. Our network was implemented using PyTorch ver 1.3.1, a Python-based framework for deep neural network. We used the mean squared error as the loss function. We trained the network over 20 epoches using the batch size of 15 and built-in Adam optimizer.

Experimental Results

We first compared the three layer configurations shown in Fig. 1 using the NTF-based layer optimization method. Throughout this comparsion, the pixel size of the displayed images I was set to that of a single color pixel (three subpixels) of the original layers. The number of layers was set to 2, and the number of updates was fixed to 50.

The results are presented in Fig. 3. As shown in the top row, the number of views reconstructed from the display was 3×3 for the case of the original layers (a), but it was 3×9 for the cases with the high-resulution layers (b) and the monochrome layer (c). The latter two configurations can increase the angular resolution three times in the horizontal direction, because the horizontal unit angle was decreased three times lower than that of the original layers. Close-ups from reconstructed images are shown in the bottom of Fig. 3. Note there that in our method (c), the colors of the target object are clearly reconstructed even though one of the layers had no color filters. The reconstruction quality was measured as PSNR values, which were 34.84 dB for the high-resolution layers (b) and 34.04 dB for the monochrome layer (c). These values were first calculated for each color channel and each viewpoint, and then, avaraged over all the channels and viewpoints. Note again that our method with the monochrome layer achieved this quality using three times less pixels than the high-resolution layers.

Next, we evaluated the performance of our method with different numbers of layers and different optimization methods. The number of layers was set to 2 or 3, and both the NTF-based and CNN-based optimization methods were implemented. In the case of three layers, we kept the middle layer as a color layer, but set the front and back layers to monochorme layers. Figure 4 shows layer patterns and reconstructed images obtained from each configuration. We can see that our method can reproduce textures and colors even though only a single layer had color filters. Us-

