FROM FOCAL STACKS TO TENSOR DISPLAY: A METHOD FOR LIGHT FIELD VISUALIZATION WITHOUT MULTI-VIEW IMAGES

Yuto KOBAYASHI, Keita TAKAHASHI, Toshiaki FUJII

Department of Electrical Engineering and Computer Science, Nagoya University, Japan
Furo-cho, Chikusa, Nagoya, 464-8603, Japan

ABSTRACT

A new type of light field display called a tensor display was investigated. Although this display consists of only a few light attenuating layers located in front of a backlight, many views can be emitted in different directions simultaneously without sacrificing the resolution of each view. The transmittance pattern of each layer is calculated from a light field, namely, a set of dense multi-view images (typically dozens) that are to be observed from different directions. However, preparing such images is often cumbersome for real objects. We propose a method that does not require multi-view images as the input; instead, a focal stack composed of only a few differently focused images is directly transformed into the layer patterns. Our method greatly reduces the data acquisition cost while also maintaining the quality of the output light field. We validated the method with experiments using synthetic light field datasets and a focal stack acquired by an ordinary camera.

***Index Terms*—** light field, focal stack, 3-D display

1. INTRODUCTION

3-D displays have been the subject of study for several years [1, 2, 3, 4, 5]. These displays can be categorized on the basis of several criteria, such as necessity of wearing glasses and the number of supported views. Glasses-free (naked-eye) displays have attracted attention because they enable a more natural viewing experience than glasses-based ones. Multi-view displays have more potentials than the conventional stereo-only displays because multi-view displays not only provide depth perception by showing different images to the left and right eyes but also present natural motion parallax along the movement of observers.

To develop glasses-free multi-view displays, researchers have devised several methods, including those that use parallax barriers [1, 6, 7, 8], specially designed lenses (lenticular screens or integral photography lenses) [2, 3, 9, 10, 11], and stacked layers [12, 13, 14, 15, 16]. In this paper, we focus on the final method, which is based on a few light-attenuating layers [13, 16]. This type of display, called a tensor display, can emit many views simultaneously in different directions without sacrificing the resolution of each view.

The structure of a typical tensor display is illustrated in Fig. 1 (right). A few light attenuating layers are stacked in front of a backlight. The transmittance of each layer pixel can be controlled individually in accordance with the contents to be displayed. Depending on the viewing direction, these layers overlap with a different shift, so the displayed images are direction dependent. More precisely, many views or a light field [17, 18], which are expected to be observed from different viewing directions, are given as the input, and then the layer patterns are optimized so as to reproduce the light field as faithfully as possible. This optimization is conducted through non-negative tensor factorization (NTF), where the transmittance values are alternately updated layer by layer.

This work was supported by JSPS Kakenhi Grant Number 15H05314

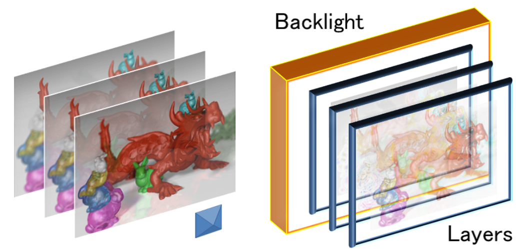


Fig. 1. Focal stack (left) and tensor display (right).

Visualizing real world 3-D scenes with this display presents a challenge in terms of data acquisition because a dense light field, i.e., a set of multi-view images (typically dozens of images) with very small viewpoint intervals, is required as the input [19]. In this paper, we demonstrate that a focal stack, which is composed of only a few differently focused images, as shown in Fig. 1 (left) , can be used as the input to this display instead of multi-view images. More specifically, if we use three semi-transparent layers for the display, we only need three images, where each image is focused on each layer. While greatly reducing the data acquisition cost, our method can also keep the quality of the output light field. We evaluated the effectiveness of the proposed method by conducting experiments using synthetic light field datasets and a focal stack acquired by an ordinary camera.

In previous works, focal stacks have been used for light field representation and depth estimation [20, 21, 22, 23]. However, to our knowledge, using a focal stack directly as the input to a light field display is a new and original contribution of this paper.

