Displaying Real-World Light-Fields with Stacked Multiplicative Layers: Requirement and Data Conversion for Input Multi-view Images

Toyohiro Saito, Yuto Kobayashi, Keita Takahashi, *Member, IEEE,* and Toshiaki Fujii, *Member, IEEE*

***Abstract*—To provide realistic 3-D perception to human observers, 3-D displays have evolved to present not only a pair of stereo images but also many images to different viewing directions. A light field display, which consists of a few light attenuating pixelized layers (e.g., LCD panels) stacked in front of a backlight, has attracted attention because of its potential to simultaneously support many viewing directions with high quality. The transmittances of the layers are determined from a set of multi-view images or a light field that is given to represent observations expected from many viewing directions. However, the relation between the configuration of the given light field and the quality of displayed images has not sufficiently been discussed in previous works. In this study, in our aim to display real world objects with high quality, we address the requirement for the target light field given as the input. We deeply analyze several factors that associate the configuration of the target light field with the quality of the displayed images, and we derive a quantitative requirement on the configuration: the disparities among the adjacent viewpoints should be limited to 0 to 1 pixels. To meet this strict requirement with real world objects, we propose using a multi-view camera and image based rendering, where we can generate virtual light fields with arbitrary configurations and densities. Our theory and method are verified by experiments using a computer simulated display.**

***Index Terms*—3-D display, light field, multi-view images, image based rendering.**

I. INTRODUCTION

D-EPTH perception is a fundamental factor for us to understand 3-dimensional (3-D) visual information. Over a century, the human mechanism of depth perception and its applications to 3-D display technologies have been studied [1], [2], [3], [4], [5]. These 3-D displays are roughly categorized into glasses-based and naked-eye based (auto-stereoscopic) displays. Although having many technical challenges, displays in the latter category are more attractive without the need for wearing special glasses. Another categorization is to divide them into stereoscopic displays and multi-view (light field) displays. The stereoscopic displays are designed only to present stereo images to the left and right eyes, individually. To provide more natural depth perception with motion parallax, not only a pair of stereo images but also many images for different viewing directions should be displayed simultaneously.

T. Saito, K. Takahashi, and T. Fujii are with Department of Electric Engineering and Computer Science, Graduate School of Engineering, Nagoya University, Japan, e-mail: tsaitou@fujii.nuee.nagoya-u.ac.jp, keita.takahashi@nagoya-u.jp, fujii@nuee.nagoya-u.ac.jp.

Manuscript received January 26, 2016

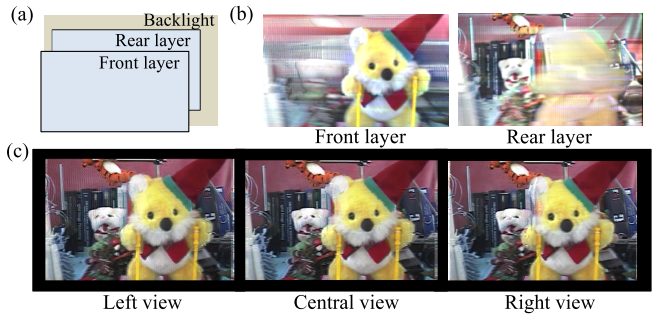


Fig. 1. (a) A light field display with multiplicative layers. (b) Transmittance for each layer. (c) Appearances from different directions.

Among those categories, this study is focused on glasses-free light field displays.

Glasses-free light field displays can be implemented in several ways. The first approach is to use a screen and many projectors [6], [7], [8], [9], [10], [11]. However, this approach is costly and requires a large space because many projectors should be placed in front of or behind the screen at a large distance. The second approach is to attach a parallax-barrier [12], [13], [14], [15] or special lens (lenticular lens or integral photography lens) [16], [17], [18], [19] to a single display panel to control the directions of emitted light rays. Although this approach is widely commercialized thanks to its simplicity, it has a fundamental limitation in terms of resolution because many images share the same display panel with a finite number of pixels. A trade-off thus exists: as the number of simultaneously supported images increase, the resolution per each image decreases accordingly. The third approach is to stack pixelized layers to produce direction-dependent outputs [20], [21], [22], [23], [24], [25], [26], [27], [28]. In [20], [22], the luminance of each layer pixel is determined by the depth of the displayed object. Meanwhile, in [21], [23], [24], [25], [26], [27], [28], the layer pixels are optimized so as to produce the target light field as faithfully as possible, which is a direct approach to our purpose. Particularly, it has been reported [23], [26], [27] that multiplicative layer displays have the potential to reproduce dense light fields with high quality.

The typical structure of a multiplicative layer display is illustrated in fig. 1. It consists of a few light-attenuating pixelized layers (e.g., LCD panels) stacked in front of a backlight. Depending on the viewing direction, the layers

overlap with a different shift, so that the displayed images are direction dependent. Therefore, the display can provide different images to different viewing directions simultaneously, and thus a light field is presented by the display. The transmittance patterns of layers are determined from a given target light field, which is equivalent to multi-view images describing what should be seen from each viewing direction. Specifically, the layer patterns are optimized in a manner of an inverse problem to approximately reproduce the target light field. This approximation accuracy, i.e., the quality of displayed images, heavily depends on the configuration of the target light field. However, this important dependency was not sufficiently discussed in previous works.

In this paper, we address what the requirements are for the configuration of the target light field to display real world objects with high quality. To be more precise, with what density should the target light field be provided? How should the light field correspond to the viewing directions of the display? To answer these questions, we present three analyses based on theory and experiments: (i) the relation between the disparities in the given light field and the depth where the objects are displayed, (ii) the upper-bound spatial-frequency of the display, which is shown to be depth-dependent, and (iii) the relation between the quality of the displayed images and the interval with which the given light field corresponds to the viewing directions of the display. As a result, we derive a quantitative requirement on the configuration: the disparities among the adjacent viewpoints should be limited to 0 to 1 pixels. To meet this very strict condition with real world objects, we propose using a multi-view camera and image based rendering, where we can generate virtual multi-view images with arbitrary density. Although the present study is currently limited to a case with two multiplicative layers, we believe it will trigger new research that will lead to a more realistic visualization of real world 3-D objects.

Before starting detailed discussions, we briefly review some related works. Spatial frequency analyses of multiplicative layered displays have been presented in [24], [27]. Although our analysis in (ii) derives an equivalent result to [24], [27], ours is much simpler and easier to be associated with the configuration of the target light field. Our analysis in (iii) seems to have some relation to sampling theories on light fields [29], [30], [31] that apply to sampling and interpolation in image based rendering. However, those theories cannot apply to our problem where we cannot control the interpolation among the viewing directions because it is performed internally during the process of layer pattern optimization. A multi-view camera and image based rendering have also been used to generate the input to a 3-D display in [32]. However, their target was an integral photography display, which has a different configuration and different frequency characteristics from a multiplicative layered display. Finally, the present paper is extended from our conference papers [33], [34].

II. LAYERED LIGHT FIELD DISPLAY

In this section, we first mention the model of a light field display consisting of two stacked multiplicative layers and then describe how to obtain layer patterns to generate the desired images for various viewing directions. Note that what are described here are not our original contribution but have been presented in [23], [25], [26]. The description is limited to a 2-D space for simplicity of analysis, but it is straightforwardly extended to an original 3-D space.

