Deriving Synthetic Filter Bank for Perfect Reconstruction of Light Field from Its Focal Stack

Akira Kubota*, Asami Ito*, and Kazuya Kodama†
* School of Science and Engineering, Chuo University, Tokyo, Japan
E-mail:kubota@elect.chuo-u.ac.jp, a14.gy46@g.chuo-u.ac.jp
Tel/Fax: +81-3-3817-1847
† National Institute of Informatics, Tokyo, Japan
E-mail: kazuya@nii.ac.jp

***Abstract*—For 3D Lambertian scenes with no occlusion, we show that there exist a filter bank such that it can perfectly reconstruct the light field from the focal stack, if the focal stack is captured through "Cauchy-type" aperture mask. All the derived filters are spatially invariant; therefore the presented filter bank method requires no depth estimation. Simulation results using a synthetic scene show that the filter bank reconstructs the light field with significant high quality.**

I. Introduction

Light filed is the distribution of light rays in a 3D free space and recorded as a 4D function, which is often represented as a 2D array of 2D images [1]. Unlike the conventional images in which the light rays are recorded at 2D pixel positions, the recorded light field has additional 2D angular information at each pixel positon; this enables us various image manipulations such as novel view generation, refocusing and extending depth-of-field (DOF).

To acquire 4D light field by a fixed single sensor, so called Plenoptic cameras equipped with micro lens array in front of the imaging plane have been used (e.g., [2]). In these systems, each lens captures 2D spatial information and the pixels under the lens capture 2D angular information; therefore the sensor pixel resolution determines the resolution of the 4D light field. This means it has a resolution trade-off between the spatial and the angular dimensions. The lower resolution of either dimension may degrade the quality of the image manipulations. For instance, in light field rendering [3], low angular resolution causes ghosting artifacts in the novel views [4].

In order to improve the resolution of the light field, a straightforward solution is to move a single camera. In this way, a light field with higher resolution can be obtained; the angular resolution is the same with the sensor pixel resolution and the spatial resolution is determined by the distance between the moving camera positions. It, however, requires a precise control of the camera position as well as an accurate camera calibration. Different from this direct capturing, an alternative way is to recover the higher resolution light field from a set of multiple images at different focal depths (called focal stack) captured by a single fixed camera.

As the early work for the light field recovery from the focal stack, Aizawa et al. presented an iterative recovery method [5] for Lambertian scenes that consist of a few layers with no occlusion. Modeling both the focal stack and the desired novel view (the novel sub-aperture image or pinhole view) as linear combinations of the unknown layer textures at different depths, they derived an equation that holds between them and solved it iteratively. This method can appropriately shift the position of object regions according to their depths without depth estimation. However, the equation cannot be completely solved; in fact, at the lower frequency, the solution (i.e., the novel view) tends to diverse much faster. So does the extended version recently presented in [6] to the scene with many layers.

In the present paper, we theoretically show that the equation can be solved if we use the optimally designed aperture mask [7], called Cauchy-type aperture, when capturing the focal stack. We then derive a set of reconstruction filters (i.e., filter bank) to be applied to images of the focal stack such that the light field can be reconstructed perfectly for Lambertian scenes with no occlusions. All the derived filters are spatially invariant; thus our filter bank method needs no depth estimation.

There are alternative approaches [8]–[12] based on computer tomography. Since the focal stack can be modeled as a set of projections of the light field [13], the light field can be roughly estimated to be the back-projected focal stack. Some methods [8]–[10] presented an optimum deconvolution filter to suppress blurring artifacts caused in the obtained novel view. Although these methods also do not involve depth estimation, they work well only if the range of the focal depth is much wider than that of the actual scene depth. In contrast, our filter bank method allows the focal depth range to be the same with the scene depth range.

II. Problem description

We represent a 4D Light field as a set of multiple images (2D array of multiview images) $f^{(s,t)}(x,y)$ captured from a view point (s,t) on st plane (called camera or aperture plane) [14]. The coordinate (x,y) denotes a pixel position of the imaging plane of each camera. Let the distance between two planes denote d.

