A FOCUS MEASURE FOR LIGHT FIELD RENDERING

Keita Takahashi[†]*, Akira Kubota*[‡]*, Takeshi Naemura*[†]

[†]The University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
{*keita,naemura*}*@hc.t.u-tokyo.ac.jp*

[‡]Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA, 15213, USA
akira@andrew.cmu.edu

ABSTRACT

Light field rendering is a fundamental method for synthesizing free-viewpoint images from a set of multi-viewpoint images. In the simplest case, the scene structure is approximated by a simple plane: a focal plane. This approximation leads to focus-like effects on synthetic images where the *focused* depth is determined by the focal plane. A serious problem is that the range of the *focused* depth is too small in most practical cases. In this paper, we propose a *focus* measure that is specialized for synthetic images by light field rendering. When a set of differently-*focused* images is generated at a given viewpoint, the proposed *focus* measure enables us to obtain a depth map and an all in-*focus* image. Our approach has some remarkable differences from other related techniques, such as depth-from-stereo and depth-from-focus methods. Experimental results show that the proposed method effectively enhances PSNR of the final synthetic images.

1. INTRODUCTION

In recent years, image based rendering (IBR) has received much attentions as a powerful approach for synthesizing free-viewpoint images with photorealistic quality [1]. Light field rendering (LFR) [2] is a fundamental method of IBR. It produces novel images from a light field: a set of multi-viewpoint images captured by densely-aligned cameras. No/little geometric information is required in this method. In fact, in the simplest case, the shape of the entire scene is approximated by a plane, which is called a focal plane. This approximation causes a focus-like effect on synthetic images where the *focused* depth is determined by the focal plane [3, 4, 5]. Objects near the focal plane appear clearly and sharply, while objects apart from the focal plane are with blurring and ghosting. The range of the *focused* depth is mainly determined by the density of input images and camera resolutions [6]. The problem is that the range is too small in most practical cases.

Assume that several differently-*focused* images are generated by LFR at a certain viewpoint. This means we synthesize multiple images with changing the depth of the focal plane. As Isaksen et al. [3] have described, integration of in-*focus* pixels into one image yields an all in-*focus* image. This process needs to determine which *focused* depth to be used for each pixel. We have proposed a *focus* measure to detect the optimal *focused* depth automatically [7]. It is specialized for synthetic images by LFR, since the *focus* in LFR has a different nature from that of physical cameras. A fast implementation method of our algorithm on programmable graphics hardwares has also been developed [8]. In this paper, we discuss the detail of the proposed *focus* measure, and report some quantitative evaluations of the final synthetic images to show the effectiveness of our approach.

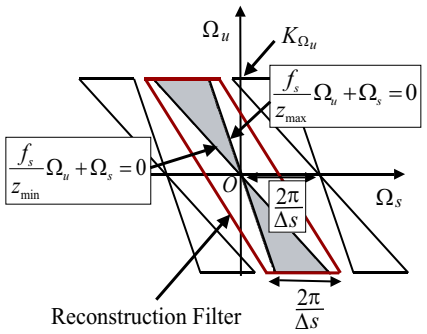


Fig. 1. Signal analysis of light fields in the frequency domain.

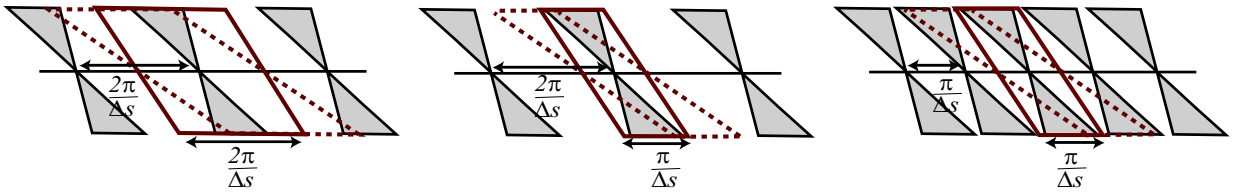
2. BACKGROUND

Our method firstly searches the optimal depth for each pixel by the *focus* measurement. This is equivalent to acquiring a depth map. Two groups of depth estimation methods are closely related with our method. The one is to use multi-viewpoint images, including stereo and silhouette methods [9]. The other is to use multi-focused images, called depth-from-focus method [10]. Our method estimates depth using multi-*focused* synthetic images, and those images are generated from multi-viewpoint images by LFR.