A top view of a light field display with two layers is illustrated in fig. 2. A backlight and two light-attenuating layers are placed in parallel to the X axis on different depths Z. Two coordinate axes, S and U, are defined on the rear and front layers, respectively, whose depths are written as $Z = z_s$ and $Z = z_u$. A light ray originating from the backlight intersects with the rear and front layers at $S = s$ and $U = u$, respectively, reducing its luminance in accordance with the transmittances of the layers at the intersection points, which are denoted by $a_t(s)$ and $b_t(u)$ $(0 \le a_t(s), b_t(u) \le 1)$ as functions of time t. The light ray emitted from the display $\tilde{l}(s,u)$ can be represented as

$$\tilde{l}(s,u) = \frac{1}{T}\sum_{t=1}^{T} a_t(s) b_t(u), \tag{1}$$

where T is the degree of time multiplexing. We assume here that the T layer patterns can be repeated rapidly so that the average over T patterns is perceived by human eyes. Note here that the coordinate system of the light field (s,u) already includes the layer interval implicitly, and the layer interval is fixed throughout the paper. Therefore, the layer interval does not appear explicitly in the formulations of this paper.

Next, we introduce a discrete representation of eq. (1), which is suitable for computation. Since the layers have discrete pixels, the coordinates s and u are also quantized with a unit length 1. We assume that these layers have same number of pixels denoted by P, and define $\mathbf{a}_t$ and $\mathbf{b}_t$ as the column vectors given by $\mathbf{a}_t = [a_t(1)\ a_t(2)\ \cdots\ a_t(P)]^{\mathrm{T}}$ and $\mathbf{b}_t = [b_t(1)\ b_t(2)\ \cdots\ b_t(P)]^{\mathrm{T}}$, respectively. Furthermore, we introduce matrices $\mathbf{A}$ and $\mathbf{B}$ by gathering $\mathbf{a}_t$ and $\mathbf{b}_t$ for all t as follows:

$$\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_T],\ \mathbf{B} = [\mathbf{b}_1\ \mathbf{b}_2\ \cdots\ \mathbf{b}_T]. \tag{2}$$

The emitted light field can also be represented as a matrix $\tilde{\mathbf{L}}$ where $\{\tilde{L}_{i,j} = \tilde{l}(i,j) | i,j \in \mathbb{N},\ i,j \le P\}$. Using these vector and matrix representations, eq. (1) is equivalent to

$$\tilde{\mathbf{L}} = \frac{1}{T}\mathbf{A}\mathbf{B}^{\mathrm{T}}. \tag{3}$$

Finally, we describe how to obtain layer patterns that generate the desired images for various viewing directions. The desired images are represented in a form of light field, $l(s,u)$ or $\mathbf{L}$, where each desired luminance is assigned to the corresponding element of $\mathbf{L}$, but not all of the elements have the assigned values. The light field emitted from the display, given by eq. (3), should be close to $\mathbf{L}$ where $\mathbf{L}$ has assigned values. The optimal layers patterns are obtained by solving a least-square problem

$$\arg\min_{\mathbf{A},\mathbf{B}} \left\| \mathbf{W} \circledast \mathbf{L} - \mathbf{W} \circledast \frac{1}{T}\mathbf{A}\mathbf{B}^{\mathrm{T}} \right\|^2,\ for\ 0 \le \mathbf{A}, \mathbf{B} \le 1, \tag{4}$$

where $\mathbf{W}$ is a binary matrix that takes 0 where the corresponding element was not given the input, and $\circledast$ is an element-wise product operator. This equation means that the target light field $\mathbf{L}$ is approximated by the product of lower rank matrices $\mathbf{A}$ and $\mathbf{B}$, in a form of a low-rank approximation or a low-rank compression, where the rank is determined by the degree of the time multiplexing T. To obtain a solution for eq. (4), non-negative matrix factorization via multiplicative update rules [35], [36] can be used. Initialized with random patterns,

$$\begin{aligned}\mathbf{A} &\leftarrow \mathbf{A} \circledast ((\mathbf{W} \circledast T\mathbf{L})\mathbf{B}) \oslash ((\mathbf{W} \circledast (\mathbf{A}\mathbf{B}^{\mathrm{T}}))\mathbf{B}) \\ \mathbf{B} &\leftarrow \mathbf{B} \circledast ((\mathbf{W} \circledast T\mathbf{L})^{\mathrm{T}}\mathbf{A}) \oslash ((\mathbf{W} \circledast (\mathbf{A}\mathbf{B}^{\mathrm{T}}))^{\mathrm{T}}\mathbf{A})\end{aligned} \tag{5}$$

are iterated until convergence. Operator $\oslash$ denotes element-wise division.

Development of real display hardware involves many technical challenges and physical limitations, which are not essential for our theoretical analysis. Therefore, all of our experiments presented in this paper were performed as computer simulations. The degree of time multiplexing T was set to 3. Note that the above formulation can be naturally extended to more than two layers [27], [36] and projection based displays [37], but these extensions are beyond the scope of this paper.

III. REQUIREMENT FOR INPUT MULTI-VIEW IMAGES

As described above, the layer patterns constituting the display are optimized by giving the target light field $l(s,u)$. The target light field is typically represented as a set of multi-view images from several viewing directions, which are taken with a constant interval and rectified with each other. A fundamental question here is what configuration is required for the input multi-view images. To answer this question, first we analyze the relation between the disparities in the multi-view images and the depth where the objects are displayed. Second, we investigate the upper-bound spatial-frequency of the display, which is shown to be depth-dependent, and derive a depth limitation on the display. Third, we investigate the relation between the quality of the displayed images and the interval with which the given light field corresponds to the viewing directions of the display. Finally, by combining these analyses, we derive a quantitative requirement on the configuration of the multi-view images: the disparities among the adjacent viewpoints should be limited to 0 to 1 pixels, which is needed to display target 3-D objects clearly throughout the depth.

Before starting detailed discussions, we first clarify the target and limitation of our analysis. To summarize, the quality of the displayed 3-D content is restricted by the three factors: (F1) the discrete representation of the light field due to the non-infinitesimal pixels of the display, (F2) insufficient sample density of the given light field, and (F3) the low-rank approximation (compression) imposed by the layer decomposition. The present paper only considers (F1) and (F2), and thus, the requirement we will derive here is only a necessary condition but not a sufficient condition for high quality 3-D visualization. More specifically, the factor (F1) solely determines the upper-bound spatial frequency described in Section III.B, while the factor (F2) is additionally considered in the analysis of Section III.C. The factor (F3) was analyzed in one of our previous papers [34], but the rank analysis of a light field is a difficult issue by itself, and thus, we would like to keep it beyond the scope of this paper. In a sense, our analysis in this paper reveals the upper-bound quality that is theoretically achievable with this display, because the factor (F3) can be diminished theoretically by increasing the degree of time multiplexing. However, in practice, the quality degradation of the displayed 3-D content is caused by the product of all three factors (F1)–(F3), and thus, the factor (F3) should be integrated into the theoretical framework in the future work.

