Physically-Correct Light-Field Factorization for Perspective Images

Shu KONDO^{†a)}, Nonmember, Yuto KOBAYASHI^{†b)}, Student Member, Keita TAKAHASHI^{†c)},
and Toshiaki FUJII^{†d)}, Members

SUMMARY A layered light-field display based on light-field factorization is considered. In the original work, the factorization is formulated under the assumption that the light field is captured with orthographic cameras. In this paper, we introduce a generalized framework for light-field factorization that can handle both the orthographic and perspective camera projection models. With our framework, a light field captured with perspective cameras can be displayed accurately.

key words: 3-D display, light field, multi-view images

1. Introduction

Autostereoscopic 3-D displays, which can be observed with naked eyes, have attracted attentions recently. Especially, we focus on light-field displays that can simultaneously emit multi-view images in different directions [1]–[5]. One promising method to implement such a display is to stack semi-transparent layers (e.g., LCD panels) in front of a backlight [3], [5] as illustrated in Fig. 1. We can observe different images depending on the view direction because the layers overlap with a different shift. To produce direction-dependent images that correspond to the target 3-D object, the transmittance patterns of layers are determined from a given multi-view images through light-field factorization.

In the original work [5], the light field is represented in a tensor form that is equivalent to a set of multi-view images captured with an orthographic camera model. Therefore, the framework of light-field factorization also implicitly assumes the orthographic camera model. In our previous work [6], [7], we took an input light field from a real 3-D object, where the images were captured with perspective cameras. However, the images were regarded as the ones with a orthographic camera model when they were associated with the display. Although the displayed results were compelling, the discrepancy in camera projection models caused shape distortions on the displayed object.

To fix this problem, we introduce a generalized framework for light-field factorization that can handle both the orthographic and perspective camera projection models. With

Fig. 1 (a) Layered light field display. (b) Pattern for each layer. (c) Appearances from different directions.

our framework, the images captured with perspective cameras can be associated with the display in a physically correct manner, resulting in accurate displayed outputs.

2. Light Field Factorization with Perspective Cameras

In this paper, the light field is parameterized as a set of multi-view images captured by a 2-D array of cameras. Each light ray is described as $L(x, y, i, j)$, where (i, j) denotes the camera's position in the 2-D array, and (x, y) denotes the pixel position. Each of the image can also be described as $I_{i,j}(x, y)$ individually. The camera's projection model can be either perspective or orthographic. We assume that the layered display has three semi-transparent layers, but our discussion can be extended to any number of layers.

2.1 Generalized Framework of Light Field Factorization

First, we introduce a generalized framework of light field factorization which can handle either a perspective or an orthographic camera model. When an orthographic projection model is adopted, this framework becomes completely the same as the nonnegative tensor factorization formulated in the original paper [5]. Note that our formulation looks very different from that of original paper [5]; our formulation is described in pixel-wise operations while matrix-tensor operations were adopted in [5].

The light field generated by the display is decomposed into the light rays that are outgoing from the display. A light ray among them, $\hat{L}(x, y, i, j)$, is depicted in Fig. 2. This light ray is synthesized as

$$\hat{L}(x, y, i, j) = L \frac{\sum_{t=1}^T \prod_{m=-1}^1 V_m^t(\mathcal{M}^{i,j \rightarrow V_m}(x, y))}{T}. \quad (1)$$

Manuscript received December 16, 2016.

Manuscript revised March 16, 2017.

Manuscript publicized June 14, 2017.

[†]The authors are with the Graduate School of Engineering, Nagoya University, Nagoya-shi, 464–8603 Japan.

a) E-mail: kondo@fujii.nuee.nagoya-u.ac.jp

b) E-mail: kobayashi@fujii.nuee.nagoya-u.ac.jp

c) E-mail: keita.takahashi@nagoya-u.jp

d) E-mail: fujii@nuee.nagoya-u.ac.jp

DOI: 10.1587/transinf.2016PCL0006

Fig. 2 Definition of ray-space.