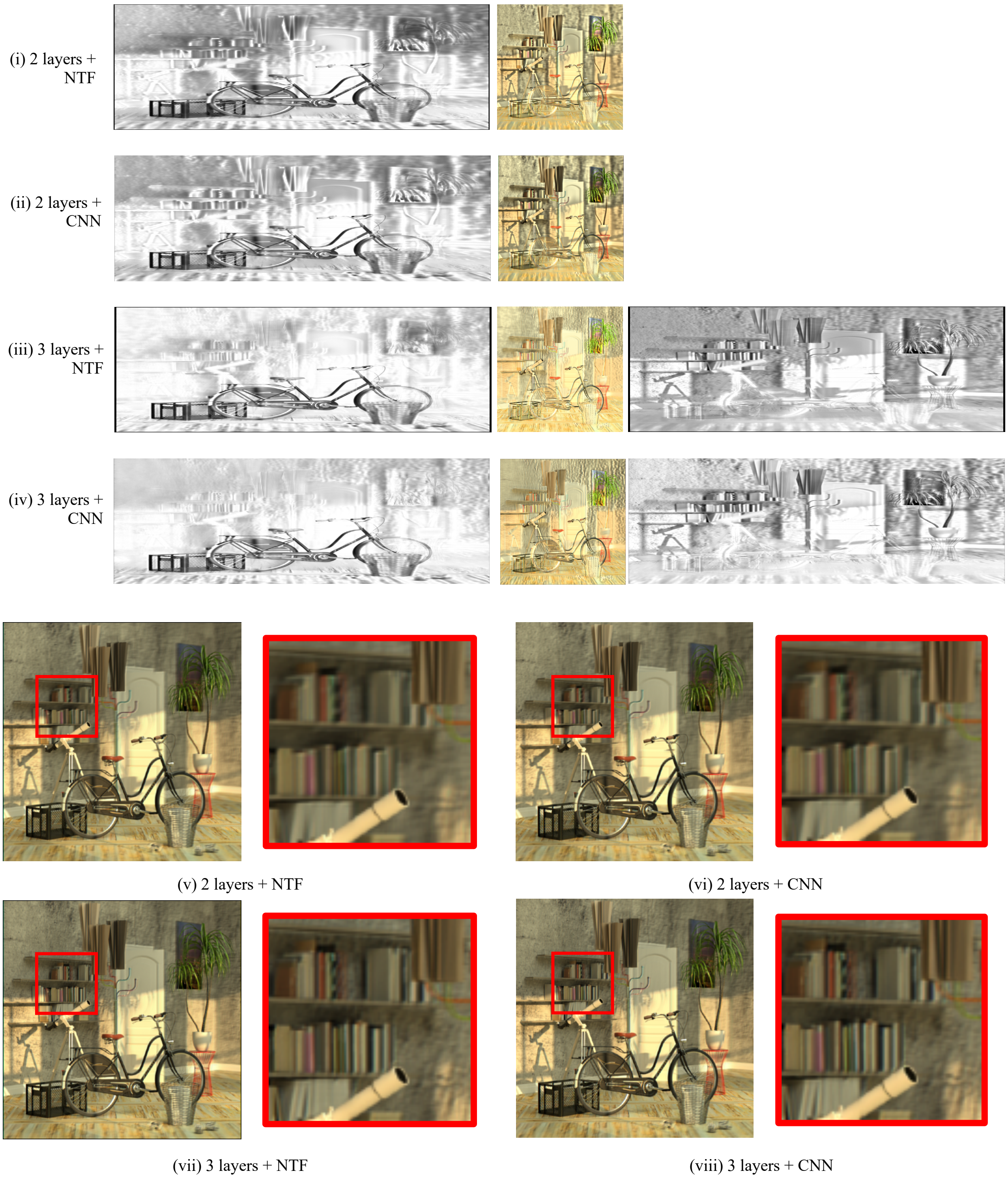

Figure 4. *Results obtained using monochrome layers; layer patterns (top) and reconstructed images (bottom).*

Table 2: Performance of monochrome layers with different number of layers and different optimization methods.

Method	PSNR ↑	SSIM ↑	time [s]
2 layers + NTF	26.29	0.8639	2.950
2 layers + CNN	26.33	0.8636	0.271
3 layers + NTF	27.95	0.9043	3.819
3 layers + CNN	27.90	0.9033	0.272

ing three layers lead to better reconstruction quality than the case with two layers. Table 2 summarizes PSNR, SSIM, and computation time for each configuration. The PSNR and SSIM values were first calculated for each color channel and each viewpoint, and then, avaraged over all the channels and viewpoints. For this experiment, we used a PC with an Intel Core i9-9900K Processer, 32 Gbyte main memory, and a NVIDIA Geforce RTX 2080 Ti graphics card. For the same number of layers, the reconstruction quality was almost the same regardless of the optimization method. However, the CNN-based method had a clear advantage in terms of the computation time over the NTF-based method.

Conclusions

We proposed a method for increasing the angular resolution of layered light-field displays using monochrome layers. We also developed NTF-based and CNN-based methods for computing the layer patterns for the given multi-view images. We confirmed that our method can successfully increase the angular resolution three times in the horizontal direction, while keeping the image quality of color reproduction. It is notable that even though the size of pixels on the layers was kept unchanged, our method achieved similar reconstruction quality to the higher resolution layers. Our future work will include implementation of our method on real hardware.

References

[1] F. E. Ives, "Parallax stereogram and process of making same." Apr. 14 1903, US Patent 725,567.

[2] H. Isono, M. Yasuda, and H. Sasazawa, "Autostereoscopic 3-D display using LCD-generated parallax barrier," *Electronics and Communications in Japan (Part II: Electronics)*, vol. 76, no. 7, pp. 77–84, 1993.

[3] K. Sakamoto and T. Morii, "Multi-view 3D display using parallax barrier combined with polarizer," in *Advanced Free-Space Optical Communication Techniques/Applications II and Photonic Components/Architectures for Microwave Systems and Displays*, vol. 6399. International Society for Optics and Photonics, 2006, p. 63990R.

[4] T. Peterka, R. L. Kooima, D. J. Sandin, A. Johnson, J. Leigh, and T. A. DeFanti, "Advances in the dynallax solid-state dynamic parallax barrier autostereoscopic visualization display system," *IEEE transactions on visualization and computer graphics*, vol. 14, no. 3, pp. 487–499, 2008.

[5] G. Lippmann, "Epreuves reversibles donnant la sensation du relief," 1908.

[6] J. Arai, F. Okano, H. Hoshino, and I. Yuyama, "Gradient-index lens-array method based on real-time integral photography for three-dimensional images," *Applied optics*, vol. 37, no. 11, pp. 2034–2045, 1998.

[7] G. Wetzstein, D. R. Lanman, M. W. Hirsch, and R. Raskar, "Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting," vol. 31, no. 4, pp. 1–11, 2012.

[8] 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 transactions on electronics*, vol. 85, no. 11, pp. 1911–1915, 2002.

