https://doi.org/10.2352/ISSN.2470-1173.2017.5.SD&A-373

Improved Depth of Field Analysis of Multilayer Displays

Hironobu Gotoda; Associate Professor; Tokyo, Japan

Abstract

The depth of field (DOF) of an auto-stereoscopic display refers to the depth range in 3D space in which objects can be depicted with small amount of blur. It provides a measurable index on the display's performance in reproducing light fields of 3D scenes. Previous studies have analyzed the maximum spatial frequencies of aliasing-free images depicted on planes parallel to the display's surface. For multilayer displays, several formulae representing the upper bounds on the maximum frequencies have been given. However, these formulae provide little information on how much blur would be present in the reproduced fields, since contributions of low frequency signals are simply neglected. Such signals are frequently damaged on multilayer displays especially when the angular range of viewing angles becomes wide. To address these drawbacks, we present a novel framework for the DOF analysis of multilayer displays. The analysis begins with a close look at the synthesis of layer images, which can be considered as solving a linear least squares problem with nonnegativity constraints. This numerical procedure is then reinterpreted in the context of multilayer displays, where part of the connections between "depth" and "blur" are observed. Finally, experimental results supporting these observations are presented.

Introduction

Depth of field (DOF) is a common term in photography referring to the distance between nearest and farthest objects in a scene that appear acceptably sharp. It is also called "focus range" or "effective focus range". The DOF of a camera varies depending on the camera type, aperture and focusing distance.

Zwicker et al [8] extended the concept of DOF to quantify the capabilities of 3D displays in reproducing stereoscopic images. They defined DOF of a 3D display as the range of planes parallel to the display's surface on which images can be displayed at maximum spatial resolution. This definition was obviously dual to photographic DOF or the range of exact refocusing in light field photography [5]. Zwicker et al. derived analytic formulae of DOF, which were applicable to displays based on parallax barriers, lenticular sheets, integral lens sheets, and multi-projector systems.

More recently, Wetzstein and Lanman et al. [2, 6] studied DOF of 3D displays based on stacks of semi-transparent flat-panel imaging devices. We will simply call such type of displays "multilayer displays" in the remainder of this paper. In a series of studies, Wetzstein and Lanman et al. applied signal processing methods to the light fields emitted from multilayer displays, analyzed the light fields in the frequency domain, and derived formulae representing the upper-bounds of the DOF of multilayer displays. However, since the formulae were based on the maximum spatial frequencies "achievable" with multilayer displays, the actual DOF corresponding to a particular light field could be substantially lower. Moreover, the formulae did not contain any factors dependent on the range of viewing angles. In practice, we often observe that visual artifacts are more present in the reconstructed light field when the angular viewing range becomes wider. Such dependence on the viewing range is missing in the derived DOF formulae.

In this paper, we present a novel approach to the DOF analysis of multilayer displays. We first take a close look at the synthesis of layer images, where light fields are transformed and mapped to images on multiple layers. This process can be considered as solving linear least squares problems with nonnegativity constraints (NNLS). Recent findings in linear algebra enable us to derive a semi-analytic formula expressing the solutions of NNLS problems. The formula provides the foundation of our DOF analysis. We will then reinterpret the formula in the context of multilayer displays, where part of the connections between "depth" and "blur" will be observed. In particular, "depth selective light field reconstruction method" will be presented as a substitute for an NNLS solver, which will bring us an intuitive view on the DOF of multilayer displays. Finally, several experimental results will be presented supporting this intuitive view.

An Analytic View on Multilayer Displays

A multilayer display is an autostereoscopic display consisting of a stack of multiple liquid crystal (LC) panels and a backlight [4]. The panels in the display are arranged in parallel with equal spacing between neighboring panels. Images shown on these panels are superimposed and delivered to viewers. The superimposition of images is direction-dependent, and forms a light field representing a 3D scene.

The optical properties of the stack of LC panels have been studied by many researchers. Nowadays most LC panels embedded in LC displays are multi-domain LC panels. When such types of LC panels are stacked, the following linear model[1]

$$b_j = \sum_{i=1}^{n} A_{ij} x_i \tag{1}$$

gives a good approximation of the ray's intensity b_j passing through pixels $x_{j1}, x_{j2}, \cdots, x_{jN}$, where N is the number of layers.

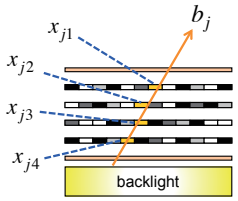

Figure 1. *Ray emitted from a 4-layer display. According to the linear model of stacking multi-domain LC cells, the ray's intensity b_j is equal to the sum of pixel's intensity $x_{j1}+x_{j2}+x_{j3}+x_{j4}$.*

Fig. 1 shows an example of a 4-layer display illustrating that b_j is equal to the sum of 4 numbers $x_{j1}+x_{j2}+x_{j3}+x_{j4}$. Equation (1) describes a more general situation, where the total number of n pixels residing in N layers are involved. The coefficient A_{ij} is either 0 or 1 depending on whether the ray b_j passes through pixel x_j or not.