Here, $V_m^t(x, y)$ denotes the transmittance of the m -th layer at time t , where $m = -1, 0, 1$ are assigned to the rear, middle, and front layers, respectively. Symbol L is the luminance of a uniform backlight. Symbol T denotes the degree of time multiplexing; T layer patterns are repeated rapidly so that the viewer perceive the average over T patterns. Operator $\mathcal{M}^{I_{i,j} \rightarrow V_m}(x, y)$ maps the coordinate (x, y) on image $I_{i,j}$ to the coordinate (u, v) on layer V_m along the path of light ray $\hat{L}(x, y, i, j)$.

To obtain the layer patterns that can approximate the given light field $L(x, y, i, j)$, we consider the optimization problem given by:

$$\{V_{-1}, V_0, V_1\} = \arg \min \sum_{x, y, i, j} \left| I_{i,j}(x, y) - \sum_{t=1}^T \prod_{m=-1}^1 \frac{V_m^t(\mathcal{M}^{I_{i,j} \rightarrow V_m}(x, y))}{T} \right|^2. \quad (2)$$

The layer patterns are iteratively optimized by repeating the alternative update for $V_m^t(u, v)$ as follows:

$$V_m^t(u, v) \leftarrow h \left(V_m^t(u, v) \frac{f_m(u, v)}{g_m(u, v)} \right), \quad (3)$$

where

$$f_m(u, v) = T \sum_{i, j} I_{i,j}(\mathcal{M}^{V_m \rightarrow I_{i,j}}(u, v)) \prod_{l \neq m} V_l^t(\mathcal{M}^{V_m \rightarrow I_{i,j} \rightarrow V_l}(u, v)), \quad (4)$$

$$g_m(u, v) = \sum_{i, j} \left(\sum_t V_m^t(u, v) \prod_{l \neq m} V_l^t(\mathcal{M}^{V_m \rightarrow I_{i,j} \rightarrow V_l}(u, v)) \right) \prod_{l \neq m} V_l^t(\mathcal{M}^{V_m \rightarrow I_{i,j} \rightarrow V_l}(u, v)), \quad (5)$$

$$h(x) = \begin{cases} \epsilon & (x < \epsilon) \\ x & (\epsilon \leq x \leq 1) \\ 1 & (1 < x). \end{cases} \quad (6)$$

Here, $\mathcal{M}^{V_m \rightarrow I_{i,j}}(u, v)$ is the projective mapping from layer V_m to image $I_{i,j}$. Operator $\mathcal{M}^{V_m \rightarrow I_{i,j} \rightarrow V_l}(u, v)$ denote the mapping

Fig. 3 Orthographic camera model (left) and perspective camera model (Right).

from layer V_m to layer V_l , along the path of the light ray that is captured in image $I_{i,j}$.

2.2 Handling Perspective Cameras

Figure 3 (left) illustrates a case when a orthographic projection model is adopted for the input light field. The mapping operators are described as

$$\mathcal{M}^{V_m \rightarrow I_{i,j}}(u, v) = (u + mi, v + mj), \quad (7)$$

$$\mathcal{M}^{V_m \rightarrow I_{i,j} \rightarrow V_l}(u, v) = (u + (m - l)i, v + (m - l)j). \quad (8)$$

Obviously, an integer pixel position (u, v) on layer V_m is always mapped onto an integer pixel position on image $I_{i,j}$ or layer V_l . In this case, the update process given by Eqs. (2)–(6) is completely the same as the multiplicative update rule of non-negative tensor factorization [5][†].

Meanwhile, when a perspective projection model is adopted as shown in Fig. 3 (right), the mapping operators are determined by the geometric relations among the layers and cameras as

$$\mathcal{M}^{V_m \rightarrow I_{i,j}}(u, v) = \left(\frac{f(u - p_i)}{D - md}, \frac{f(v - q_i)}{D - md} \right), \quad (9)$$

$$\mathcal{M}^{V_m \rightarrow I_{i,j} \rightarrow V_l}(u, v) = \left(\frac{(D - ld)(u - p_i)}{D - md}, \frac{(D - ld)(v - q_i)}{D - md} \right), \quad (10)$$

where D is the distance between the cameras and the central layer, and d is the distance among the layers. Symbol f and (p_i, q_i) denotes the focal length and the viewpoint of the image $I_{i,j}$. In this case, an integer pixel position (u, v) on layer V_m is mapped onto a fractional pixel position on image $I_{i,j}$ or layer V_l . Therefore, we use linear interpolation to obtain pixel values at fractional positions. In this case, the update process given by Eqs. (2)–(6) is no longer interpreted as non-negative tensor factorization, but can be handled similarly to the case with an orthographic projection model only with modifications on the mapping operators.