There are two remarkable points in our approach. Firstly, our method generates depth maps directly at arbitrary viewpoints. Depth maps are also available by depth-from-stereo and depth-from-focus methods. However, the viewpoint is limited to the position where the images were actually captured. On the other hand, our method uses synthetic images, the viewpoint of which can be determined arbitrarily. Therefore, depth maps are generated directly for the desired viewpoint without any visibility checking or explicit 3D reconstruction.

The second point is that we adopt a novel *focus* measure. In general, high-frequency energy is regarded as a sufficient focus measure for depth estimation, since in-focus regions are supposed to have the more high-frequency energy [10]. However, it dose not hold true in synthetic images by LFR. In the *defocused* regions, not only blurring but also ghosting artifacts, which have much high frequency energy, are caused. Isaksen et al. [3, 4] proposed to use aperture filtering with a very large radius in order to reduce ghosting artifacts, and to apply the conventional focus measure. But sufficient results were not shown in those articles. On the other hand, we adopt an absolutely different *focus* measure specialized for LFR, which is based on the frequency domain analysis [11].

0-7803-8554-3/04/$20.00 ©2004 IEEE.



(a) Normal filter. (b) Wide-aperture filter. (c) Camera-skipped filter.

Fig. 2. Shape of reconstruction filters: (a) normal filter, (b) wide aperture filter, and (c) camera skipped filter. Solid parallelograms show a case when the target region is in *focus* ($z_0 = z_{opt}$). Dashed parallelograms show a case when the target region is out of *focus* ($z_0 \neq z_{opt}$).

(a) Normal filter. (b) Wide-aperture filter. (c) Camera-skipped filter.

Fig. 3. Synthesized images by different reconstruction filters. *In-focus* region (further building) looks similar regardless of the filters, while *defocused* regions have different artifacts from each other.

3. PRINCIPLE OF OUR APPROACH

3.1. Frequency Domain Analysis of Light Fields

Consider that cameras to acquire input images are placed on a two-dimensional (2-D) grid. A light field is parameterized by (s, t, u, v): (s, t) denotes positions of cameras and (u, v) denotes positions of pixels on each image. For simplicity, we discuss a 2-D subspace (s, u) of the original 4-D light field.

Chai et al. [11] have assumed that non-Lambertian reflections and occlusions could be negligible, and analyzed light fields in the frequency domain (Ω_s, Ω_u). They have revealed that the signal spectrum of a light field is restricted to the shadowed area in Fig. 1. Here, the depth of the scene (distance from the (s, t) plane) is represented as $z_{\min} \leq Z \leq z_{\max}$. f_s denotes the focal length of cameras, and K_{Ω_u} denotes the maximum frequency determined by the complexity of the scene textures or image resolutions. Since the light field is sampled discretely, replications of the original spectrum appear in constant intervals as shown in Fig.1. When the interval between cameras is Δs, the repeated cycle in the frequency domain becomes $2\pi/\Delta s$.

They have also discussed the image synthesis process, in which the light-ray data are interpolated and re-sampled. When the original spectrum does not overlap with neighboring replications (it is the anti-aliasing condition), the ideal box filter shown in Fig.1 would reconstruct the continuous signal without aliasing artifacts. Here, the spectrum is band-limited inside the box region. The slope of the box corresponds to the depth of the focal plane, which is represented as z_0. The filter is optimized when z_0 is equal to z_{opt} that is defined as

$$\frac{1}{z_{opt}} = \frac{1}{2}\left(\frac{1}{z_{\max}} + \frac{1}{z_{\min}}\right). \tag{1}$$

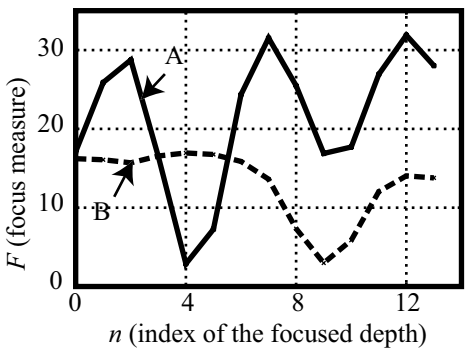


Fig. 4. Examples of the *focus* measure values.