We need to control the light field emitted from a multilayer display. The only way to control the light field is to adjust the pixels' intensity (i.e., transmittance of light). Let $b \in R^m$ be the light field (i.e., array of rays' intensities) to be reproduced by the multilayer display, $x \in R^n$ be the array of pixels' intensities in the display, and $A \in R^{m\times n}$ be the matrix representing the geometrical relationships between rays and pixels. The vector x can be determined by solving the following least squares problem

$$\min_{x\in R^n} \|Ax-b\| \qquad \text{subject to} \qquad 0 \le x \le 1 \tag{2}$$

where we have assumed that each element x_i of x is normalized to fit in the range $x_i \in [0,1]$.

In our experience with multilayer displays, the upper bound constraint $x \le 1$ is rarely broken. Therefore, even if we solve the modified problem,

$$\min_{x\in R^n} \|Ax-b\| \qquad \text{subject to} \qquad 0 \le x \tag{3}$$

the solution of (3) is the same as that of (2). On the other hand, the lower bound constraint $x \le 0$ is frequently broken. In most cases, the solution of the unconstrained problem (4).

$$\min_{x\in R^n} \|Ax-b\| \tag{4}$$

contains negative x_i's, and is different from the solution of (3).

We therefore adopt (3) as the basis of our analysis. Recent progress in linear algebra has revealed that a semi-analytic expression of the solution of (3) exists as described in the next subsection.

Solving the Nonnegative Least Squares Problem

The nonnegative least squares problem (3) can be solved using the modulus transformation $x = z+|z|$, where $z \in R^n$ is a real vector. Note that the vector z does not have to be a nonnegative vector. Even if some elements of z are negative, the transformation ensures that the corresponding elements of x are all 0. This means that x is guaranteed to be nonnegative by the modulus transformation.

Zheng et al. [7] developed several numerical methods to solve (3) using the modulus transformation. Referring to some earlier work, they argued that the solution of (3) can be obtained by solving the fixed-point equation

$$(I+A^TA)z = (I-A^TA)|z|+A^Tb \tag{5}$$

and setting $x = z+|z|$. Suppose that the fixed point z^* of equation (5) is actually found. It then follows that

$$A^TA(z^*+|z^*|) = A^Tb+(|z^*|-z^*) \tag{6}$$

$$A^TAx^* = A^Tb+c \tag{7}$$

$$x^* = (A^TA)^{-1}(A^Tb+c) \tag{8}$$

where $c = |z^*|-z^*$ is a nonnegative vector. It is worth to mention that vector c can be determined only after the solution of (5) is found. Equation (8) gives a semi-analytic expression of the solution of (3) although it contains an unknown nonnegative vector c. If one can estimate the amount of c, the analysis of the solution x^* becomes possible using equation (8).

Novel Framework of Depth of Field Analysis

The semi-analytic expression (8) of the solution of nonnegative least squares problem (3) provides many clues to the DOF analysis of multilayer displays. The expression is simple, and consists of only three terms: $(A^TA)^{-1}$, A^Tb and c. In the following subsections, we will interpret the first two terms (i.e., $(A^TA)^{-1}$ and A^Tb) in the context of multilayer displays. We will also try to interpret the last term (i.e., c), which turns out to be a difficult task. In our experience with multilayer displays, the magnitude of c is modest in many cases. Therefore the dominant factors affecting the quality of the light fields emitted from displays are A^TA and A^Tb.

Light Accumulation on Layers

The matrix-vector product A^Tb can be interpreted as the accumulation of light on layers. To see this, let us review the following equation.

$$b_j = \sum_{i=1}^{n} A_{ij}x_i \tag{9}$$

Each ray b_j passes through several pixels. The coefficient A_{ij} describes whether the ray pass through pixel x_i (i.e., $A_{ij}=1$) or not (i.e., $A_{ij}=0$). The coefficients $A_{0j}, A_{1j}, \cdots A_{nj}$ forms the j-th row vector of matrix $A \in R^{m\times n}$. If we look at the i-th column of matrix A, we notice that the $\{0,1\}$ pattern in the column represent which ray passes through pixel x_i.

Now consider the i-th value of vector A^Tb. The value can be obtained by computing the inner product of the i-th column vector of A and light vector b. The inner product represents the sum of all rays' intensities $\sum_j A_{ji}b_j$ passing through pixel x_i.