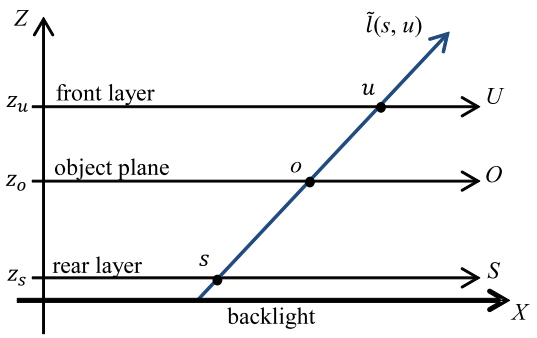


Fig. 2. Configuration of a light field display with two layers.

A. relation between disparity and depth

We derive the relation between the disparity of an object in the input images and the depth where the object is displayed.

Let I_m be the m-th input image and $I_m(n)$ be the n-th pixel. We assume that the input images are assigned to the viewing directions of the display with a constant interval k. Specifically, $I_m(n)$ is corresponded to the display's light field $l(s,u)$ by

$$l(n, n-km) = I_m(n), \tag{6}$$

in which (n,m) are associated with (s,u) in accordance with the rule

$$s = n,\ u = n - km. \tag{7}$$

Note that the direction of the light ray can be defined by $u-s$, so that the m-th input image corresponds to the direction $-km$ of the display. Given the above assignment, each input image pixel $I_m(n)$ is associated with a light ray emitted by the display.

In the configuration of fig. 2, the light rays emitted into a specific direction, which constitute a "directional view" of the display, are parallel with each other. Therefore, in a strict sense, each of the input images $I_m(n)$ should be captured through an orthographic projection. However, we make a compromise in the physical accuracy by using perspective images as the input instead of orthographic images, due to the limitation of the available data and our current technology. This limitation comes from the fact that perspective cameras are much more commonly used than orthographic cameras to capture real images. Moreover, the image-based rendering method that will be introduced in Section IV supports only the perspective projection model. Therefore, throughout this paper,

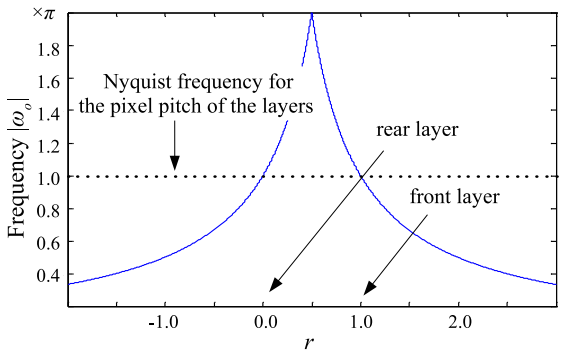


Fig. 3. Upper bound for spatial frequency of the light field display.

all the input images are captured as perspective images, but they are used directly as orthographic directional views for the display. This is based on the assumption that a perspective camera can approximate an orthographic camera if the viewing angle is narrow and the viewpoint is located far from the objects. This approximation induces nonlinear distortions along the depth on the displayed 3-D contents, but the disparities among the input images are preserved up to scale among the directional views emitted from the display. Therefore, we still have a natural depth sensation as demonstrated in the supplementary video [41]. The distortions along the depth should be addressed in the future work, because physically accurate visualization of the original 3-D volume is important and even essential to some applications.

Meanwhile, a light ray specified by (s,u), or equivalently intersecting with the two layers at $(X,Z)=(s,z_s)$ and (u,z_u), can be expressed as a straight line written as

$$(z_u-z_s)X-(u-s)Z-sz_u+uz_s=0. \tag{8}$$

By substituting eq. (7) into eq. (8), we obtain

$$(z_u-z_s)X+kmZ-nz_u+(n-km)z_s=0. \tag{9}$$

This is the path of the light ray in (X,Z) space that corresponds to the input image pixel $I_m(n)$.

Next, we analyze the position where the object is displayed in (X,Z) space. We consider an object point whose disparity among the adjacent input images is d. The same object point is observed from many input images with disparity d, resulting in a correspondence

$$I_m(n_0+md)\Leftrightarrow I_0(n_0)\quad \text{for } \forall\ m, \tag{10}$$

where n_0 is the position of the object point on 0-th input image I_0. Therefore, the pairs (n,m) satisfying

$$n=n_0+md \tag{11}$$

represent the input pixels recording the same object point. By substituting eq. (11) into (9), we obtain the group of line equations in (X,Z) space

$$(z_u-z_s)\,(X-n_0)+m\,\{kZ-dz_u-(k-d)z_s\}=0 \tag{12}$$

which always pass through a point

$$(X^*,\ Z^*)=\left(n_0,\ \frac{d}{k}z_u+(1-\frac{d}{k})z_s\right) \tag{13}$$

regardless of m. Equation (12) represents a group of light rays corresponding to the same object point. Therefore, the fixed point $(X^*,\ Z^*)$ is the position where the image of the object point is formed (or the object point is displayed) in (X,Z) space.

B. depth dependent upper bound of spatial frequency

We analyze the spatial frequency of an object displayed at a certain depth to derive the relation between the object depth and the frequency upper bound of the displayed object. This upper bound comes from the discrete representation of the light field, i.e., non-infinitesimal pixel size of the display. On the basis of this relation, we introduce a limitation on the object depth, which leads to a requirement on the configuration of the input images.

In [24], [27], the spatial frequency upper-bound due to the discrete representation of the light field have already been analyzed. In these works, the spectral support of the display is derived by analyzing the light field in the Fourier transform domain. Their sophisticated analysis can apply to more than three layers. Meanwhile, our analysis here provides a different viewpoint to the same issue. We only analyze whether each of the display panel has a sufficient resolution to make an aliasing-free sampling of a unit sinusoidal wave that is located at a certain depth. Not surprisingly, our analysis leads to the same conclusion as those in [24], [27]. Although our analysis is limited to two layers we believe our analysis is much simpler and easier to understand than [24], [27].

As shown in fig. 2, we assume that a planar object is displayed on a constant depth $Z=z_o$. A plane called the object plane is located at this depth, and a new axis O is defined to parameterize horizontal positions on the plane. For efficiency of analysis, a new value r is introduced to represent the relative depth of the object as

$$r=\frac{z_o-z_s}{z_u-z_s}. \tag{14}$$

When $0.0\le r\le 1.0$, the object plane is located between the layers. Otherwise, it is located outside the layers. Using r, the intersection point of a light ray specified by (s,u) and the object plane is described as