2. PROPOSED METHOD

2.1. Light field parameterization

A light field is defined as a 4-D function to describe all the light rays that travel straight in a free space [17, 18]. For simplicity of analysis, we only consider a 2-D flatland model with a 2-D light field (as shown in Fig. 2), but extension to a 3-D space (a 4-D light field) is straightforward. In this paper, we adopt a plane + angle parameterization: a reference plane ($z = 0$) is defined, and a light ray is parameterized by the intersection point with the reference plane (u)

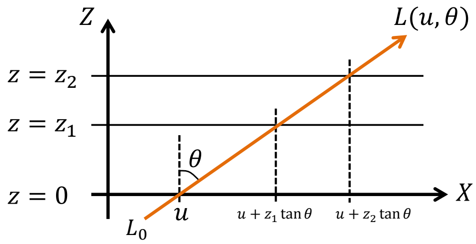


Fig. 2. Configuration of a light field.

and the direction with respect to the z axis (θ). The luminance of each light ray is described as $L(u, \theta)$. In accordance with this parameterization, the light rays that pass through a scene point (x, z) should satisfy

$$u = x - z \tan\theta. \tag{1}$$

2.2. Modeling a Tensor Display

As shown in Fig. 2, a few light attenuating layers are stacked in front of a backlight. A light ray originating from the point u on the reference plane and outgoing in the θ direction is described as

$$L(u, \theta) = \prod_{k \in \mathcal{K}} a_k \left(u + z_k \tan\theta\right) L_0. \tag{2}$$

Here, $a_k(x)$ denotes the transmittance of the k-th layer located at z_k and $\mathcal{K}$ denotes the set of layer indices. L_0 is the luminance of the backlight and can be omitted under the assumption that the light intensity is normalized.

As mentioned in [13, 16], using time-multiplex over several consecutive frames can improve the quality of the displayed light field. Time-multiplex is not considered in this paper, but will be explored in our future work.

2.3. Layer pattern optimization for a given light field

Given a light field that should be emitted from the display, $\bar{L}(u, \theta)$, the goal of layer pattern optimization is described as

$$\arg \min_{a_k(x)(k \in K)} \int_{\Theta} \int_{U} \left\| \bar{L}(u, \theta) - L(u, \theta) \right\|^2 du d\theta, \tag{3}$$

where U and Θ are the effective ranges for u and θ, respectively. This optimization is non-convex, so we adopted an alternative approach to solve it.

Suppose that we want to obtain the pattern for a specific layer $a_l(x)$ under the assumption that the other layer patterns $a_k(x)$ ($k \in \mathcal{K} \backslash \{l\}$) are known. From Eq. (2), we instantly obtain

$$\begin{aligned} L(x - z_l \tan\theta, \theta) &= \prod_{k \in \mathcal{K}} a_k \left(x + (z_k - z_l) \tan\theta\right) \\ &= a_l(x) A_l(x, \theta), \end{aligned} \tag{4}$$

where

$$A_l(x, \theta) = \prod_{k \in \mathcal{K} \backslash \{l\}} a_k \left(x + (z_k - z_l) \tan\theta\right). \tag{5}$$

Here, $L(x - z_l \tan\theta, \theta)$ represents all the outgoing light rays that depend on the transmittance of $a_l(x)$. It should be noted that $A_l(x, \theta)$ can be obtained without using $a_l(x)$. Therefore, under the assumption that $a_k(x)$ ($k \in \mathcal{K} \backslash \{l\}$) are fixed, Eq. (3) is equivalent to

$$\arg \min_{a_l(x)} \int_{\Theta} \left\| \bar{L}(x - z_l \tan\theta, \theta) - a_l(x) A_l(x, \theta) \right\|^2 d\theta \tag{6}$$

and we obtain the analytical solution that is written as

$$a_l(x) = \frac{\int_{\Theta} \bar{L}(x - z_l \tan\theta, \theta) A_l(x, \theta) d\theta}{\int_{\Theta} \left\| A_l(x, \theta) \right\|^2 d\theta}. \tag{7}$$

On the basis of the above, we can derive an algorithm (Algorithm 1) that optimizes the layer patterns for the given light field $\bar{L}(u, \theta)$. Although it looks different at first glance, this algorithm is completely the same as the multiplicative update rule used in [16] when time-multiplex is disabled.