Fig. 2(a) illustrates the accumulation of light via the computation of A^Tb. Optically equivalent accumulation of light using a refractive lens is shown in Fig. 2(b). Here, the aperture of the refractive lens (b) corresponds to the angular viewing range of the multilayer display (a). As the viewing range becomes wider, more

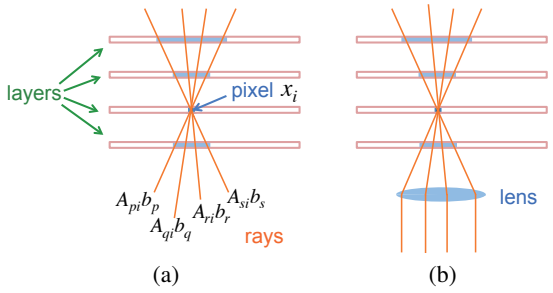

Figure 2. *Accumulation of light on the same pixel. As shown in (a), the weighted sum of rays' intensities $\sum_j A_{ji}b_j$ represent how much light accumulates on pixel x_i. Such accumulation of light can be re-interpreted as focusing light on pixel x_i using a refractive lens.*

Figure 3. *Accumulation images formed on layers, where $layer = 1$ indicates the nearest layer, and $layer = 4$ the farthest from the viewer; H and W represent the range of horizontal and vertical viewing angles, respectively.*

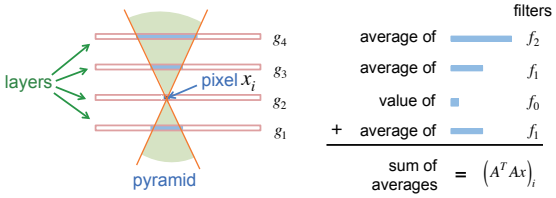

Figure 4. *The weighted sum of pixels' intensities within a pyramid, whose apex is located at pixel x_i as shown in (a), is equal to the i-th element of $A^T Ax$. The weighted sum can be computed by averaging the pixels' intensities in each layer, and adding the averaged values together (b).*

focal blur is to be introduced in the images formed by light accumulation. Such images are called "accumulation images" in the rest of the paper. Accumulation images are formed only on layers. Fig. 3 shows series of accumulation images under two viewing conditions. Top row images correspond to the condition where the horizontal viewing angle ranges between $-5.0°$ and $5.0°$, and vertical angle between $-2.5°$ and $2.5°$. Bottom row images correspond to the condition where the viewing ranges are 1.8 times wider than the top ones in both horizontal and vertical directions.

The formula (8) states that the accumulation images (as represented by $A^T b$) are the input sources for computing the solution vector x^*. If these images are already blurred, the resulting solution vector x^* will also be blurred, leading to blurry reconstruction of the light field.

Conversion from Accumulation to Layer Images

Given the vector x representing the array of pixels' intensities, the matrix-matrix-vector product $A^T Ax$ can be interpreted as collecting the pixels within a pyramid illustrated in Fig. 4, and computing the weighted sum of intensities of these pixels. In Fig. 4(a), a 2D slice of a pyramid is drawn, whose apex is located at pixel x_i. The pyramid is a solid consists of lines passing through x_i, and the base of the pyramid is an rectangle, whose shape is determined by the horizontal and vertical ranges of viewing angles. The weight applied to each pixel is inversely proportional to the number of pixels lying on the same layer as shown in Fig. 4(b). The derivation of such interpretation of $A^T Ax$ is almost obvious. Here we only remark that $(Ax)_j$ corresponds to the sum of pixels' intensities that lie along ray b_j.

We then examine the equation (7), which is equivalent to (8). As explained, $A^T Ax$ can be rewritten as the weighted sum of pixels' intensities. The weighted sum can be obtained by first applying an appropriate mean filter to each layer, and then merging the resultant images. Let $g_1, g_2, \cdots, g_N$ be the "layer images" (i.e., images displayed on layers) represented by vector x, and let $f_0, f_1, f_2, \cdots, f_{N-1}$ be the set of mean filters applicable to the layer images. In the illustrated example shown in Fig. 4, filters $\{f_0, f_1, f_2\}$ are applied to images $\{g_1, g_2, \cdots, g_4\}$, which gives $f_1 * g_1 + f_0 * g_2 + f_1 * g_3 + f_2 * g_4$. Here, $f_r * g_s$ represent the convolution of filter f_r and layer image g_s. Computing $f_1 * g_1 + f_0 * g_2 + f_1 * g_3 + f_2 * g_4$ is equivalent to computing the weighted sum $(A^T Ax)_i$ for each pixel x_i on the second layer. Generalizing this observation will lead to

$$[A^T Ax]_k = \sum_{i=1}^{N} f_{\mathrm{abs}(k-i)} * g_i \tag{10}$$

where $[y]_k$ represent the 2D array of real numbers extracted from vector $y \in R^n$ lying on the k-th layer. The equation (7) is now rewritten as

$$\sum_{i=1}^{N} f_{\mathrm{abs}(k-i)} * g_i = [A^T b]_k = h_k. \tag{11}$$

Notice that $[A^T b]_k$ represents the accumulation image formed on the k-th layer, which is denoted as h_k for the brevity of explanation.