$$o=(1-r)s+ru. \tag{15}$$

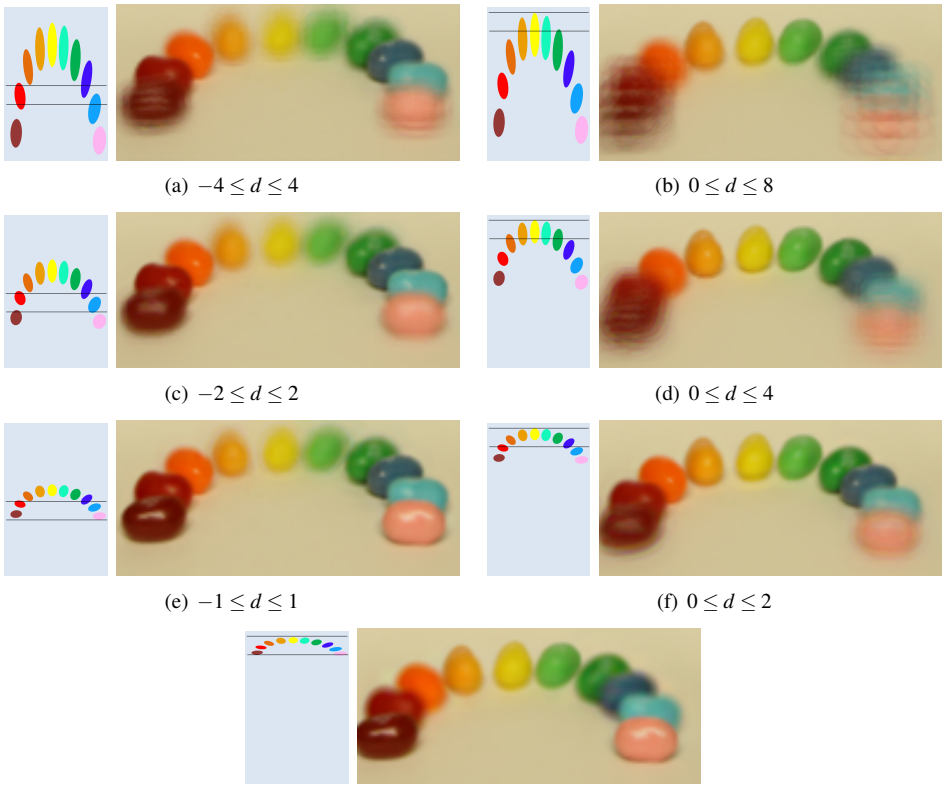
Now, we analyze the spatial frequency of the displayed object. A unit sinusoidal wave on the object plane is considered because any luminance pattern on the object plane can be represented as a sum of sinusoidal waves with different frequencies. According to eq. (15), a unit sinusoidal wave $e^{j\omega_o o}$ can be written as

$$e^{j\omega_o o}=e^{j(1-r)\omega_o s}e^{jr\omega_o u}. \tag{16}$$

Equation (16) indicates that $e^{j\omega_o o}$ is observed as $e^{j(1-r)\omega_o s}$ and $e^{jr\omega_o u}$ along the S and U axes, respectively. The frequencies along the S and U axes are written as

$$\omega_s=(1-r)\omega_o,\ \omega_u=r\omega_o. \tag{17}$$

On the other hand, the upper-bound frequencies that can be reproduced with discrete layer pixels are determined by



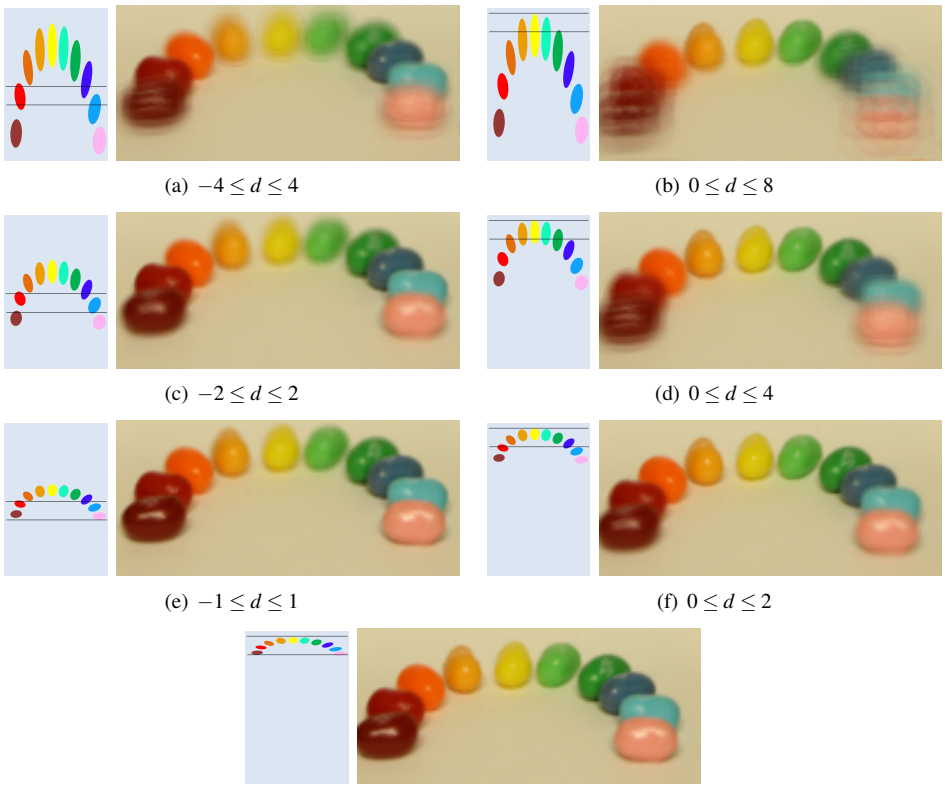
(a) $-4 \leq d \leq 4$ (b) $0 \leq d \leq 8$

(c) $-2 \leq d \leq 2$ (d) $0 \leq d \leq 4$

(e) $-1 \leq d \leq 1$ (f) $0 \leq d \leq 2$

(g) $0 \leq d \leq 1$

Fig. 4. The layout of the displayed objects and displayed images with rank 3 approximation.



(a) $-4 \leq d \leq 4$ (b) $0 \leq d \leq 8$

(c) $-2 \leq d \leq 2$ (d) $0 \leq d \leq 4$

(e) $-1 \leq d \leq 1$ (f) $0 \leq d \leq 2$

(g) $0 \leq d \leq 1$

Fig. 5. The layout of the displayed objects and displayed images without the effect of low rank approximation.

the Nyquist-Shannon theorem. For each layer, the frequency bound is described as $|\omega_s| \leq \pi$ or $|\omega_u| \leq \pi$. Using eq. (17), the frequency bound on the object plane is given by

$$|\omega_o| \leq \min\left\{\frac{1}{|r|}, \frac{1}{|1-r|}\right\}\pi. \tag{18}$$

This equation shows the upper-bound frequency that can be reproduced by the display at the relative depth r, which is illustrated in fig. 3. This upper bound takes the maximum at $r = 1/2$ and decreases as r diverges from $1/2$.

On the basis of frequency-domain analysis, we set a limitation on the depth where the objects are displayed. We want to represent all the details contained in the input images. In other words, the upper-bound frequency of the displayed object should be greater than the Nyquist frequency of the input images. In accordance with the assignment of eq. (6), the pixel sizes of input images and the layers are the same. Therefore, the upper-bound frequency of the displayed object should be greater than π in fig. 3. This condition is satisfied with the depth range $0.0 \leq r \leq 1.0$, or equivalently

$$z_s \leq z_o \leq z_u. \tag{19}$$

This equation means that the objects should be displayed within the two layers.