The focal stack as an input for the light field reconstruction is captured from the origin of the st plane by changing the focal depth. Letting $g_n^{(0,0)}(x,y)$ be the image captured when

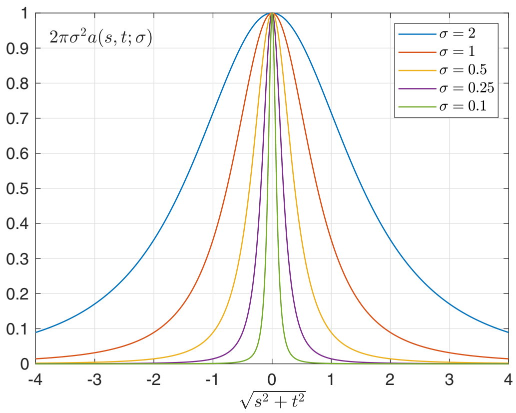

Fig. 1. Characteristics of Cauchy-type pupil function (The maximum amplitude was normalized to 1).

the focal depth is z_n $(n = 1, 2, ..., N)$, it is generated from the light field $f^{(s,t)}(x, y)$ [14]:

$$g_n^{(0,0)}(x,y) = \iint a(s,t;\sigma) f^{(s,t)}(x-\alpha_n s, y-\alpha_n t)\, ds\, dt \quad (1)$$

where $a(s, t; \sigma)$ is the aperture mask on the st plane with its scaling factor σ which determines the amount of blur and $\alpha_n = d/z_n$. The aperture mask $a(s, t; \sigma)$ satisfies $\iint a(s, t; \sigma)\, ds\, dt = 1$ for arbitrary σ.

For the aperture mask, we adopt the pupil function

$$a(s,t;\sigma) = \frac{1}{2\pi\sigma^2}\left[\left(\frac{s}{\sigma}\right)^2 + \left(\frac{t}{\sigma}\right)^2 + 1\right]^{-3/2}, \quad (2)$$

called Cauchy-type pupil function (see the characteristics in Fig.1). This function has been designed for all-in-focus image generation from the focal stack in our recent work [7]. Its 2D Fourier transform (i.e., its optical transfer function) can be simply represented by a rotationally symmetric exponential function

$$A(\xi,\eta;\sigma) = \exp\{-2\pi\sigma\sqrt{\xi^2+\eta^2}\}, \quad (3)$$

where ξ and η denote the horizontal and the vertical frequencies, respectively.

Our goal is to derive a filter bank $\{k_n^{(s,t)}(x, y)\}$ that perfectly reconstructs the light field $f^{(s,t)}(x, y)$ directly from the acquired focal stack $\{g_n^{(0,0)}(x, y)\}$ by

$$f^{(s,t)}(x,y) = \sum_{n=1}^{N} k_n^{(s,t)}(x,y) * g_n^{(0,0)}(x,y), \quad (4)$$

where $*$ represents 2D convolution. Note that all the derived filters are linear and spatially invariant; hence the proposed filtering method requires no depth estimation and is independent of the scene geometry.

III. Filter Bank for Light Field Reconstruction

A. Image formation model using layered scene representation

In this subsection, we model the focal stack and the light field based on the layered scene representation used in the previous work [5]–[7].

We represent a Lambertian scene with no occlusion as a set of multiple layers at different depth z_n from the camera/aperture plane. Consider only m'th layer and let $f_m^{(s,t)}(x, y)$ denote all the light rays emerging from the surface of m'th layer. Since the layer surface is Lambertian, we have

$$f_m^{(s,t)}(x,y) = f_m^{(0,0)}(x+\alpha_m s, y+\alpha_m t), \quad (5)$$

where again $\alpha_m = d/z_m$. Substituting this into Eq. (1), we obtain

$$\begin{aligned} g_n^{(0,0)}&(x,y) \\ &= \iint a(s,t;\sigma) f_m^{(0,0)}(x-(\alpha_n-\alpha_m)s, y-(\alpha_n-\alpha_m)t)\, ds\, dt \\ &= \iint a(s,t;\sigma|\alpha_n-\alpha_m|) f_m^{(0,0)}(x-s, y-t)\, ds\, dt \\ &= a(x,y;\sigma|\alpha_n-\alpha_m|) * f_m^{(0,0)}(x,y). \end{aligned} \quad (6)$$