[9] D. Lanman, M. Hirsch, Y. Kim, and R. Raskar, "Content-adaptive parallax barriers: optimizing dual-layer 3D displays using low-rank light field factorization," in *ACM SIGGRAPH Asia 2010 papers*, 2010, pp. 1–10.

[10] D. Lanman, G. Wetzstein, M. Hirsch, W. Heidrich, and R. Raskar, "Polarization fields: dynamic light field display using multi-layer LCDs," in *Proceedings of the 2011 SIGGRAPH Asia Conference*, 2011, pp. 1–10.

[11] ——, "Beyond parallax barriers: applying formal optimization methods to multilayer automultiscopic displays," in *Stereoscopic Displays and Applications XXIII*, vol. 8288. International Society for Optics and Photonics, 2012, p. 82880A.

[12] 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.

[13] F.-C. Huang, D. P. Luebke, and G. Wetzstein, "The light field stereoscope." in *SIGGRAPH Emerging Technologies*, 2015, pp. 24–1.

[14] D. Lanman and D. Luebke, "Near-eye light field displays," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, pp. 1–10, 2013.

[15] H. Hua and B. Javidi, "A 3D integral imaging optical see-through head-mounted display," *Optics express*, vol. 22, no. 11, pp. 13 484–13 491, 2014.

[16] W. Song, Y. Wang, D. Cheng, and Y. Liu, "Light field head-mounted display with correct focus cue using micro structure array," *Chinese Optics Letters*, vol. 12, no. 6, p. 060010, 2014.

[17] C. Yao, D. Cheng, and Y. Wang, "Design and stray light analysis of a lenslet-array-based see-through light-field near-eye display," in *Digital Optics for Immersive Displays*, vol. 10676. International Society for Optics and Photonics, 2018, p. 106761A.

[18] G. Kramida, "Resolving the vergence-accommodation conflict in head-mounted displays," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 7, pp. 1912–1931, 2015.

[19] T. Ueno and Y. Takaki, "Super multi-view near-eye display to solve vergence–accommodation conflict," *Optics express*, vol. 26, no. 23, pp. 30 703–30 715, 2018.

[20] Y. Kobayashi, K. Takahashi, and T. Fujii, "Using high-resolution binary layers and a low-resolution multibit backlight for a layered light-field display," *Optical Engineering*, vol. 57, no. 6, p. 061607, 2018.

[21] A. Cichocki, R. Zdunek, and S.-i. Amari, "Nonnegative matrix and tensor factorization [lecture notes]," *IEEE signal processing magazine*, vol. 25, no. 1, pp. 142–145, 2007.

[22] V. D. Blondel, N.-D. Ho, P. Dooren *et al.*, "Weighted nonnegative matrix factorization and face feature extraction," in *In Image and Vision Computing*. Citeseer, 2008.

[23] K. Maruyama, K. Takahashi, and T. Fujii, "Comparison of layer operations and optimization methods for light field display," *IEEE Access*, vol. 8, pp. 38 767–38 775, 2020.

[24] Computer Graphics Laboratory, Stanford University, "The (new) stanford light field archive," http://lightfield.stanford.edu, 2018.

[25] Heidelberg Collaboratory for Image Processing, "4D light field dataset," http://hci-lightfield.iwr.uni-heidelberg.de/, 2018.

Author Biography

Kotaro Matsuura received his B.E. in electrical engineering from Nagoya University, Japan, in 2020. He is currently a graduate student at the Graduate School of Engineering, Nagoya University, Japan. His research interest includes 3-D display and related rendering techniques.

Keita Takahashi received the B.E., M.S., and Ph.D. degrees in information and communication engineering from The University of Tokyo, in 2001, 2003, and 2006, respectively. From 2006 to 2011, he was a Project Assistant Professor with The University of Tokyo. From 2011 to 2013, he was an Assistant Professor with The University of Electro-Communications. He is currently an Associate Professor with the Graduate School of Engineering, Nagoya University, Japan. His research interests include computational photography, image-based rendering, and 3D displays.

Toshiaki Fujii received the B.E., M.E., and Dr.E. degrees in electrical engineering from The University of Tokyo, in 1990, 1992, and 1995, respectively. From 2008 to 2010, he was with the Graduate School of Science and Engineering, Tokyo Institute of Technology. Since 1995, he has been with the Graduate School of Engineering, Nagoya University, where he is currently a Professor. His current research interests include multidimensional signal processing, multicamera systems, multiview video coding and transmission, free-viewpoint television, and their applications.