Finally, Fourier transforms are applied to (11). This result is a system of N simultaneous equations

$$\begin{pmatrix} H_1 \\ H_2 \\ \vdots \\ H_N \end{pmatrix} = \begin{pmatrix} F_0 & F_1 & \cdots & F_{N-1} \\ F_1 & F_0 & \cdots & F_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ F_{N-1} & F_{N-2} & \cdots & F_0 \end{pmatrix} \begin{pmatrix} G_1 \\ G_2 \\ \vdots \\ G_N \end{pmatrix} \tag{12}$$

where $F_k(u,v) = \mathscr{F}(f_k(i,j))$, $G_k(u,v) = \mathscr{F}(g_k(i,j))$, and $H_k(u,v) = \mathscr{F}(h_k(i,j))$ represent Fourier transforms of the mean filter f_k, layer image g_k, and accumulation image h_k, respectively.

The equation (12) gives a convenient tool for analyzing the conversion from accumulation images $\{h_1, h_2, \cdots, h_N\}$ to layer images $\{g_1, g_2, \cdots, g_N\}$ in the frequency domain. The conversion is governed by a $N \times N$ matrix in which elements in $\{F_0, F_1, \cdots, F_{N-1}\}$ are arranged in a symmetric manner. The matrix are dependent on the window size of the mean filter denoted by f_1.

Fig. 5 shows three plots of the condition number of the matrix appearing in (12). Note that the matrix depends on the spatial frequency (u,v), and the condition number can be evaluated independently for each frequency. The horizontal axis represent the ratio of the frequency to the maximum frequency in percentage(%), and vertical axis represent the condition number. Three plots corresponding to different window sizes of the mean filter f_1 are shown here, where $w = 5, 10$, and 15 indicate that the filter's window sizes are 5%, 10%, and 15% of the image size, respectively. The image size used for this simulation is $640x360$, and the number of layers is 4 (i.e., $N = 4$).

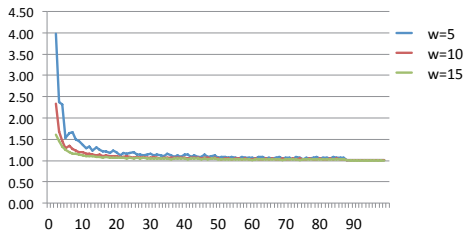

Figure 5. *Distribution of the condition number of the matrix appearing in equation (12). The horizontal axis represent the ratio of spatial frequency (u,v) to the maximum spatial frequency $(u_{\max}, v_{\max})$, and vertical axis represent the condition number. The number is evaluated independently for each frequency. Three plots are shown here, which differ in the window size of the mean filter f_1 used to make up the matrix. For example, $w = 5$ indicates that the window size is 5% of the image size.*

In Fig. 5, plots corresponding to the first 2 frequencies (i.e., lowest 2 frequencies) are dropped from the graph since the condition numbers are extremely large. However, in all other frequencies including the highest, the condition numbers stay within the rage of $[1,4]$. This implies that equation (12) is stable in almost all frequencies, and no cut-off of frequency components will take place in the conversion from the accumulation to layer images.

Evaluating Nonnegativity Constraint

Among the three terms included in formula (8), we have already discussed $(A^TA)^{-1}$ and A^Tb. We have seen that A^Tb may introduce blur in the reconstructed light field, while $(A^TA)^{-1}$ will retain high-frequency components as much as possible. The remaining task is to evaluate the last term c.

Unfortunately, we have not arrived at any convincing models of c. We only know that c is certainly related to the nonnegativity constraint in (3). Once the solution to (3) is obtained, we could evaluate the magnitude of c, and its influence to the reconstructed light field. However, before the solution of (3) is computed, we have little information to evaluate the magnitude and influence of c. Described below are a few attempts to locate the regions in 3D scenes, where c_i will likely take a positive value.

Let us start our discussion by analyzing the solutions of a few sample scenes. Since $c \in R^n$ is a vector whose dimension is the same as the number of pixels in the display, it is possible to map c on the layers. The superimposition of the mapped images can be observed from various directions, which provide stereoscopic views.

Fig. 6 shows four sample scenes and their corresponding images representing vector c. Here only the central views are shown. Each vector c is computed from the solution of an NNLS problem (3) , where the same 4-layer display is used to determine the matrix A. Details of the display's configuration are omitted here since the general tendency of c is the same. Bright pixels in the bottom row images indicate that magnitude of c is large at some layers.

