Head-Tracking Layered Light-Field Display

Kotaro Matsuura, **Chihiro Tsutake**, **Keita Takahashi**, **Toshiaki Fujii**

Corresponding Author: Kotaro Matsuura (koutarou.m@fujii.nuee.nagoya-u.ac.jp)
Department of Information and Communication Engineering, Graduate School of Engineering, Nagoya University, Japan
Keywords: layered light-field display, light field, wide viewing area, head-tracking display

ABSTRACT

To extend the viewing angle while keeping the image quality, we incorporate head-tracking into a layered light field display. Our display is designed to show only a local light field at a time, which will cover the viewing angle around the currently estimated head position.

1 Introduction

A layered light field display [1, 2, 3] is composed of several semi-transparent panels (such as LCD panels) stacked in front of a backlight. The light rays emitted from a single point of the backlight pass through a different set of pixels depending on the outgoing directions, resulting in having direction-dependent luminances. Thanks to this mechanism, the display can output different images toward different directions, which are referred to as directional views.

The layered architecture has several advantages over others in displaying 3-D images. In particular, unlike other architectures for naked eyes, such as parallax barriers [4, 5] and lenticular screens [6, 7], the layered architecture can maintain the spatial resolution of each directional view regardless of the number of viewpoints. However, it can not escape from the trade-off between the number of viewpoints and the image quality of each directional view. The image quality for each view deteriorates as the number of viewpoints increases, due to the limited capacity of the layers with finite numbers of pixels.

The number of viewpoints is an important factor for user experience because it directly corresponds to the angular range of the viewing zone (viewing angle) in front of the display. Therefore, we aim to widen the viewing angle while keeping the image quality. To circumvent the inherent trade-off between the number of viewpoints and image quality, we incorporate head tracking into a layered light field display, as shown in Fig. 1. The key idea behind our proposal is that we can maintain high image quality for a *local light field*, which is a partial set of views taken from the entire target light field. Using the result of head-tracking, we switch the local light fields to be displayed, so as to support the viewing angle around the current head position. Unlike conventional head-tracking 3-D displays [8, 9, 10], our method can tolerate errors and time delays of the head-tracking system to some extent, because the viewing angle supported by the display

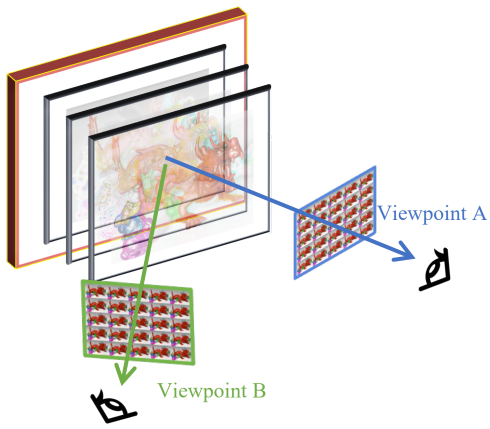

Fig. 1: Head-tracking layered light-field display

is not limited to a single direction, but includes a range of directions corresponding to the size of the local light field. Experimental results are presented to show the feasibility and effectiveness of our proposal.

2 Layered Light-Field Display

2.1 Display Principle

We assume that the display is composed of three semi-transparent layers located at depths $z \in \{-1, 0, 1\}$. Given a set of layer patterns $\{P_z\}$, each of the light rays is generated as the product of the layers' transmittance values along the path. More speficifically, the display emit a set of light rays, $L(u, v, x, y)$, as

$$L(u, v, x, y) = \prod_z P_z(x + zu, y + zv) := \Phi(\{P_z\}) \quad (1)$$

where (x, y) denotes spatial location (pixel position) on the central layer, and (u, v) denotes the outgoing direction of the light ray. Without losing generality, the perpendicular direction against the layers is denoted as $(u, v) = (0, 0)$.

The set of light rays, $L(u, v, x, y)$, can be regarded as a light field composed of a set of directional views, where (u, v) denotes the viewpoint, (x, y) denotes the pixel position of each directional view. We assign $(u, v) = (0, 0)$ to the central viewpoint of the light field. For brevity of

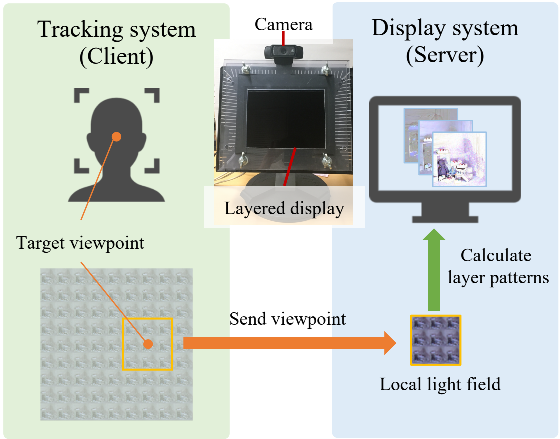

Fig. 2: System configuration (client and server system)

description, we use an operator Φ to describe the mapping process from the layer patterns to the emitted light field.