In [11], $z_{\max}$ and $z_{\min}$ are defined for the entire scene. In most practical cases, the depth variation is too large to satisfy the anti-aliasing condition. However, we consider a relatively small region of the scene, and re-define $z_{\max}$ and $z_{\min}$ as local values for the region. Assume that the depth variation of the region $(1/z_{\min} - 1/z_{\max})$ is small enough to satisfy:

$$K_{\Omega_u} f_s \cdot \left(\frac{1}{z_{\min}} - \frac{1}{z_{\max}}\right) \leq \frac{1}{2} \cdot \frac{2\pi}{\Delta s}. \tag{2}$$

3.2. How to Measure *Focus* in LFR

Our proposal is to use the difference between some kinds of reconstruction filters for the *focus* measurement in LFR. We describe how it works based on the frequency domain analysis.

We adopt three kinds of reconstruction filters shown in Fig. 2(a) – (c). (a) The normal filter [11] is the most fundamental one. The width is $2\pi/\Delta s$. (b) The wide-aperture filter [3, 4, 5] is nar-

Fig. 5. All in-*focus* images are generated for arbitrary viewpoints using the proposed *focus* measure.

rower than $2\pi/\Delta s$. Synthetic images by this filter look as if they were taken by a wide-aperture camera. The width is set to $\pi/\Delta s$ in our method. (c) The camera-skipped filter firstly skips the input images alternately and then applies the normal filter to the sub-sampled light field. The repeating interval of the spectra and the width of filter are $\pi/\Delta s$.

When the region is *in focus* ($z_0 = z_{opt}$), every filter lets through the same amount of the frequency components as shown by the solid parallelograms in Fig. 2 (a) – (c). In this case, the same synthetic result is obtained theoretically regardless which filter is used. However, when the region is *out of focus* ($z_0 \neq z_{opt}$), different filters would produce different results. The dashed parallelograms in Fig. 2 show the shapes of filters when $z_0 < z_{opt}$. The frequency components passed through by the filters are obviously different from each other. Here, the leakage of the original spectrum causes blurring, and interfusion of the replicated spectra causes ghosting artifacts on the synthetic images.

The above discussion explains the synthetic results of Fig. 3 well. We use 81 (9×9) static images from "the multiview image database courtesy of University of Tsukuba, Japan" as the input. Those synthetic images have the same viewpoint and the same *focused* depth: the farthest building, which appears clear and sharp, is in *focus*. But different reconstruction filters are used to interpolate light-ray data. Notice that the *focused* regions are almost the same in all the images regardless which filter is used. While, the *defocused* regions show the differences in characteristics between the reconstruction filters. Therefore, we use subtractions between those images for the *focus* measurement. This is based on the fact that if a region is *in focus*, the absolute of the subtraction is relatively small (theoretically zero).

4. ALGORITHM

Depth candidates z_n ($n = 0, 1 ..., N-1$) where the focal plane is placed should be determined in advance. We adopt the following equation which divides the disparity space in a constant interval:

$$\frac{1}{z_n} = \frac{n}{N-1}\left(\frac{1}{Z_{\min}} - \frac{1}{Z_{\max}}\right) + \frac{1}{Z_{\max}}. \tag{3}$$

where $Z_{\max}$ and $Z_{\min}$ denote the maximum/minimum depth of the entire scene.

Assume that a set of differently-*focused* images is generated by LFR at a certain viewpoint. $I_k^{(n)}(x, y)$ represents a synthetic image with the filter h_k and the *focused* depth z_n. (x, y) denotes positions of pixels. In our method, h_1, h_2 and h_3 are assigned to the normal filter, the wide-aperture filter and the camera-skipped filter respectively. Therefore, we generate $3N$ images in total by the LFR method.