Because we assume the scene has no occlusion (i.e., all the layers do not occlude each other), we can model the focal stack as the sum of the above obtained results for all the layers as

$$g_n^{(0,0)}(x,y) = \sum_{m=1}^{N} a(x,y;\sigma|\alpha_n-\alpha_m|) * f_m^{(0,0)}(x,y). \quad (7)$$

The function $a(x, y; \sigma|\alpha_n - \alpha_m|)$ is known as a point spread function (PSF), which is a scaled version of the aperture mask used. Note that the PSF becomes Dirac's delta function $\delta(x, y)$ when $m = n$.

We can also model the light field $f^{(s,t)}(x, y)$ using $f_m^{(0,0)}(x, y)$. From now on, assuming $t = 0$, we only consider the horizontal view point without loss of generality. The light field $f^{(s,0)}(x, y)$ can be expressed by

$$f^{(s,0)}(x,y) = \sum_{m=1}^{N} f_m^{(0,0)}(x-\alpha_m s, y). \quad (8)$$

B. Image formation model in the Fourier domain

We represent the above image formation models in the Fourier domain. Taking Fourier transform of Eq. (7), we obtain the following matrix-vector formula:

$$\boldsymbol{g} = H\boldsymbol{f}. \quad (9)$$

The two vectors are defined as

$$\boldsymbol{g} = \begin{pmatrix} G_1^{(0,0)}(\xi,\eta) \\ G_2^{(0,0)}(\xi,\eta) \\ \vdots \\ G_N^{(0,0)}(\xi,\eta) \end{pmatrix} \quad \text{and} \quad \boldsymbol{f} = \begin{pmatrix} F_1^{(0,0)}(\xi,\eta) \\ F_2^{(0,0)}(\xi,\eta) \\ \vdots \\ F_N^{(0,0)}(\xi,\eta) \end{pmatrix}, \quad (10)$$

where $G_n^{(0,0)}(\xi, \eta)$ and $F_n^{(0,0)}(\xi, \eta)$ represent the 2D Fourier transforms of $g_n^{(0,0)}(x, y)$ and $f_n^{(0,0)}(x, y)$, respectively. If we change the focal depth z_n such that α_n changes in equal interval of p, the matrix H can be given by the following symmetric Toeplitz form (especially called Kac-Murdock-Szegö (KMS) matrix [15]):

$$H = \begin{pmatrix} 1 & A & A^2 & \dots & A^{N-1} \\ A & 1 & A & \dots & A^{N-2} \\ A^2 & A & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & A \\ A^{N-1} & \dots & \dots & A & 1 \end{pmatrix}, \tag{11}$$

where the function A denotes the Fourier transform of PSF $a(x, y; \sigma p)$ as

$$A(\xi, \eta; \sigma p) = \exp\{-2\pi\sigma p\sqrt{\xi^2 + \eta^2}\}. \tag{12}$$

Similarly, taking the Fourier transform of Eq. (8) gives

$$F^{(s,0)}(\xi, \eta) = \boldsymbol{w}^\top \boldsymbol{f}, \tag{13}$$

where $\boldsymbol{w}$ is defined as

$$\boldsymbol{w} = \begin{pmatrix} \exp(j2\pi\xi\alpha_1 s) \\ \exp(j2\pi\xi\alpha_2 s) \\ \vdots \\ \exp(j2\pi\xi\alpha_N s) \end{pmatrix}$$

and $^\top$ denotes transpose operation.

C. Deriving filter bank

Using the image formation models in the Fourier domain, we derive the filter bank that reconstructs the light field directly from the focal stack.