Algorithm 1 Obtain layer patterns from a given light field

Input: $\bar{L}(u, \theta)$
Output: $a_k(x)$ ($k \in \mathcal{K}$)
Initialize $a_k(x)$ ($k \in \mathcal{K}$) with random numbers
Do until convergence
  For $k = 1, \ldots \|\mathcal{K}\|$
    update: $a_k(x) = \dfrac{\int_{\Theta} \bar{L}(x - z_k \tan\theta, \theta) A_k(x, \theta) d\theta}{\int_{\Theta} \left\| A_k(x, \theta) \right\|^2 d\theta}$
  End
End

2.4. Layer pattern optimization using a focal stack

Here, we propose a new method of layer pattern optimization using a focal stack. First, we pose an assumption that $A_l(x, \theta)$ given by Eq. (5) is smooth along θ within the limited effective range Θ. Accordingly, we rewrite $A_l(x, \theta)$ as

$$A_l(x, \theta) = A_l(x) + \epsilon_l(x, \theta) \tag{8}$$

where the amplitude of $\epsilon_l(x, \theta)$ is assumed to be sufficiently small with respect to that of $A_l(x)$. Given this assumption, Eq. (7) can be approximated as

$$a_l(x) \simeq \frac{\int_{\Theta} \bar{L}(x - z_l \tan\theta, \theta) d\theta}{\int_{\Theta} A_l(x, \theta) d\theta}. \tag{9}$$

Moreover, we rewrite the numerator of Eq. (7) as $I_l(x)$:

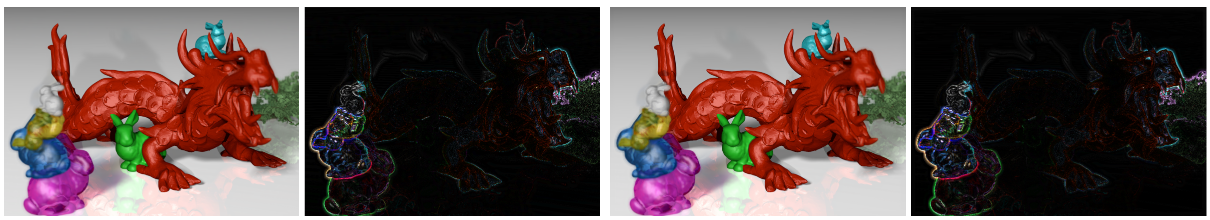
$$I_l(x) = \int_{\Theta} \bar{L}(x - z_l \tan\theta, \theta) d\theta. \tag{10}$$

Equation (10) indicates that the light rays originating from a scene point (x, z_l) and going into different directions θ (the right hand-side) converge on a single point x in $I_l(x)$ (the left hand-side). Therefore, $I_l(x)$ can be regarded as an image whose focus is set to the depth $z = z_l$. An important observation here is that the original light field $\bar{L}(u, \theta)$ is no longer necessary—only the image $I_l(x)$ is required to obtain the numerator of Eq. (7).

Accordingly, we can modify the previous algorithm into Algorithm 2, which requires a focal stack $I_k(x)$ ($k \in \mathcal{K}$) as the input instead of a light field $\bar{L}(u, \theta)$. This significantly reduces the cost of data acquisition, as a focal stack consists of only a few images (the same number as the display layers), while a light field typically consists of dozens of images. Moreover, Algorithm 2 requires less computational cost than Algorithm 1 for each iteration.

(a) Light field (5×5 views) and its central view (b) Focal stack with 3 depths generated from (a)

(c) Layer patterns obtained from the light field (a) (d) Layer patterns obtained from the focal stack (b)



(e) Simulated output using the layer patterns (c) and its error (f) Simulated output using the layer patterns (d) and its error

Fig. 3. Overview of the experiment using a synthetic light field dataset.

Table 1. Light field datasets from [24]

Dataset	Views	Resolution
DragonsAndBunnies Fishi Messerschmitt	5×5	840×593
Dice GreenDragon RedDragon	7×7	512×384

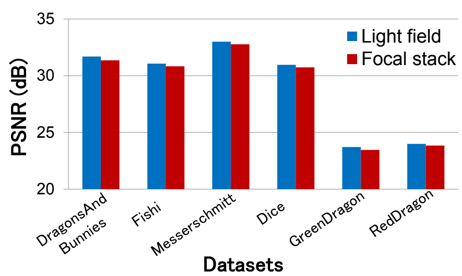


Fig. 4. Quality of output light fields.

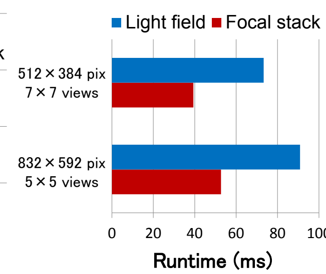


Fig. 5. Computational time.