From the results shown in Fig. 6, one could speculate that the magnitude of c would become large at such pixels where large directional variations exist in the intensity of light. One could also conjecture that bright objects occluding dark background would be the sources of positive c. To validate these hypotheses, two

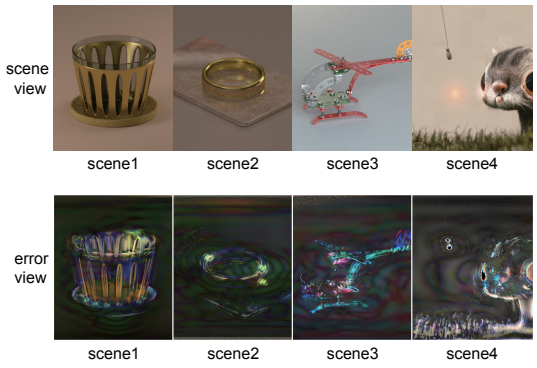

Figure 6. *Visualization of vector c in 3D space. Top row images show the central views of 4 sample scenes directly extracted from the light fields, while bottom row images show the corresponding views obtained by solving the least squares problem using a 4-layer display, and mapping the resultant vector c onto the same display. The intensity of each pixel x_i is determined by the magnitude of c_i, where larger c_i results in brighter intensity of x_i. The vector c is computed independently for each color channel.*

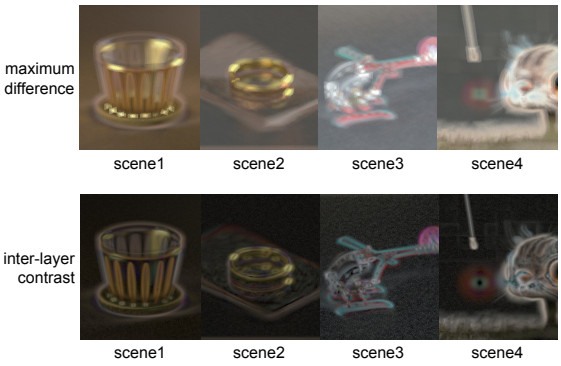

Figure 7. *Estimation of vector c using two methods. The maximum difference method measures the difference between the maximum and minimum intensities of rays passing through the same pixel, while the inter-layer contrast method measures the differences in the accumulation images between pairs of layers neighboring to each other.*

methods have been developed for estimating the magnitude of c. One is the maximum difference method, which, for each pixel in a multilayer display, measures the difference between the maximum and minimum intensity of light passing through that pixel. The other method is called inter-layer contrast method, which measures the difference in the accumulation images between any pairs of layers neighboring to each other. In the latter method, the pixels are first traced along each ray, then differences between consecutive pairs of pixels are calculated, and finally averaged over all rays passing through the same pixel. In averaging the differences, we only deal with cases where bright pixels appear nearer than darker pixels.

Fig. 7 shows the results obtained by applying the maximum difference and inter-layer contrast methods to the sample scenes as in Fig. 6. Compared to the maximum difference method, the inter-layer method produces darker and less blurry images. This is probably due to the non-linear nature of the inter-layer contrast

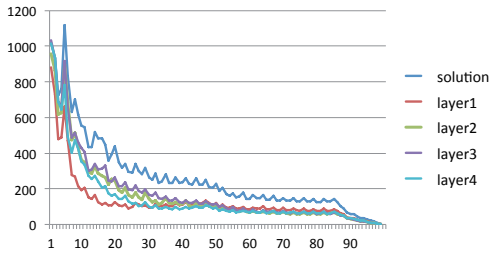

Figure 8. *Frequency analysis of the solution vector and accumulation images for scene1, which appeared in Fig. 6. The horizontal axis represent the ratio of spatial frequency to the maximum spatial frequency, and vertical axis represent the magnitude of frequency components.*

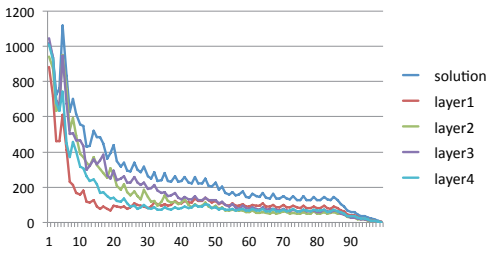

Figure 9. *Frequency analysis of the solution vector and modulated accumulation images for scene1. The modulation done here is the weighted sum of accumulation images denoted by equation (13). Compared to the signals in Fig. 8, the signals of layer2 and layer3 are slightly enhanced, while those of layer 1 and layer4 are slightly reduced.*

method, which chooses only bright-dark pairs of pixels. However, there are distinct differences between the bottom row images shown in Figs. 6 and 7. First, images in Fig. 6 exhibit higher contrast than those in Fig. 7. Secondly, some irregular patterns are visible in the dark, background areas in Figs. 6, while any such patterns are invisible in Fig. 7.

To summarize, the inter-layer method provides only a partial solution to estimating the magnitude of nonnegativity vector c. Developing a handy and more accurate estimation method is left as the task for future research.