The above depth limitation can be translated into a limitation on disparities among the input images. By substituting Z^* of eq. (13) into z_o of eq. (19), we obtain

$$z_s \leq \frac{d}{k}z_u + (1-\frac{d}{k})z_s \leq z_u, \tag{20}$$

which can be simplified to

$$0 \leq \frac{d}{k} \leq 1. \tag{21}$$

To show the correctness of our analysis, we performed an experiment through computer simulation of a light field display with two layers. As the input, we used a dense light field from [38] with modifications on the image size and viewing angle. The original viewpoints are 17×17, and the range of disparities among the adjacent viewpoints was 0 to 1 pixels. We sub-sampled the original viewpoints of the input images by the factors of 2, 4, and 8, to change the range of disparities among the adjacent viewpoints. The farthest object had 0 pixel disparity among the adjacent viewpoints in all cases. Meanwhile, the maximum disparity (the disparity of the nearest object) among the adjacent viewpoints was changed to 2, 4, and 8 pixels by the sub-sampling. Although the number of viewpoints are different, the outermost viewpoints are the same in all the cases. Therefore, the disparity range between the leftmost and rightmost (topmost and bottommost) images are the same among all the cases. Similarly, we also prepared multi-view images with 9×9 viewpoints where the range of disparities among the adjacent viewpoints was -1 to 1 pixels. We sub-sampled the original viewpoints by the factors of 2 and 4, to make the disparity range -2 to 2 pixels and -4 to 4 pixels, respectively, while keeping the outermost viewpoints. To summarize, in the seven cases mentioned above, the disparity range among adjacent viewpoints are different case to case, but the absolute disparity ranges (the difference between the maximal and minimal disparities) among the outermost viewpoints are the same.

These multi-view images were converted to layer representations using eq. (4), where the degree of time multiplexing (the rank for the light field approximation) was set to 3. We set $k = 1$, which means the input images were assigned to the viewing directions of the display without intervals. For each of the seven cases, the layout of the displayed objects, which is given by eq. (13), and a displayed image observed from a specific viewpoint are illustrated in fig. 4. From top to bottom, the absolute disparity ranges among the adjacent viewpoints were 8, 4, 2, and 1 pixels, and the range of the object space with respect to the layers changed accordingly. While the layer interval is kept constant, the scene objects are virtually expanded/compressed along the depth direction when visualized on the display. The displayed images are observed from the direction perpendicular to the display layers, but the viewpoint was translated by half a pixel length both in horizontal and vertical directions. Among those seven displayed images, only the bottommost one (fig. 4(g)) satisfies eq. (21), which results in a relatively clear image throughout the depth, except for little amount of blur on the intermediate depth, which will be discussed later. Meanwhile, as for the other displayed images, the objects displayed outside the two layers are blurry and ghosted, which indicates lower frequency upper bounds outside the layers.

The blur observed around the intermediate depth in fig. 4(g) cannot be explained by the spatial frequency analysis presented above, where the upper bound frequency was derived solely from the discrete representation of the light field. We think this blur comes from another factor, the low rank approximation of the light field. In [34], we have analyzed a light field generated by a single planar object, and revealed that the rank of the light field depends on not only the texture complexity of the target object but also its depth, i.e. the distance from the display layers. As the object diverges from the display layers, the rank of the light field increases, and thus, the quality of the displayed image degrades because of the low rank approximation enforced in the layer pattern optimization. Accordingly, the intermediate depth between the two layers is a weak point of the display in terms of low rank approximation.

To verify the discussion above, we also present displayed images without the effect of the low rank approximation in figure 5, where the configuration was kept the same as fig. 4. Here and only here, we assumed that all the input images given to the display should perfectly be reproduced as the directional views of the display. However, due to the viewpoint shift in a sub-pixel level, the displayed images still exhibit quality degradation due to the discrete representation of the light field. More specifically, each of the light rays constituting a displayed image is interpolated from the discrete set of light rays that constitute the directional views of the display, and thus, the displayed image is blurred. In fig. 5, we can see that the objects located in-between the layers are clearly visible, while the objects displayed outside the two layers are blurry. In contrast to fig. 4(g), no blurs are observed inside the space between the layers in fig. 5(g). Those observations clearly support the correctness of our analysis on the frequency upper

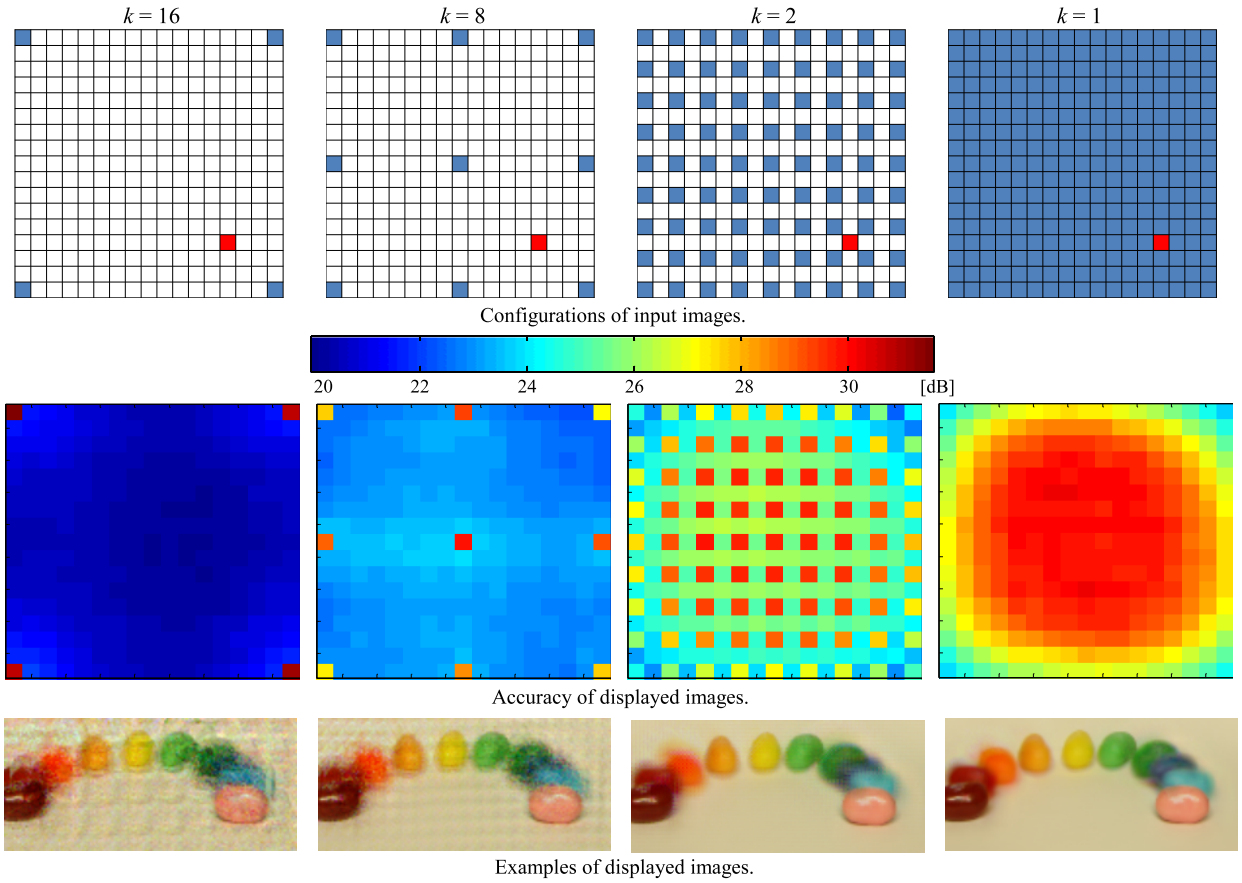


Fig. 6. Density of input light field and the quality of displayed images.

bound.