2.2 Obtaining Layer Patterns

To display a target light field, we need to obtain appropriate layer patterns that will reproduce the target light field as accurately as possible.

To find a set of layer patterns from a given light field, Wetzstein et al. [1] proposed an analytical method based on non-negative tensor factorization (NTF). However, due to the iterative nature of the algorithm, this method takes significant computation time until convergence.

Maruyama et al. [11] proposed a learning-based method using a convolutional neural network (CNN), which enables much faster computation. In their method, a CNN, denoted as f, is optimized to compute a set of layer patterns from an input light field, which can reproduce the light field via the mapping Φ. The network is actually optimized over a large amount of data ($\mathcal{D}$), so as to minimize the error between the input (L^*_{train}) and reproduced ($\Phi(f(L^*_{\text{train}}))$) light fields.

$$f = \underset{f}{\operatorname{argmin}} \sum_{L^*_{\text{train}} \in \mathcal{D}} ||L^*_{\text{train}} - \Phi(f(L^*_{\text{train}}))||^2 \quad (2)$$

This training process takes several days depending on the network architecture and the amount of training dataset. Meanwhile, once the network has been trained, we can obtain a set of layer patterns for the target light field (L^*) via a single inference operation

$$\{P_z\} = f(L^*) \quad (3)$$

which is drastically faster than the iterative method.

3 Head-Tracking Layered Light Field Display

3.1 Overview

It is difficult for a layered light-field display to present many directional views simultaneously while maintaining the quality of each view. Therefore, we choose to reduce the number of views displayed for each time. More specifically, we incorporate head tracking into our display system, and we designed the display to show only a subset of views around the current head position. In other words, the display shows only a local light field instead of the entire light field at a time, which will ease keeping the image quality for each view. Moreover, our system can tolerate the errors and time delays of the head-tracking system to some extent, because the viewing angle supported by the display is not limited to a single direction, but covers a range of directions corresponding to the size of the local light field.

3.2 System

As depicted in Fig. 2, our system is composed of two PCs, a video camera (Logicool Carl Zeiss Tessar HD 1080p), and our prototype layered display [12, 13]. The camera is attached on the display. Head tracking is conducted on the first PC (client) equipped with Intel Core i5-4590 CPU. For simplicity of implementation, we regard a flashlight of a smartphone as the head position in the image, which will be replaced by actual head tracking with a more sophisticated tracker in the future. The estimated head position is translated into a viewpoint index as the target viewpoint, which is sent to the second PC (server) equipped with Intel Core i7-6700K (CPU) and NVIDIA GeForce GTX 1080 (GPU). The second PC is connected to the display and is responsible to provide the display with an appropriate set of layer patterns. For each time, the set of layer patterns should be adapted for a local light field that covers the viewing angle around the current head position.

3.3 Displaying a Local Light Field

The entire pipeline for displaying a local light field is illustrated in Fig. 3. Given the target viewpoint (u_t, v_t), we extracted a local light field centered on the target viewpoint, $L^*_{u_t,v_t}$, from the original light field L^*_{wide}. The viewpoint coordinate of the local light field is given with an offset (u_t, v_t) as

$$L^*_{u_t,v_t}(u, v, x, y) = L^*_{\text{wide}}(u + u_t, v + v_t, x, y) \quad (4)$$

The local light field is then fed to a pre-trained CNN to obtain a set of layer patterns as

$$\{P_{z,u_t,v_t}\} = f(L^*_{u_t,v_t}). \quad (5)$$

We can expect that the local light field generated from $\{P_{z,u_t,v_t}\}$ is of high quality because the number of viewpoints of the local light field is small. To align the central viewpoint to the display's configuration, we apply pixel shift to the set of layer patterns as

$$\bar{P}_{z,u_t,v_t}(x, y) = P_{z,u_t,v_t}(x - zu_t, y - zv_t) \quad (6)$$

We finally transfer the layer patterns $\{\bar{P}_{z,u_t,v_t}\}$ to the display hardware, which in turn outputs the local light field but