Algorithm 2 Obtain layer pattern from a focal stack

Input: $I_k(x)$ $(k \in \mathcal{K})$
Output: $a_k(x)$ $(k \in \mathcal{K})$
Initialize $a_k(x)$ $(k \in \mathcal{K})$ with random numbers
Do until convergence
    For $k = 1, \dots \|\mathcal{K}\|$
        update: $a_k(x) = \dfrac{I_k(x)}{\int_\Theta A_k(x, \theta) d\theta}$
    End
End

3. EXPERIMENTS

It is obvious that our method using a focal stack (Algorithm 2) can significantly reduce the data acquisition cost compared to the conventional method using the original light field (Algorithm 1). In this section, we experimentally demonstrate that our method can achieve reasonable quality in reproducing the light field compared to the conventional method. In the experiments, the number of display layers was set to three.

The first experiment was conducted using light field datasets [24], the specifications of which are listed in Table 1. An overview of this experiment is shown in Fig. 3. As shown in Fig. 3(a), the original light field consists of 5×5 images. From this light field, we generated three refocused images, each of which is focused on each display's layer, as seen in Fig. 3(b). Shown in Figs. 3(c) and 3(d) are the optimized layer patterns obtained by the conventional and proposed methods, respectively. The former was calculated directly

(a) Focal stack obtained from a camera

(b) Layer patterns

(c) Simulated outputs

Fig. 6. Experiment with a real scene (refer also to demo video).

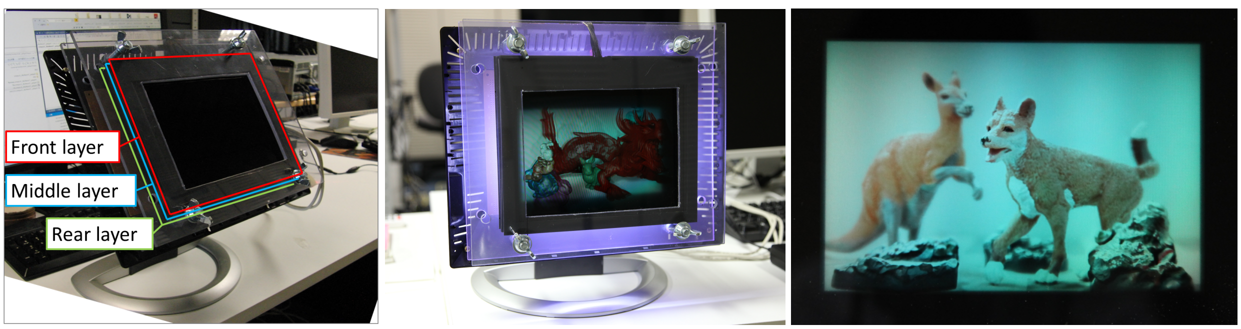


Fig. 7. Our prototype display.

from the original light field in Fig. 3(a), while the latter was from the focal stack in Fig. 3(b). Both of them look similar. Shown in Figs. 3(e) and 3(f) are the simulated images that would be observed when three stacked layers are seen from the central viewpoint. Their errors from the ground truth are also presented (magnified by 4 for visualization). Again, similar results are obtained by the conventional (Fig. 3(e)) and proposed methods (Fig. 3(f)).

A quantitative evaluation of the reproduced light fields over six datasets is presented in Fig. 4. Here, the stacked layers were observed by simulation from the same viewpoints as those of the original light field dataset and the reproduction quality was measured by PSNR against the original light field. Our method (focal stack) achieves comparable quality to the conventional method (light field). Figure 5 shows the average computational times for each iteration with the conventional method (light field) and our method (focal stack). We used a PC that has an Intel Core i7 CPU with 8GB of RAM and a NVIDIA Geforce GT 730 video card. Both methods were parallelized over x using CUDA. The computational time was reduced to 50 – 60 % by our method.

Finally, we demonstrate light field visualization of a real scene by using a focal stack acquired with an ordinary camera (Canon EOS 5D Mark II). Three images focused on different depths are presented in Fig. 6(a). From this focal stack, we generated the layer patterns shown in Fig. 6(b). Figure 6(c) shows simulated output images that would be observed from different directions. We also displayed these layer patterns on our prototype display [25], as shown in Fig. 7, and confirmed that natural 3D perception was obtained. Note again that we used only three focused images to achieve this; no light field was captured at all.