C. input images assignment to viewing directions

It is obvious from eq. (21) that the depth limitation is represented as a ratio between the disparity d in the input images and the interval of viewing directions k in the display. Therefore, with any given d, we can satisfy this condition by choosing appropriate values for k; as d increases we should increase k accordingly. However, there arises a natural question; does the value of k affect the quality of the displayed images? If k is greater than 1, the input images are given only to every k viewing directions. Other viewing directions are without input images. The display's output to such "unconstrained" directions is automatically generated through the optimization process of eq. (4) without guarantee of correctness.

To analyze the relation between k and the quality of the displayed images, we performed another experiment using the same dataset as section III-B. The original dataset had 17×17 viewpoints, and the disparity among the adjacent viewpoints was limited to 0 to 1 pixels. To change the range of disparities among the adjacent viewpoints , we sub-sampled the viewpoints as illustrated on the top row in fig. 6, where the remaining input images are marked in blue. The maximum disparity among the adjacent viewpoints was 16, 8, 2, and 1 from left to right. To satisfy eq. (21), we changed the value of k accordingly, 16, 8, 2, and 1 from left to right. More specifically, the display had 17×17 viewing directions that corresponded to the 17×17 viewpoints of the original input images. When an input image was skipped by the sub-sampling, the corresponding viewing direction became unconstrained. In all cases, all of the objects were displayed in-between the layers.

The center row of fig. 6 shows the accuracy of images displayed to the 17×17 viewing directions. The accuracy was measured by PSNR against the original 17×17 input images. When all of the viewing directions were filled with input images, the quality of the displayed images was kept high over all viewing directions. Meanwhile, when input images were not assigned to several viewing directions, the quality for these unconstrained directions was not good. As k increased, the quality for unconstrained directions became worse. The bottom row of fig. 6 shows the displayed images observed from the viewpoint marked in red in the top row. The visual quality with $k = 1$ was fine; however, strong ghosting artifacts appeared with larger values of k, although eq. (21) was satisfied.

From this experiment, we conclude that for high quality visualization, the input images should be assigned to all viewing directions within the effective viewing range without intervals, which means $k = 1$. By substituting $k = 1$ into eq. (21), we derive a more strict condition required for the input images as

$$0 \leq d \leq 1. \tag{22}$$

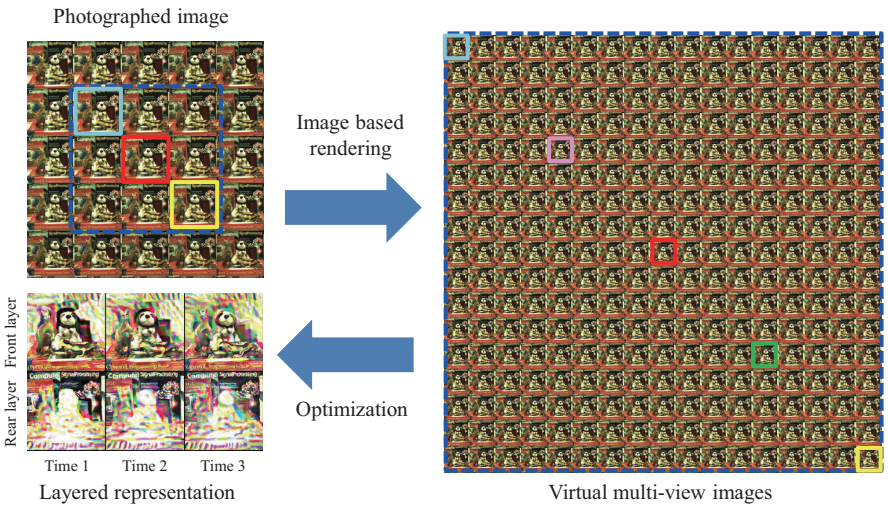


Fig. 7. Flowchart of light field data conversion.

(a) Camera and objects (b) Camera (c) Objects

Fig. 8. Experimental setup with 5×5 multi-view camera and objects.

Fig. 9. Left: configurations of real (top) and virtual (bottom) cameras. Right: several real (top) and virtual (bottom) images are also presented, where the boundary color of each image indicates the viewpoint in the left pane and the corresponding image in fig. 7

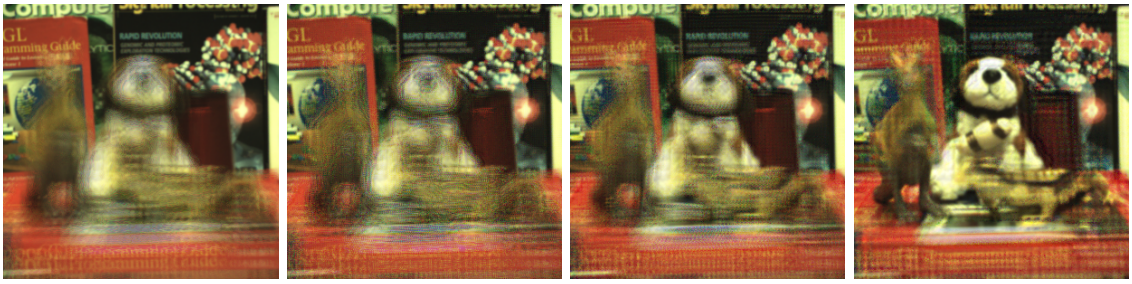


Fig. 10. Displayed images when sparsely photographed real multi-view images are directly input to the display. From left to right, k takes 1, 2, 4, and 8.

Fig. 11. Results when densely generated virtual multi-view images are directly input to the display. Left: accuracy of the 17×17 viewpoints in PSNR. Right: displayed images observed from different viewpoints marked in the left column.

IV. DATA CONVERSION USING IMAGE BASED RENDERING

In Section III, we have made the two claims for high quality 3D visualization: (C1) the depth range of the object space should be limited inside the two layers (Section III.B), and (C2) all of the directional views in the effective viewing range should have the assigned input images (Section III.C). These claims lead to eq. (22), which requires that the disparities among the adjacent input images are limited to only one pixel. In other words, we need to prepare significantly dense multi-view images as the input. This requirement does not become a problem for computer graphics contents because any number of input images with any interval can be easily rendered, as was done in [23], [26], [27]. However, in this study, we want to display real world contents that are captured by physical cameras. It is not always feasible to directly take photographs that meet this strict requirement.