At the frequency where H^{-1} exists, eliminating the unknown $\boldsymbol{f}$ from equations (9) and (13) yields

$$F^{(s,0)}(\xi, \eta) = \boldsymbol{w}^\top H^{-1} \boldsymbol{g}. \tag{14}$$

Each element of the row vector $\boldsymbol{w}^\top H^{-1}$ specifies the frequency characteristic of the reconstruction filter $k_n^{(s,0)}(x, y)$ to the image $g_n^{(0,0)}(x, y)$; therefore in the Fourier domain we can obtain the filter bank as a form of

$$\left(K_1^{(s,0)}(\xi, \eta), K_2^{(s,0)}(\xi, \eta), \dots, K_N^{(s,0)}(\xi, \eta)\right) = \boldsymbol{w}^\top H^{-1}. \tag{15}$$

In all the frequency except the direct current (DC), since $A(\xi, \eta; \sigma p) \neq 1$, the matrix H is invertible and its inverse is given as

$$H^{-1} = \frac{1}{1 - A^2} \begin{pmatrix} 1 & -A & & & \\ -A & 1 + A^2 & -A & & \\ & \ddots & \ddots & \ddots & \\ & & -A & 1 + A^2 & -A \\ & & & -A & 1 \end{pmatrix}. \tag{16}$$

Consequecncy the filter bank is derived based on Eq. (15) as follows:

$$K_1^{(s,0)}(\xi, \eta) = \frac{1 - A\exp(-j2\pi\xi p s)}{1 - A^2} \exp(j2\pi\xi\alpha_1 s) \tag{17}$$

$$K_n^{(s,0)}(\xi, \eta) = \frac{A^2 - 2A\cos(2\pi\xi p s) + 1}{1 - A^2} \exp(j2\pi\xi\alpha_n s) \quad (n = 2, \dots, N-1) \tag{18}$$

$$K_N^{(s,0)}(\xi, \eta) = \frac{1 - A\exp(j2\pi\xi p s)}{1 - A^2} \exp(j2\pi\xi\alpha_N s) \tag{19}$$

At the DC where H^{-1} does not exist, Eq. (9) cannot be solved for $\boldsymbol{f}$. Despite of this, we can identify the DC component of the filters by taking the limit to zero of the above obtained results. By applying l'Hospital's theorem, we can obtain the limit value at the DC as follows:

$$\lim_{\zeta \to 0} K_1^{(s,0)}(\zeta, \theta) = \frac{1}{2} - j\frac{s\cos\theta}{2\sigma} \tag{20}$$

$$\lim_{\zeta \to 0} K_n^{(s,0)}(\zeta, \theta) = 0, \quad (n = 2, \dots, N-1) \tag{21}$$

$$\lim_{\zeta \to 0} K_N^{(s,0)}(\zeta, \theta) = \frac{1}{2} + j\frac{s\cos\theta}{2\sigma} \tag{22}$$

Here $\zeta = \sqrt{\xi^2 + \eta^2}$ and $\theta = \arctan(\eta/\xi)$.

These results show that all the filters do not diverse and are stable [1]. The imaginary parts of the DC components of $K_1^{(s,0)}$ and $K_N^{(s,0)}$ depend on the directional frequency θ and thus are not determined uniquely; however it is acceptable that they are set to be zero because we can assume that these filters in the spatial domain have real values.

IV. Simulation

Using synthetically generated focal stack data, we reconstructed the light field by the proposed filter bank and evaluated the reconstruction accuracy in peak-signal to noise ratio (PSNR).

Assuming a test scene that consists of three planes, on each of which a different part of image *Baboon* is mapped, we synthetically generated nine differently focused images (256 × 256 resolution) as the focal stack based on the image formation model (9). The focal stack generated for the case of $\sigma p = 1$ [pixel] is shown in the left column in Fig. 2. The images g_1, g_5 and g_9 are focused on the nearest, the middle and the farthest depth layer, respectively. For instance, the farthest layer in g_1 is blurred with $\sigma p = 8$ [pixels] and both nearest and farthest layers in g_5 is blurred with $\sigma p = 4$ [pixels].