More results are presented in the supplemental video [26].

4. CONCLUSION

A method for light field visualization on a tensor display that uses a focal stack instead of multi-view images as the input was proposed. Our method greatly reduces the data acquisition cost compared to the conventional method while maintaining the quality of the output light field. In future work, we will develop an end-to-end system where a real scene captured as a focal stack can be reproduced in 3D on our display hardware in real-time.

5. REFERENCES

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

[2] G. Lippmann, "Epreuves reversibles donnant la sensation du relief," *Journal of Physics*, vol. 7, no. 4, pp. 821–825, 1908.

[3] T. Okoshi, *Three-Dimensional Imaging Techniques*, Academic Press, 1976.

[4] S. Pastoor and M. Wopking, "3-D displays: A review of current technologies," *Elsevier Displays*, vol. 17, no. 2, pp. 100–110, 1997.

[5] B. Javidi and F. Okano, Eds., *Three-dimensional television, video, and display technologies*, Springer, 2002.

[6] H. Isono, M. Yasuda, and H Sasazawa, "Autostereoscopic 3-D LCD display using LCD-generated parallax barrier," *Japan Display*, pp. 303–306, 1992.

[7] K. Sakamoto and T. Morii, "Multiview 3D display using parallax barrier combined with polarizer," *Proc. SPIE*, vol. 6399, 2006.

[8] 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 Trans. on Visualization and Computer Graphics*, vol. 14, no. 3, pp. 487–499, 2008.

[9] R. Borner, "Autostereoscopic 3D-imaging by front and rear projection and on flat panel displays," *Elsevier Displays*, vol. 14, no. 1, pp. 39–46, 1993.

[10] M. McCormick, "Integral 3D imaging for broadcast," *Int. Display Workshop*, vol. 14, no. 1, pp. 39–46, 1993.

[11] J. Arai, F. Okano, M. Kawakita, M. Okui, Y. Hanio, M. Yoshimura, M. Furuya, and M. Sato, "Integral three dimensional television using a 33-megapixel imaging system," *Journal of Display Technology*, vol. 6, no. 10, pp. 422–430, 2010.

[12] 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.

[13] 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 Transactions on Graphics*, vol. 29, no. 6, 2010.

[14] D. Lanman, G. Wetzstein, M. Hirsch, W. Heidrich, and R. Raskar, "Polarization fields: dynamic light field display using multi-layer LCDs," *ACM Transactions on Graphics*, vol. 30, no. 6, 2011.

[15] D. Lanman, G. Wetzstein, M. Hirsch, W. Heidrich, and R. Raskar, "Beyond parallax barriers: applying formal optimization methods to multilayer automultiscopic displays," *Proc. SPIE*, vol. 8288, 2012.

[16] 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, no. 4, 2012.

[17] M. Levoy and P. Hanrahan, "Light field rendering," *ACM SIGGRAPH*, pp. 31–42, 1996.

[18] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," *ACM SIGGRAPH*, pp. 43–54, 1996.

[19] 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," *accepted to OSA/IEEE Journal of Display Technology*, 2016.

[20] Kiriakos N. Kutulakos and Samuel W. Hasinoff, "Focal stack photography: High-performance photography with a conventional camera," *In Proc. IAPR Conference on Machine Vision*, pp. 332–337, 2009.

[21] A. Levin and F. Durand, "Linear view synthesis using a dimensionality gap light field prior," *Computer Vision and Pattern Recognition (CVPR)*, pp. 1831–1838, 2010.

[22] K. Kodama and A. Kubota, "Efficient reconstruction of all-in-focus images through shifted pinholes from multi-focus images for dense light field synthesis and rendering," *IEEE Transactions on Image Processing*, vol. 22, no. 11, 2013.

[23] X. Lin, J. Suo, G. Wetzstein, Q. Dai, and R. Raskar, "Coded focal stack photography," *Computational Photography (ICCP), 2013 IEEE International Conference on. IEEE*, pp. 1–9, 2013.

[24] "Synthetic Light Field Archive," https://www.media.mit.edu/∼gordonw/SyntheticLightFields/.

[25] K. Takahashi, Y. Kobayashi, and T. Fujii, "Displaying real world light fields using stacked LCDs," *to appear in International Display Workshops*, 2016.

[26] "supplemental video of this paper," http://www.fujii.nuee.nagoya-u.ac.jp/∼takahasi/Research/LFDisplay2/.