3. Experimental Result

To verify our method, we used an input light field captured

[†]To confirm this, rewrite the matrix-tensor operations in [5] as element-wise operations.

Fig. 4 (a) Captured space, (b) multi-view images, and (c) close-ups of the captured and reconstructed multi-view images.

Fig. 5 Absolute errors (magnified by 4) between the captured and reconstructed multi-view images with the orthographic camera model (left) and the perspective camera model (right).

with perspective cameras, and evaluated the reconstructed light field on the computer-simulated light field display. We compared two light-field factorization models: the ones with orthographic and perspective camera models.

First, we shot a set of multi-view images using a 7×7 camera array as shown in Fig. 4 (a)(b). The distance from the camera array to an object was about 800, the angle of views was 10° , the intervals among the cameras were 2, and each image had 1100×1100 pixels. Second, we factorized the acquired light field into semi-transparent layer patterns assuming either the orthographic or perspective camera model. For both of the models, the degree of time multiplexing was set to $T = 2$ and the number of update was 5. Finally, using the layer patterns, we simulated the appearance of the display and observed it from the original viewpoints of the captured images. The reconstruction accuracy was evaluated as the difference between the acquired and observed multi-view images. Figure 4 (c) shows several parts of the captured and reconstructed images. Figure 5 shows the reconstruction errors from the captured images.

As shown Figs. 4 and 5, using the perspective camera model in light field factorization leads to better reconstruction quality than using the orthographic camera model. The

Table 1 Reconstruction quality with different number of layers (left) and different distance from the camera array to the object (right).

N	PSNR (dB)	D	FOV ($^\circ$)	PSNR (dB)
2	24.81	600	13.3	24.15
3	25.33	800	10.0	25.33
5	25.54	1200	6.7	23.36

PSNR over the entire multi-view images was 17.46 dB for the orthographic camera model and 25.33 dB for the perspective camera model. This result seems to be trivial because the original light field was captured with perspective cameras. However, it is thanks to our framework that the images captured with perspective cameras can be handled in a physically correct manner.

Furthermore, we show that our method works for a wide range of configurations. We changed the number of layers N and the distance between the object and the camera array D from the original configuration of Fig. 4. The results are summarized in Table 1. When we changed the number of layers, we kept the nearest and farthest layers fixed, and placed the other layers with constant intervals. As the number of layers increased, the reconstruction quality also improved at the cost of the increased computational complexity, which is $O(N^2)$. When we changed the distance D , we also changed the viewing angle (FOV) of each camera to keep the object size almost constant in the image. For all cases, our method can reconstruct multi-view images with reasonable quality.

4. Conclusion

We have introduced a generalized framework for light-field factorization that can handle both the orthographic and perspective camera projection models simply by changing the mapping operators. With our framework, images captured with perspective cameras can be associated with the display in a physically correct manner, which was impossible with the previous works. Experimental results have shown that

multi-view images captured with perspective cameras were displayed accurately thanks to our framework.

Acknowledgments

This work was supported by JSPS Kakenhi Grant Number 15H05314.

References

- [1] 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. Vis. Comput. Graphics*, vol.14, no.3, pp.487–499, 2008.
 - [2] J. Arai, F. Okano, M. Kawakita, M. Okui, Y. Haino, 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.
 - [3] 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 Trans. Graphics*, vol.30, no.4, article 95, 2011.
 - [4] D. Lanman, G. Wetzstein, M. Hirsch, W. Heidrich, and R. Raskar, "Polarization fields: Dynamic light field display using multi-layer LCDs," *ACM Trans. Graphics*, vol.30, no.6, article 186, 2011.
 - [5] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor displays: Compressive light field synthesis using multilayer displays with directional backlighting," *ACM Trans. Graphics*, vol.31, no.4, article 80, 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, no.11, pp.1290–1300, 2016.
 - [7] K. Takahashi, Y. Kobayashi, and T. Fujii, "Displaying real world light fields using stacked LCDs," *International Display Workshops in Conjunction with Asia Display, DES4/3D8-1*, pp.1300–1303, 2016.
-