Experiments

In the previous section, we have identified three key factors that affect the quality of the light field reproduced by a multilayer display. These three factors are symbolically referred to as $A^T b$, $(A^T A)^{-1}$, and c. While the c factor is very difficult to evaluate, the $A^T b$ and $(A^T A)^{-1}$ factors are not. Leaving aside the c factor, it has been suggested the computation of $A^T b$ is the main source of blur. In fact, the $(A^T A)^{-1}$ operator has been shown to not cut-off any high frequency signals. However, this does not necessarily mean that conversion from the accumulation to layer images, which $(A^T A)^{-1}$ is responsible for, is always done in the highest possible quality. The $(A^T A)^{-1}$ operator may change low frequency signals, which could introduce blur in the resultant layer images

Processing Accumulation Images

The equation (8) indicates that the solution vector x^* can be obtained by applying the $(A^T A)^{-1}$ operator to the combination of accumulation image (represented by $A^T b$) and error image (represented by c). Here we focus on how the accumulation image is decomposed, modulated, and merged into the layer image (i.e., the solution vector x^*).

Fig. 8 shows the frequency signals of layer image (referred to as "solution" in the figure) and accumulation images (referred to as "layer1" $\sim$ "layer4") corresponding to the scene1 that already appeared in Fig. 6. Note that "solution" keeps the highest position in all frequencies. As discussed previously, the $(A^T A)^{-1}$ operator decomposes the layer images into several frequency components, and apply the inverse of the filter matrix shown in equation (12). The filter matrix is frequency-dependent implying that the number of matrices are very large. However, if we simply average the elements of all inverse matrices, we will obtain a single matrix such as

$$\begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{pmatrix} = \begin{pmatrix} 1.23 & -0.34 & 0.05 & 0.06 \\ -0.34 & 1.68 & -0.40 & 0.05 \\ 0.05 & -0.40 & 1.68 & -0.34 \\ 0.07 & 0.05 & -0.34 & 1.23 \end{pmatrix} \begin{pmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \end{pmatrix} \quad (13)$$

where $(l_1, \cdots, l_4)$ represent "accumulation images" and $(m_1, \cdots, m_4)$ represent "modulated accumulation images". The values in the matrix are computed based on the size of the viewing ranges and distance between layers. Frequency signals of the modulated accumulation images obtained via (13) are shown in Fig. 9. Here we observe that "layer2" and "layer3" come near to "solution".

Fig. 10 compares 4 reconstruction methods of the light fields representing respectively 4 sample scenes. The reconstruction is done on the same 4-layer display, where the resolution of each layer is 640×360, and layer separation (i.e., the smallest distance between pairs of layers) is $0.1 \times$ width of the display. Part of the central views are extracted from the light fields, and shown in the figure. The "computational" reconstruction corresponds to the ordinary multilayer reconstruction, where the solution x of (3) is first computed, and then the resultant layer images are simply merged into the view. The other three methods uses the accumulation images rather than the layer images. In the "averaged" reconstruction, the accumulation images are simply averaged along each ray path. This method is essentially equivalent to the classical photo-stereo-synthesis method invented by Lumière [3]. The "depth selective" reconstruction is an improvement of the "averaged" reconstruction, where depth maps of the scenes are used as supplementary information. Using the depth maps, nearest two layers are identified for each ray path, and pixels' intensities of accumulation images belonging to the selected layers are weighted and fused. Finally, the "modulated" reconstruction is a variant of "depth selective" reconstruction, where modulated accumulation images, obtained by (13), are used instead of the original accumulation images.

Examining the PSNR (peak signal-to-noise ratio) scores in Fig. 10, we notice that the "computational" reconstruction always gives the best results, while the "averaged" reconstruction result in blurry images always in the worst quality. The "depth selective" and "modulated" reconstruction stands in the midway

Figure 10. *Comparison of 4 reconstruction methods of the light fields representing 4 sample scenes. The "computational" reconstruction refers to the ordinary light field reconstruction, which computes the solution of equation (3), and maps the solution to multiple layers. On the other hand, the "averaged" reconstruction puts the accumulation images on those layers. The other two methods make use of scenes' depth maps as supplementary information to determine the layers' weight before fusing the images. The weight is computed for each pixel and for each view direction. The "depth selective" reconstruction fuses accumulation images, while the "modulated" reconstruction fuses modulated accumulation images.*