As a practical solution to meet the requirement, we propose using a multi-view camera and image based rendering as illustrated in fig. 7. The procedure of our method is described below with specific configurations of our implementation, but our idea is also applicable to other configurations. First, real multi-view images were photographed by a multi-view camera. We used a ViewPLUS ProFUSION 25 which had 5×5 viewpoints arranged in a 2-D grid, and each camera had 640×480 pixels. Our experimental setup is visualized in fig. 8, and examples of photographed images with rectification are presented on the top row of fig. 9. Second, these real images were converted to denser virtual multi-view images via image based rendering to satisfy the required condition of eq. (22). Specifically, we generated 17×17 virtual multi-view images from the real 5×5 images. Examples of these virtual images and their viewpoint arrangement are illustrated on the bottom row of fig. 9. Here, the outermost virtual viewpoints are located on the outer edge of the inner 3×3 real viewpoints. Note that in figs. 7 and 9, the colors attached around the images indicate the viewpoints, and the same color represents the same viewpoint. Therefore, the 17×17 virtual viewpoints cover the same disparity range as those of the inner 3×3 viewpoints of the real images. In other words, 3×3 real images are densified into 17×17 virtual images. We used a layer-based method [32], [39], [40] for image based rendering. Finally, the generated 17×17 virtual multi-view images were reconverted into layer patterns by using the images as the target light field $\mathbf{L}$ in the optimization of eq. (4). Our method can easily be applied to moving objects because a set of multi-view images is captured at a time, which is impossible with sequential move-and-capture methods using a single camera [38].

To show the importance of our data conversion method, we performed experiments and compared two cases where (i) actually photographed 3×3 (inner 3×3 out of originally captured 5×5) images and (ii) virtually generated 17×17 images were used as the input to the display.

For the first case, the photographed images were rectified and cropped, which were minimal operations to use them as the input, resulting in 320×320 pixels for each image. The range of disparities among the adjacent viewpoints was approximately 0 to 8 pixels. In this case, we have several options for the value of k in assigning the input images to the directional views of the display. Changing the value of k makes a trade-off between the two claims (C1) and (C2). When k is set to 1, all the directional views are assigned the input images, but the object space largely exceeds the layer interval, which

causes quality degradation as was analyzed in Section III.B. Meanwhile, when k is set to 8, the object space is limited inside the two layers, but many directional views are left unconstrained without assigned input images, which causes aliasing artifacts as was discussed in Section III.C. Anyway, two claims (C1) and (C2) cannot be satisfied simultaneously, because eq. (22) is not satisfied. Several displayed images with different k are presented in fig. 10, where k was set to 1, 2, 4, and 8 from left to right. Significant noises, which seem to be moiré and color-channel crosstalk, are observed on those image. Note that these noises are not caused by display hardware issues, because those images are generated by computer simulation. For the first three images from left, the object space was larger than the layer interval, resulting in blurry outputs outside the layers. Meanwhile, only the rightmost one with $k = 8$ satisfied the condition of eq. (21); the entire object space was located within the layers, resulting in the best image quality among the four images. However, due to the large interval among the constrained directional views, the image quality was still insufficient. To conclude, as far as using the input images that do not satisfy eq. (22), we cannot meet the two claims (C1) and (C2) simultaneously, and thus, cannot achieve sufficient visual quality.

On the other hand, the second case, where the resolution of the virtual images were also set to 320×320 pixels, meets the required condition of eq. (22); the range of disparities among the adjacent viewpoints was approximately 0 to 1 pixels. In this case, the two claims (C1) and (C2) can be satisfied simultaneously; these input images were assigned to the viewing directions of the display without intervals ($k = 1$), and the depth range of the object space is limited inside the two layers. The accuracy of the displayed images observed from 17×17 viewing directions and three examples of these images are presented in fig. 11. We achieved satisfactory quality in this case; all the objects were clearly visualized throughout the depth from any viewing direction within the range supported by the input images.

A supplemental video is available from [41].

V. CONCLUSION

To visualize real world 3-D objects with high quality, we studied light field displays with two stacked multiplicative layers. We presented three analyses based on theory and experiments. First, we derived the relation between the disparities in the given light field and the depth where the objects are displayed. Second, we analyzed the upper-bound spatial-frequency of the display, which was shown to be dependent of the depth where an object is displayed. Finally, we investigated the relation between the quality of the displayed images and the interval with which the given light field corresponded to the viewing directions of the display. From those analyses, we derived a condition required for the given light field: the disparity range among the adjacent viewpoints of the given light field should be limited to 0 to 1 pixels. To meet this very strict condition with real world objects, we proposed using a multi-view camera and image based rendering, where we can generate virtual multi-view images with arbitrary density, and verified its effectiveness through experiments.

Our future work includes several directions. Our theory and analyses will be extended to different light field displays such as one with more than two layers and a projection based display. Our method will be verified not only with computer simulations but with physical display hardware. Another interesting direction is to use other types of cameras such as Lytro [42] and coded mask cameras [43], [44], [45], [46], [47] to capture input light field data. Finally, we want to develop an end-to-end system where real world 3-D objects are captured, converted, and displayed in 3-D with satisfying quality.

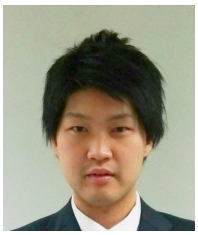
ACKNOWLEDGMENT

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

REFERENCES

[1] Okoshi, T., "Three-dimensional displays," Proceedings of the IEEE, vol. 68, no. 5, pp. 548–564, 1980.
[2] Pastoor, B., and Wopking, M., "3-D displays: A review of current technologies," Elsevier Displays, vol. 17, no. 2, pp. 100–110, 1997.
[3] Travis, A. R. L., "The display of three-dimensional video images," Proceedings of the IEEE, vol. 85, no. 11, pp. 1817–1832, 1997.
[4] Javidi, B., and Okano, F., "Three-dimensional television, video, and display technologies," Springer Science & Business Media, 2002.
[5] Hong, S. H., Jang, J. S., and Javidi, B., "Three-dimensional volumetric object reconstruction using computational integral imaging," Optics Express, vol. 12, no. 3, pp. 483–491, 2004.
[6] Borner, R., "Auto stereoscopic 3D-imaging by front and rear projection and on flat panel displays," Elsevier Displays, vol. 14, no. 1, pp. 39–46, 1993.
[7] Matusik, W., and Pfister, H., "3D TV: A Scalable System for Real-Time Acquisition, Transmission and Autostereoscopic Display of Dynamic Scenes," ACM Transactions on Graphics, vol. 23, issue 3, pp. 814–824, 2004.
[8] Takaki, Y., and Nago, N., "Multi-projection of lenticular displays to construct a 256-view super multi-view display," Optics express, vol. 18, no. 9, pp. 8824–8835, 2010.
[9] Hong, J., Kim, Y., Park, S. G., Hong, J. H., Min, S. W., Lee, S. D., and Lee, B., "3D/2D convertible projection-type integral imaging using concave half mirror array," Optics express, Vol. 18, Issue 20, pp. 20628–20637, 2010.
[10] Iwasawa, S., Kawakita, M., and Inoue, N., "REI: an automultiscopic projection display," Proceedings of 3DSA, Selected paper 1, 2013.
[11] Lee, J. H., Park, J., Nam, D., Choi, S. Y., Park, D. S., and Kim, C. Y., "Optimal projector configuration design for 300-Mpixel multi-projection 3D display," Optics express, vol. 21, Issue 22, pp. 26820–26835, 2013.
[12] Ives, F. E., "Parallax stereogram and process of making same," U.S. Patent US72556A, 1903.
[13] Isono, H., Yasuda, M., and Sasazawa, H., "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.
[14] Sakamoto, K., and Morii, T., "Multiview 3D display using parallax barrier combined with polarizer," Proc. SPIE 6399, Advanced Free-Space Optical Communication Techniques/Applications II and Photonic Components/Architectures for Microwave Systems and Displays, 63990R, 2006.
[15] Peterka, T., Kooima, R. L., Sandin, D. J., Johnson, A., Leigh, J., and DeFanti, T. A., "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.
[16] Lippmann, G., "Epreuves reversibles donnant la sensation du relief," Journal of Physics, vol. 7, no. 4, pp. 821–825, 1908.
[17] McCormick, M., "Integral 3D imaging for broadcast," In Proc. 2nd Int. Display Workshop, vol. 3, pp. 77–80, 1995.
[18] Arai, J., Okano, F., Kawakita, M., Okui, M., Hanio, Y., Yoshimura, M., Furuya, M., and Sato, M., "Integral three-dimensional television using a 33-megapixel imaging system," Journal of Display Technology, vol. 6, no. 10, pp. 422–430, 2010.