The reconstructed light field from the focal stack is shown in the middle column in Fig. 2. They are the novel views from different view point s. In this simulation setting, according to the view point s, the maximum disparity $\alpha_1 s$ varies from -64 to 64 [pixesl] and the minimum disparity $\alpha_9 s$ varies from -32 to 32 [pixesl]. As can be seen, compared with the grand truth in the left column, all the novel views are reconstructed with sufficient quality and the three planes in the views are properly shifted without visible artifacts.

[1] We found that the filters diverse and are unstable for normally used Gaussian or pillbox apertures.

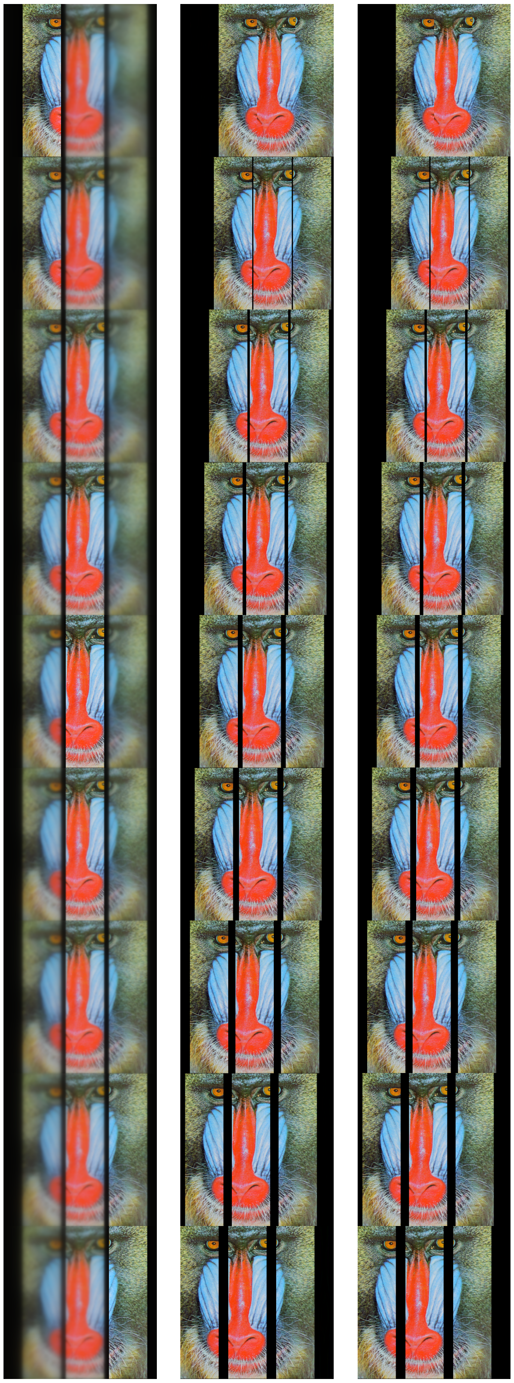

Fig. 2. Simulation results. The left column: Synthetically generated focal stack (top to bottom: g_1 to g_9). The middle column: Reconstructed light field from the focal stack (top to bottom: horizontally shifted novel views from left to right) The right column: The grand truth of the light field

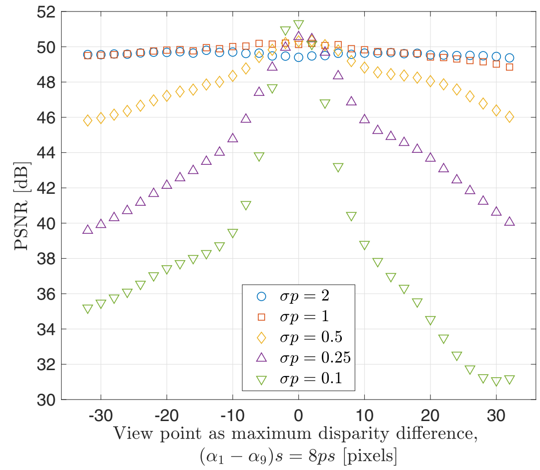

Fig. 3. PSNR of the reconstructed light field.