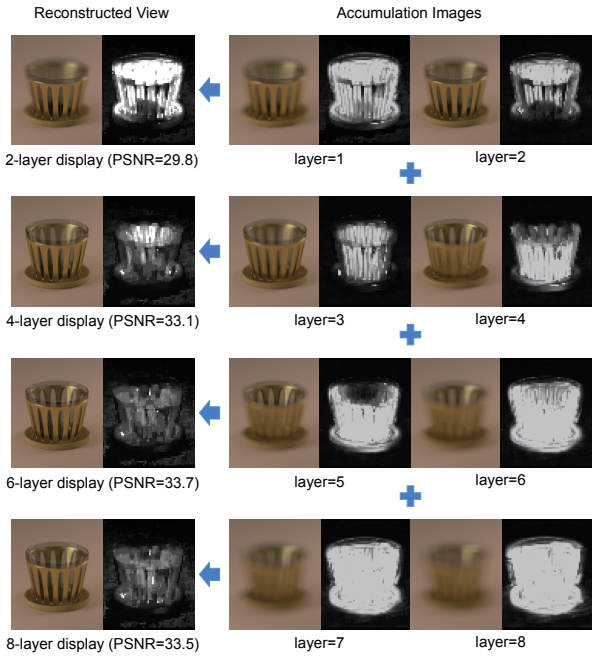

Figure 11. *Incremental addition of layers and its effect on the reconstructed light field. Starting from the 2-layer display, layers are incrementally added to the rear side of the display, yielding 4-layer, 6-layer, and 8-layer displays. These displays share the same set of accumulation images, where N-layer uses the first N images for reconstructing the light field. Each color image is associated with a black-and-white image, called "blur map", which represent the amount of blur visible in the color image. As seen in the PSNR scores, the 6-layer display marks the peak performance. Since the blur maps of the 7-th and and 8-th layers are almost the same as that of the 6-th layer, the PSNR scores of the 6-layer and 8-layer displays keeps the same level.*

between these two methods, where the "modulated" reconstruction tends to give slightly better results than the "depth selective"; however, their differences are almost negligible. It is worthwhile to mention that unless depth maps are available, neither the "depth selective" nor "modulated" reconstruction method will work. Despite such practical limitations, these methods are theoretically interesting since their performance is almost parallel to that of the ordinary "computational" reconstruction method. This observation seems to provide direct insight into the DOF analysis of multilayer displays

Accumulation Blur and Depth of Field

Suppose that the scene depth is constant, and every point in the scene lies on the i-th layer. In this case, the output of the "depth selective" reconstruction method will be the light field generated by putting the i-th accumulation image on the i the layer, and letting the other layers transparent. The quality of the i-th accumulation image determine the quality of the light field.

This simple observation leads to the following question: if the placement of layers is sufficiently dense and covers the entire depth range, will the ordinary multilayer reconstruction method perfectively recover the light field of the scene? The answer to this question will be negative unless all objects in the scene are non-specular, i.e., exhibiting Lambertian reflectance. In addition, it should also be recalled that we have not obtained any convincing model of the "error image" representing vector c. Nevertheless, it is interesting to investigate how the coverage of scene's depth range by multiple layers affects the quality of the reconstructed light field.

Fig. 11 reports on such an investigation, where the number of layers incrementally increases while keeping constant separation between neighboring layers. Since the layer separation is $0.1 \times$ width, the N-layer display covers the depth range of $[0, 0.1(N-1)\text{width}]$. In the left column of Fig. 11, the central views of the reconstructed light field are shown along with black-and-white images. These images exhibit the amount of blur in the corresponding views. The blur is computed for each small area within each view based on the comparison with the filtered view of the original light field (i.e., vector b). Filters used for this purpose are mean filters, and the filter's window size is adjusted to minimize the difference between the reconstructed and filtered views. The resultant black-and-white image is called "blur map" of the view. White pixels in a blur map indicate that the amount of blur is large at these pixels. The right column of Fig. 11 shows a series of accumulation images. The central view of the 2-layer display is computationally reconstructed based on the first two

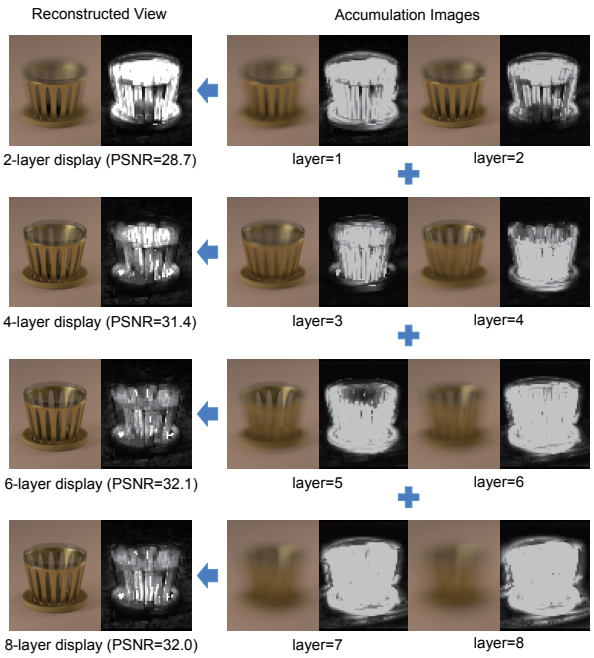

Figure 12. *Incremental addition of layers and its effect on the reconstructed light field. The difference between Fig. 11 and Fig. 12 lies in the range of viewing angles, where Fig. 11 sets the range parameter to* $[H \times V] = [-5.0^\circ, 5.0^\circ] \times [-2.5^\circ, 2.5^\circ]$*, while Fig. 12 sets it to* $[-9.0^\circ, 9.0^\circ] \times [-4.5^\circ, 4.5^\circ]$*. In terms of the PSNR scores, the 6-layer display marks the peak performance as in Fig. 11. However, the peak is lower than that of Fig. 11, which is accounted for by the amount of larger blur present in accumulation images.*

accumulation images, while the one of the 4-layer display is computed based on the first four images. Note that the first two accumulation images are shared by the 2-layer and 4-layer displays.