[19] JCT3V-C0210, "3D Holoscopic Video Test Material," FP7 3D VIVANT Consortium, 2013.
[20] Suyama, S., Takada, H., and Ohtsuka, S., "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.
[21] Gotoda, H., "A multilayer liquid crystal display for autostereoscopic 3D viewing," Proc. SPIE 7524, Stereoscopic Displays and Applications XXI, 75240P, 2010.
[22] Yoon, S., Baek, H., Min, S.-W., Park, S.-G., Park, M.K., Yoo, S.-H., Kim, H.-R., and Lee, B., "Implementation of active-type Lamina 3D display system," Opt. Express, vol. 23, no. 12, pp. 15848–15856, 2015.
[23] Lanman, D., Hirsch, M., Kim, Y., and Raskar, R., "Content-adaptive parallax barriers: optimizing dual-layer 3D displays using low-rank light field factorization," ACM Transactions on Graphics, vol. 29, issue 6, article 163, 2010.
[24] Wetzstein, G., Lanman, D., Heidrich, W., and Raskar, R., "Layered 3D: tomographic image synthesis for attenuation-based light field and high dynamic range displays," ACM Transactions on Graphics, vol. 30, issue 4, article 95, 2011.
[25] Lanman, D., Wetzstein, G., Hirsch, M., Heidrich, W., and Raskar, R., "Polarization fields: dynamic light field display using multi-layer LCDs," ACM Transactions on Graphics, vol. 30, issue 6, article 186, 2011.
[26] Lanman, D., Wetzstein, G., Hirsch, M., Heidrich, W., and Raskar, R., "Beyond parallax barriers: applying formal optimization methods to multilayer automultiscopic displays," Proc. SPIE 8288, Stereoscopic Displays and Applications XXIII, 82880A, 2012.
[27] Wetzstein, G., Lanman, D., Hirsch, M., and Raskar, R., "Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting," ACM Transactions on Graphics, vol. 31, issue 4, article 80, 2012.
[28] Huang, F. C., Chen, K., and Wetzstein, G., "The light field stereoscope: immersive computer graphics via factored near-eye light field displays with focus cues," ACM Transactions on Graphics, vol. 34, issue 4, article 60, 2015.
[29] Chai, J. X., Tong, X., Chan, S. C., and Shum, H. Y., "Plenoptic sampling," ACM Transactions on Graphics, pp. 307–318, 2000.
[30] Zhang, C., and Chen, T., "Spectral analysis for sampling image-based rendering data," IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 11, pp. 1038–1050, 2003.
[31] Do, M. N., Marchand-Maillet, D., and Vetterli, M., "On the bandwidth of the plenoptic function," IEEE Transactions on Image Processing, vol. 21, no. 2, pp. 708–717, 2012.
[32] Taguchi, Y., Koike, T., Takahashi, K., and Naemura, T., "TransCAIP: A live 3D TV system using a camera array and an integral photography display with interactive control of viewing parameters," IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 5, pp. 841–852, 2009.
[33] Saito, T., Takahashi, K., Tehrani, M. P., and Fujii, T., "Data conversion from multi-view cameras to layered light field display for aliasing-free 3D visualization," Proc. SPIE 9391, Stereoscopic Displays and Applications XXVI, 939111, 2015.
[34] Takahashi, K., Saito, T., Tehrani, M. P., and Fujii, T., "Rank analysis of a light field for dual-layer 3D displays," Proc. IEEE International Conference on Image Processing, pp. 4634–4638, 2015.
[35] Ho, N. D., Van Dooren, P., and Blondel, V., "Weighted nonnegative matrix factorization and face feature extraction," Image and Vision Computing, pp. 1–17, 2008.
[36] Cichocki, A., Zdunek, R., Phan, A. H., and Amari, S. I., "Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation," John Wiley & Sons, 2009.
[37] Hirsch, M., Wetzstein, G., and Raskar, R., "A compressive light field projection system," ACM Transactions on Graphics, vol. 33, issue 4, article 58, 2014.
[38] "The (New) Stanford Light Field Archive, Computer Graphics Laboratory, Stanford University," http://lightfield.stanford.edu/.
[39] Takahashi, K., and Naemura, T., "Layered light-field rendering with focus measurement," Signal Processing: Image Communication, vol. 21, no. 6, pp. 519–530, 2006.
[40] Taguchi, Y., Takahashi, K., and Naemura, T., "Real-time all-in-focus video-based rendering using a network camera array," Proc. IEEE 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, pp. 241–244, 2008.
[41] "Light Field Display Project," http://www.fujii.nuee.nagoya-u.ac.jp/~takahasi/Research/LFDisplay/
[42] "Lytro," https://www.lytro.com.
[43] Veeraraghavan, A., and Rasker, R., "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," ACM Transactions on Graphics, vol. 26, issue 3, article 69 2007.
[44] Liang, C. K., Lin, T. H., Wong, B. Y., and Chen, H. H., "Programmable aperture photography: Multiplexed light field acquisition," ACM Transactions on Graphics, vol. 27, issue 3, article 55, 2008.
[45] Nagahara, H., Zhou, C., Watanabe, T., Ishiguro, H., and Nayar, S. K., "Programmable aperture camera using LCoS," in Proc. ECCV, pp. 337–350, 2010.
[46] Babacan, S. D., Ansorge, R., Luessi, M., Mataran, P. R., Molina, R., and Katsaggelos, A. K., "Compressive light field sensing," IEEE Transactions on Image Processing, vol. 21, no. 12, pp. 4746–4757, 2012.
[47] Marwah, K., Wetzstein, G., Bando, Y., and Raskar, R., "Compressive light field photography using overcomplete dictionaries and optimized projections," ACM Transactions on Graphics, vol. 32, issue 4, article 46, 2013.



Toyohiro Saito received his B.E. and M.E. degrees in electrical engineering and computer science from Nagoya University, Japan, in 2014 and 2016. His research interests were 3-D displays, image based rendering, and light field acquisition. He is currently working at Nippon Steel & Sumitomo Metal Corporation.

Yuto Kobayashi received his B.E. degree in electrical engineering from Nagoya University, Japan, in 2016. He is currently a graduate student in electrical engineering and computer science at Nagoya University. He is working on light field acquisition and rendering for 3-D displays.

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. 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. His research interests include computational photography, image-based rendering, and 3-D display.

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. 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. His current research interests include multi-dimensional signal processing, multi-camera systems, multi-view video coding and transmission, free-viewpoint television, and their applications.