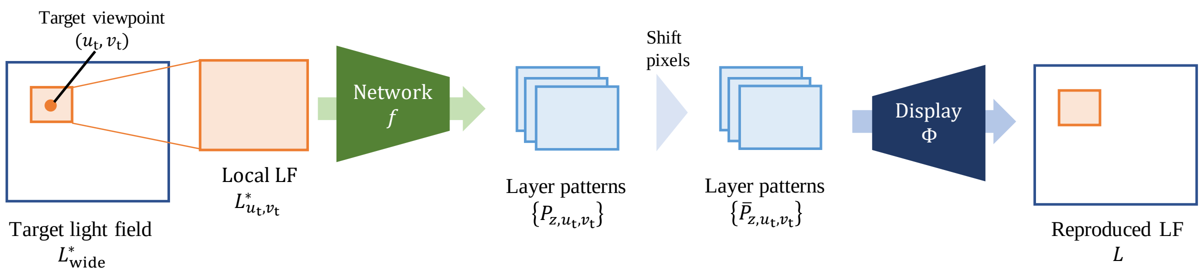

Fig. 3: Process pipeline for displaying a local light field

aligned with the coordinate of the original light field.

$$L = \Phi(\{\bar{P}_{z,u_t,v_t}\}) \quad (7)$$

At the point of this paper, the target light field was static. Therefore, we pre-computed the layer sets $\{\bar{P}_{z,u_t,v_t}\}$ for all the possible target viewpoints, which enabled real-time interaction with the head motion. When the layer patterns were computed online for a local light field with 5×5 views, the frame rate was only limited to 2.8 fps.

4 Experiments

We conducted a simulative experiment to compare our head-tracking-based method against a counterpart without head tracking. For our method, we trained a CNN that can accept 5×5 views to generate a corresponding set of layer patterns. The same network was used for any target viewpoint. Meanwhile, for the counterpart method, we trained a different CNN that was designed to convert 17×17 views into a set of layer patterns. The network architectures and training conditions are summarized in Table 1. We then prepared a target light field with 17×17 views for evaluation. The pre-traiend CNNs were used to obtain the layer patterns for the target light field, from which we computationally generated images for each viewpoint using Eq. (1). We used two target viewpoints $(u_t, v_t) = (5, 5)$ and $(-5, -5)$, and evalutaed the image quality for all of the 17×17 viewpoints.

Figure 4 (left) shows quantitative image quality in PSNR (top) and computationally generated images at $(u, v) = (-7, -7)$ and $(7, 7)$ (center and bottom). The counterpart method was designed to reconstruct the entire light field with 17×17 views, resulting in moderate image quality over all the viewpoints. Meanwhile, our method concentrated on a local light field with 5×5 views at a time, resulting in significantly better image quality around the target viewpoint. Although the image quality degraded as the viewpoint diverges from the target viewpoint, it was still moderate even outside the range of the local light field. Therefore, our method can tolerate some amount of inaccuracy of the target viewpoint, which may be caused by the error and time delay of the head tracking method.

We also tested our method on real hardware as shown in Fig. 4 (right). In this setup, the target light field had 45×45 viewpoints, for which a single set of layer patterns cannot achieve accurate reconstruction. However, thanks to the head-tracking-based scheme, our method can concentrate on displaying only 5×5 views at a time, which helped to maintain the image quality while allowing a significant degree of freedom for the head position.

Table. 1: Network architecture (top) and conditions (bottom)

layer	in	chns	act
input	L^*		
conv2D-1	input	$n^2/64$	ReLU
conv2D-1a	conv2D-1	64/64	ReLU
conv2D-1b	conv2D-1a	64/64	
Add-1	conv2D-1 + conv2D-1b		ReLU
conv2D-2a	Add-1	64/64	ReLU
conv2D-2b	conv2D-2a	64/64	
Add-2	Add-1 + conv2D-2b		ReLU
⋮	⋮		
conv2D-24a	Add-23	64/64	ReLU
conv2D-24b	conv2D-24a	64/64	
Add-24	Add-23 + conv2D-24b		ReLU
conv2D-L	Add-24	64/3	Hard Sigmoid
output	conv2D-L		

	Ours	Counterpart
Input viewpoints	$n = 5$	$n = 17$
Number of training samples	1,968,000	762,000
Number of training epochs	5	5
Loss function (MSE)	5^2 views	17^2 views

5 Conclusion

The number of viewpoints is an important factor for a light field display because it directly corresponds to the viewing angle in front of the display. To extend the viewing angle while keeping the image quality, we incorporate head-tracking into a layered light field display. Our display pipeline is designed to show only a local light field that will cover the viewing angle around the currently estimated head position. Experimental results were presented to show the