The PSNR of the reconstructed light field for various blur amount σp is shown in Fig. 3. For large blur setting ($\sigma p \geq 0.5$), the PSNR were significantly high over all the range of horizontal view point. This shows that the presented filter bank can accurately reconstruct the light field. On the other hand, for very small blur setting ($\sigma p = 0.1$), the PSNR become lower when the view point was farther from the origin. This is because the filter characteristics tend to have larger value, leading to the increase of the quantization errors in generating the focal stack.

V. Conclusion

In this paper, we derived the filter bank that allows perfect reconstruction of the high-resolution light field directly from the focal stack that is captured through Cauchy-type aperture mask. The simulation results showed that high quality reconstruction is possible by the derived filter bank.

In future, we conduct experiments using real captured focal stack to evaluate the reconstruction quality.

REFERENCES

[1] Gaochang Wu, Belen Masia, Adrian Jarabo, Yuchen Zhang, Liangyong Wang, Qionghai Dai, Tianyou Chai, and Yebin Liu, "Light field image processing: An overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, pp. 926–954, 2017.

[2] M. Levoy R. Ng, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Stanford Tech Report CTSR 2005-02*, 2005.

[3] M. Levoy and P. Hanrahan, "Light field rendering," in *proc. 23rd Annu. Conf. Comput. Graph. Interactive Techn.*, 1996, pp. 31–42.

[4] J.-X. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum, "Plenoptic sampling," in *Proc. 27th Annu. Conf. Comput. Graph. Interactive Techn.*, 2000, pp. 307–318.

[5] K. Aizawa, K. Kodama, and A. Kubota, "Producing object-based special effects by fusing multiple differently focused images," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 10, pp. 323–330, Mar. 2000.

[6] J. R. Alonso, A. Fernandez, and J. A. Ferrari, "Reconstruction of perspective shifts and refocusing of a three-dimensional scene from a multi-focus image stack," *Applied Optics*, vol. 55, pp. 2380–2386, Mar. 2016.

[7] A. Kubota, "Synthesis filter bank and pupil function for perfect reconstruction of all-in-focus image from focal stack," in *proc. of SPIE Int. Conf. on Quality Control by Artificial Vision*, 2017.

[8] F. Perez, A. Perez, M. Rodriguez, and E. Magdaleno, "Lightfield recovery from its focal stack," *Math Imaging Vis.*, vol. 56, pp. 573–590, 2016.

[9] K. Kodama and A. Kubota, "Efficient reconstruction of all-in-focus images through shifted pinholes from multi-focus images for dense light field synthesis and rendering," *IEEE Trans. on Image Processing*, vol. 22, pp. 4407–4421, 2013.

[10] A. Levin and F. Durand, "Linear view synthesis using a dimensionality gap light field prior," in *proc. IEEE Conf. Comput, Vis. Pattern Recognition*, 2010, pp. 1831–1838.

[11] J. Park et al., "Light ray field capture using focal plane sweeping and its optical reconstruction using 3d displays," *Opt. Exp.*, vol. 22, pp. 25444–25454, 2014.

[12] J. M. Trujillo-Sevilla et. al, "Restoring integral images from focal stacks using compressed sensing techniques," *Display Technology*, vol. 12, pp. 701–706, 2016.

[13] R. Ng, "Fourier slice photography," *ACM Trans. Graphics*, vol. 24, pp. 735–744, 2005.

[14] A. Isaksen, L. McMillan, and S. J. Gortler, "Dynamically reparameterized light fields," in *Proc. Comput. Graph. Interactive Techn.*, 2000, pp. 279–306.

[15] U. Grenander and G. Szego, *Toeplitz forms and their applications*, Uni. Calif. Press, New York, 1958.