In the left column of Fig. 11, we observe the gradual quality improvement in the reconstructed view from the 2-layer to 6-layer display. In terms of PSNR, the largest score is marked by the 6-layer display. The score decreases slightly from the 6-layer to 8-layer display. This is probably due to the numerical errors introduced while solving the least squares problems. Looking at the accumulation images and associated blur maps in the right column of Fig. 11, we notice that accumulation images of layers 6 to 8 are all blurry. The blur maps of these layers are almost the same. This seems to imply that even if we acquire accumulation images of layers 7 and 8 in addition to 6, we will not obtain any extra information. These observations coincide with the algorithm of the "depth selective" reconstruction, although the "computational" method is actually used to reconstruct the light fields.

Fig. 12 shows similar results to those of Fig. 11. Almost all experimental conditions are the same except for the range of viewing angles, where Fig. 11 sets the range parameter to $[H \times V] = [-5.0^\circ, 5.0^\circ] \times [-2.5^\circ, 2.5^\circ]$, while Fig. 12 sets it to $[-9.0^\circ, 9.0^\circ] \times [-4.5^\circ, 4.5^\circ]$. We notice that the observations made for Fig. 11 also apply to Fig. 12. We also find that the PSNR scores in Fig. 12 are all lower than those in Fig. 11. This could be explained from the fact that any accumulation image in Fig. 12 is more blurry than the corresponding image in Fig. 11.

Concluding Remarks

In this paper, we presented a novel framework to analyze the depth of field (DOF) of multilayer displays. Three factors affecting the quality of the reconstructed light fields were identified from the computational viewpoint, and re-interpreted in the context of multilayer displays. Among the three factors, the $A^T b$ factor, which modeled the accumulation of light on layers, was found to be the the key in analyzing the DOF. With the assistance of the depth selective reconstruction method, which approximated the ordinary computational method usually adopted for multilayer displays, the DOF-like effects of a sample scene were analyzed. Although analytic formulae of DOF have not been established yet, the combination of the "light accumulation on layers" and "depth selective reconstruction" brings us an intuitive understanding on how the DOF of multilayer displays is determined.

Several issues raised in the paper are addressed only partially, and need to be investigated further. In particular, modeling and evaluation of the c factor is indispensable for making the DOF analysis more accurate. In addition, simple and analytic formulae of DOF will also be required in practical applications. Building reconfigurable displays dynamically adjustable to specific 3D scenes would be an interesting topic, where the DOF analysis will be utilized most effectively.

References

[1] M. Date, T. Hisaki, H. Takada, S. Suyama and K. Nakazawa, Luminance addition of a stack of multidomain liquid-crystal displays and capability for depth-fused three-dimensional display application, Applied Optics, 44, 6. (2005).

[2] D. Lanman, G. Wetzstein, M. Hirsch and R. Raskar, Depth of field analysis for multilayer automultiscopic displays, 9th International Symposium on Display Holography, 1. (2012).

[3] L. Lumière, Répresentation photographique d'un solide dans l'espace. Photo-stéreo-synthèse, Comptes rendus hebdomadaires des séances de l'académie des sciences, 171. (1920).

[4] H. Gotoda, A multilayer liquid crystal display for autostereoscopic 3D viewing, Proc. SPIE, 7524-24. (2010).

[5] R. Ng, Fourier slice photography, ACM Transactions on Graphcis, 24, 3. (2005).

[6] 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 Transactions on Graphics, 30, 4. (2011).

[7] N. Zheng, K. Hayami, and J.-F. Yin, Modulus-Type Inner Outer Iteration Methods for Nonnegative Constrained Least Squares Problems, SIAM Journal on Matrix Analysis and Applications, 37, 3. (2016).

[8] M. Zwicker, W. Matusik, F. Durand and H. Pfister, Antialiasing for automultiscopic 3D displays, Eurographics Symposium on Rendering. (2006).

Author Biography

Hironobu Gotoda received his BSc (1989), MSc (1991), and PhD in information science (1994) all from the University of Tokyo. He then joined the Research Division of National Center for Science Information Systems, in Tokyo, Japan. He is currently an associate professor at the National Institute of Informatics also in Tokyo. His research interest includes 3D object modeling, search, rendering, and display. He is a member of ACM, IEEE, IPSJ, IEICE, OSA, and IS&T.