Our proposal is to use differences between reconstruction filters, instead of high frequency energy, for the *focus* measurement. Firstly, we take subtractions between images that are generated by different reconstruction filters at the same z_n as follows:

$$Sub^{(n)}(x, y) = \sum_{k<l} a_{k,l} |I_k^{(n)}(x, y) - I_l^{(n)}(x, y)|. \tag{4}$$

where $a_{k,l}$ denotes a weighting coefficient. Then, $Sub^{(n)}(x, y)$ are summed in a window region, the size of which is $(2M + 1)^2$:

$$F^{(n)}(x, y) = \sum_{-M \leq i,j \leq M} Sub^{(n)}(x + i, y + j). \tag{5}$$

This is defined as the *focus* measure at a point (x, y).

For each pixel (x, y), the index of the optimal depth is given by $n_o(x, y)$ that yields the minimum of $F^{(n)}(x, y)$:

$$n_o(x, y) = \arg \min_{0 \leq n \leq N-1} F^{(n)}(x, y). \tag{6}$$

Shown in Fig. 4 are examples of the *focus* measure values for different two pixels which are plotted in Fig. 3(a). In this case, we set $N = 14$, $a_{1,2} = 1, a_{1,3} = 1, a_{2,3} = 0$ in Equation (4), and $M = 7$ in Equation (5). Like these examples, the minimum of $F^{(n)}(x, y)$, which determines the optimal depth, is given definitely in most cases. Acquiring $n_o(x, y)$ for all pixels, we obtain the depth map at the given viewpoint.

Finally, the all *in-focus* image $I(x, y)$ is generated by integration of *in-focus* pixels into one image, as follows:

$$I(x, y) = I_1^{(n_o(x,y))}(x, y). \tag{7}$$

Shown in Fig. 5 are some final synthetic images at different viewpoints. Those images are all in-*focus*, since all objects are clear and sharp compared to Fig. 3. Our algorithm seems to be successful to synthesize visually-acceptable images at arbitrary viewpoints, but quantitative evaluations are also desired to show its effectiveness.

5. EXPERIMENTS

We have conducted some experiments for quantitative evaluations of our approach. Firstly, we generate a CG scene, and capture a set of input images at 81 (9×9) viewpoints in a constant interval. Then, we produce depth maps and all in-*focus* images at new viewpoints following the procedure described in Section 4. Finally, we compare the final all in-*focus* images with the original CG scene

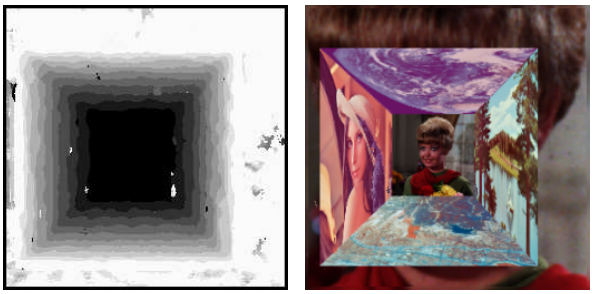
Table 1. Conditions of experiments with a CG scene.

CG scene	$10 \leq Z \leq 20$
input images	$Z = 0$ interval: 0.9, number: 9×9 pixels: 256×256, FOV: 27°
synthetic images	$Z = -20$ pixels: 256×256, FOV: 12°

by PSNR. Detail configurations are shown in Table 1. Here, Z-axis is determined to be orthogonal to the plane where cameras are aligned to capture input images. We set $M = 3$ in Equation (5).

Shown in Fig. 6 are examples of (a) depth maps, where the pixel nearer to us has the higher luminance, and (b) the final synthetic images. Figure 7 shows the relation between the number of depth candidates N and PSNR of synthetic images. In this graph, averages and the standard deviations which are calculated for 81 different viewpoints are plotted. The viewpoints are distributed constantly in a 0.8×0.8 square region at $Z = -20$. The solid line represents the case when $a_{1,2} = 1$, $a_{1,3} = 1$, $a_{2,3} = 0$ in Equation (4). In this case, three filters are used for the *focus* measurement. On the other hand, the dashed line is the case when $a_{1,2} = 1$, $a_{1,3} = 0$, $a_{2,3} = 0$: only two filters (the normal filter and the wide-aperture filter) are taken into consideration. When $N = 1$, synthetic results are the same as the conventional LFR method with the focal plane at the optimal depth defined by Equation (1).