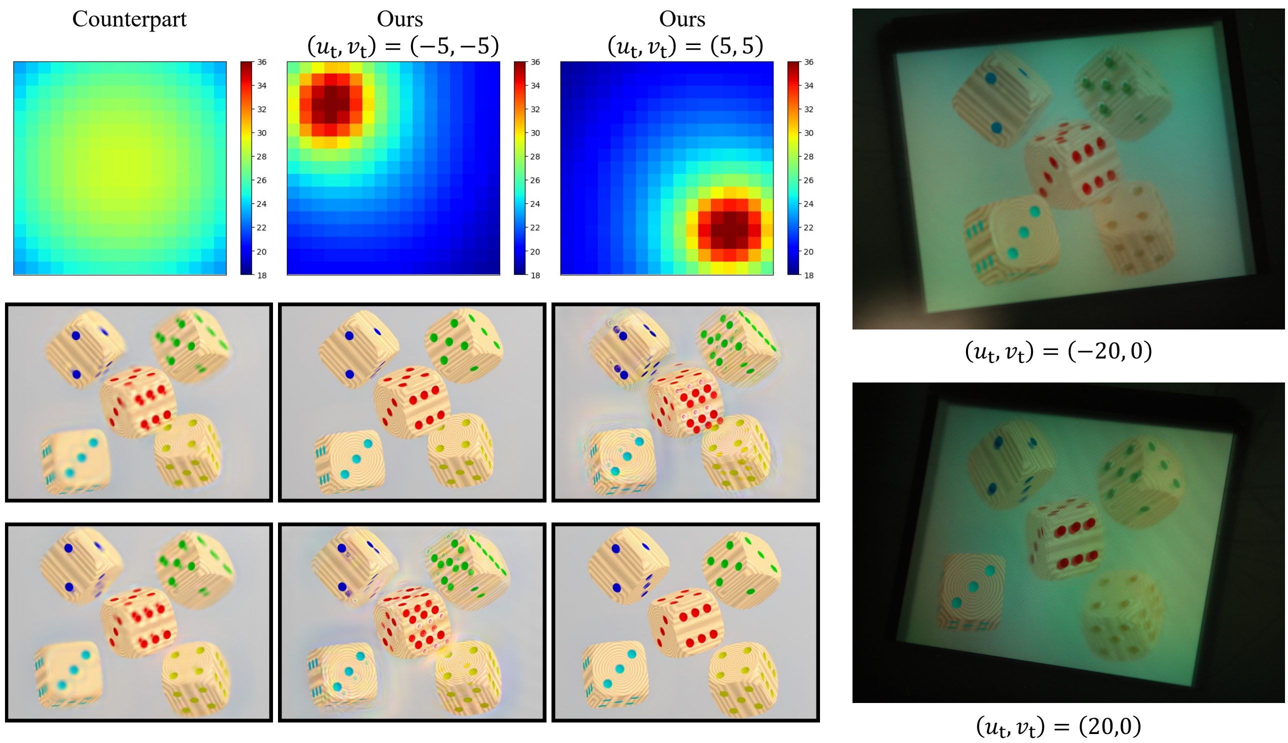

Fig. 4: Left: image quality for each viewpoint (top) and images at $(u, v) = (-7, -7)$ (center) and $(u, v) = (7, 7)$) (bottom) obtained by computer simulation. Right: images displayed on real hardware, where entire light field had 45×45 views.

feasibility and effectiveness of our proposal. In our future work, we will replace the head tracking method with a better one, and speed up the computation for the layer patterns to handle dynamic light fields online.

References

[1] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor Displays: Compressive Light Field Synthesis using Multi-layer Displays with Directional Backlighting," *ACM Trans. Graph. (Proc. SIGGRAPH)*, vol. 31, no. 4, pp. 1–11, 2012.

[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," in *ACM SIGGRAPH Asia 2010 papers*, 2010, pp. 1–10.

[3] 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.

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

[5] 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.

[6] G. Lippmann, "Epreuves reversibles donnant la sensation du relief," *J.de Phys*, vol. 7, pp. 821–825, 1908.

[7] 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.

[8] A. Schwartz, "Head tracking stereoscopic display," *IEEE Transactions on Electron Devices*, vol. 33, no. 8, pp. 1123–1127, 1986.

[9] P. Surman, S. Wang, X. Zhang, L. Zhang, and X. W. Sun, "Head tracked multiview 3D display," in *2015 Visual Communications and Image Processing (VCIP)*, 2015, pp. 1–1.

[10] T. Huang, B. Han, X. Zhang, and H. Liao, "High-performance autostereoscopic display based on the lenticular tracking method," *Optics express*, vol. 27, no. 15, pp. 20 421–20 434, 2019.

[11] 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.

[12] 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, 2017.

[13] 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.