From those results, we have found that PSNR increases with an increase of N, and it is gradually saturated. This is a favorable fact to show the effectiveness of our approach. Fig. 7 also shows that the combination of three kinds of filters yields higher synthetic quality compared to that of two kinds of filters.



(a) Depth map (b) Final synthetic image

Fig. 6. These images are (a) a depth map and (b) an all in-*focus* image at a novel viewpoint when $N = 11$.

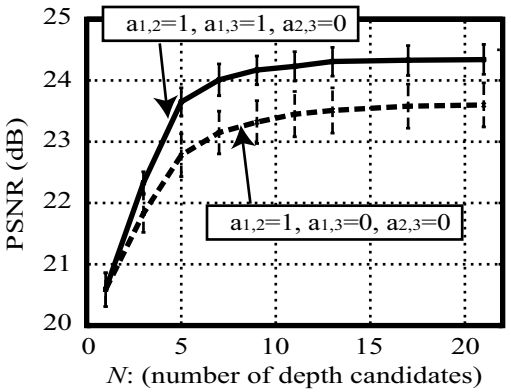


Fig. 7. PSNR of synthetic images: our method effectively increases PSNR with an increase of the number of depth candidates.

6. CONCLUSIONS

In this paper, we discussed a *focus* measure for LFR that uses differences between some kinds of reconstruction filters. Combined with the LFR method, it enables us to generate depth maps and all in-*focus* images at arbitrary viewpoints. Experimental results showed that the larger N (the number of depth candidates) yields the higher PSNR on synthetic images. However, the total computation cost is approximately proportional to N. For the future, we should consider the tradeoff between the quality of synthesis and the processing time in practical cases.

Acknowledgement: Thanks to Prof. Hiroshi Harashima of The University of Tokyo, Japan, for his helpful discussions.

7. REFERENCES

[1] H. -Y. Shum, S. B. He, and S. -C. Chan. "Survey of Image-Based Representations and Compression Techniques", *IEEE Trans. CSVT* 13, 11, pp. 1020 – 1037. 2003.

[2] M. Levoy and P. Hanrahan. "Light Field Rendering", *Proc. ACM SIGGRAPH*, pp. 31 – 42. 1996

[3] A. Isaksen, M. Leonard, and S. J. Gortler. "Dynamically Reparameterized Light Fields", *MIT-LCS-TR-778*. 1999.

[4] A. Isaksen, L. McMillan, and S. J. Gortler. "Dynamically Reparameterized Light Fields", *Proc. ACM SIGGRAPH*, pp. 297 – 306. 2000.

[5] J. Stewart, J. Yu, S. J. Gortler, and L. McMillan. "A New Reconstruction Filter for Undersampled Light Fields", *Proc. Eurographics Symposium on Rendering*, pp. 150 – 156. 2003.

[6] K. Takahashi, T. Naemura, and H. Harashima. "Depth of Field in Light Field Rendering", *Proc. IEEE ICIP*, 1, pp. 409 – 412. 2003.

[7] K. Takahashi, A. Kubota, and T. Naemura. "All in-Focus View Synthesis from Under-Sampled Light Fields", *Proc. VRSJ ICAT*, pp. 249 – 256. 2003.

[8] K. Sugita, K. Takahashi, T. Naemura, and H. Harashima: "Focus Measurement on Programmable Graphics Hardware for All in-Focus Rendering from Light Fields", *Proc. IEEE VR*, pp. 255 – 256. 2004.

[9] S. Baker, T. Sim, and T. Kanede. "When Is the Shape of a Scene Unique Given Its Light-Field: A Fundamental Theorem of 3D Vision?", *IEEE Trans. PAMI*, 25, 1, pp. 100 – 109. 2003.

[10] S. K. Nayer and Y. Nakagawa. "Shape from Focus", *IEEE Trans. PAMI*, 16, 8, pp. 824 – 831. 1994.

[11] J.-X. Chai, X. Tong, S. -C. Chany, and H. -Y. Shum. "Plenoptic Sampling", *Proc. ACM SIGGRAPH*, pp. 307 – 318. 2000.