Reconstructing Dense Light Field from a Multi-Focus Images Array

Akira Kubota[1,2] Kiyoharu Aizawa[1] Tsuhan Chen[2]

[1]*Dept. of Electrical Engineering, University of Tokyo*
[2]*Dept. of Electrical and Computer Engineering, Carnegie Mellon University*
akira@andrew.cmu.edu aizawa@hal.t.u-tokyo.ac.jp tsuhan@cmu.edu

Abstract

This paper presents a novel method for synthesizing a novel view from two sets of differently focused images taken by a sparse camera array for a scene of two approximately constant depths. The proposed method consists of two steps. The first step is a view interpolation to reconstruct an all-focused dense light field of the scene. The second step is to synthesize a novel view by light-field rendering technique from the reconstructed dense light field. The view interpolation can be achieved simply by linear filters that are designed to convert the defocus effects to the parallax effects without estimating the depth map of the scene. The proposed method can effectively create a dense array of pin-hole cameras (i.e., all-focused images), so that the final novel view is better than traditional method using sparse array of cameras. Experimental results on the real images from four aligned cameras are shown.

1. Introduction

As an alternative to the conventional geometry-based methods, various image-based rendering (IBR) methods [1] have been investigated for synthesizing a novel view. IBR methods do not require any geometrical information of the scene, but requires many images. The minimum number of the images needed for anti-aliased rendering is analyzed by the Plenoptic sampling theories [2, 3] in the frequency domain.

In this paper, we propose a method for synthesizing a novel view from the images taken by sparsely located cameras for a scene of two depths. The concept of our approach is illustrated in Fig. 1. Instead of using the pin-hole camera in the conventional IBR, we use aperture camera for taking input images with different foci on two depths. We utilize the defocus blur information appeared in the images as implicit depth information, avoiding depth map estimation. The proposed view synthesis method consists of two steps. In the first step, we interpolate the intermediate views densely between the original cameras to reconstruct an all-focused dense light field. The view interpolation can be achieved simply by linear filters that are designed to convert the defocus effects to the parallax effects. In the second step, we synthesize a novel view by light-field rendering (LFR) [4] from the reconstructed all-focused dense light field. By reconstructing the dense light field once, we can synthesize the novel views without aliasing at arbitrary positions and directions.

Main advantage of the proposed method is that the camera array is allowed to be sparse compared with the conventional IBR, because our method can create a dense light field from a very sparse array of captured images, which means we can reduce the number of cameras necessary for anti-aliased LFR. The capturing camera does not need to be all-focus. Our method can convert these partially focused images into all-focused images.

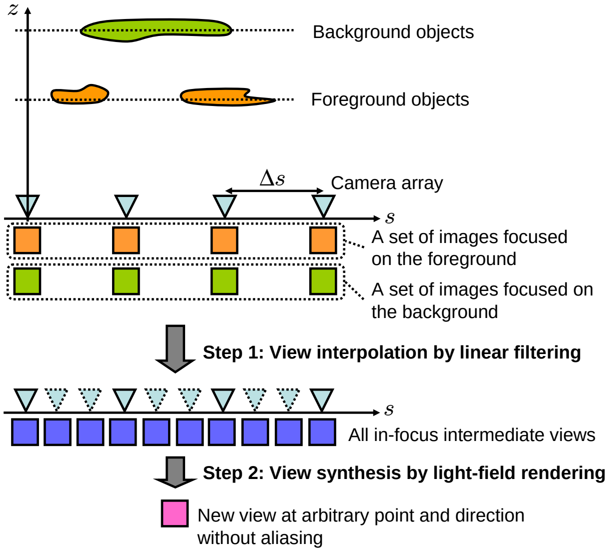

z
Background objects
Foreground objects
Δs
Camera array
s
A set of images focused on the foreground
A set of images focused on the background
Step 1: View interpolation by linear filtering
s
All in-focus intermediate views
Step 2: View synthesis by light-field rendering
New view at arbitrary point and direction without aliasing

Fig. 1. The concept of our approach.

Our previous method [6] was not able to render the parallax effect on the novel view at different depth and direction precisely, since a novel view is rendered directly from the differently focused images without reconstructing the intermediate views. In addition, the filters used to synthesize the novel view were not stable in the low frequency range and diverged at DC, resulting in the undesirable luminance variation. To avoid this problem, in this paper, we introduce a frequency-dependent shifting for the view interpolation in such that the DC component is not interpolated, which does not affect much the rendering quality.

2. Reconstructing dense light field

As shown in Fig. 1, in our method, we assume that the scene has foreground and background objects at different depths and that the cameras are aligned on the s axis with interval of Δs. We acquire two set of images by a 1-D array of cameras by changing the focus on the foreground and the background of the scene, respectively. We interpolate the intermediate views that are focused on the both objects at arbitrary position s through the linear processing of the acquired images. Once the intermediate views are obtained, novel views can be synthesized by LFR.

2.1. Parameterization and modeling

We express the acquired images focused on the foreground and background at camera position s_i (i.e. $\Delta s \cdot i$, i = 0, 1,...,N) as $g_1(s_i, u, v)$ and $g_2(s_i, u, v)$, respectively, by parameterizing them using s_i and the pixel location (u, v). They are considered as the light field data sparsely sampled along the s axis. Let us define $f_1(s_i, u, v)$ and $f_2(s_i, u, v)$ as the in-focus foreground and background texture visible from s_i, respectively. We model the acquired images by a linear combination of their textures with blur function as follows [5]:

$$\begin{cases} g_1(s_i, u, v) = f_1(s_i, u, v) + h_2(u, v) * f_2(s_i, u, v) \\ g_2(s_i, u, v) = h_1(u, v) * f_1(s_i, u, v) + f_2(s_i, u, v) \end{cases}, \tag{1}$$

where $*$ means a 2-D convolution operation, h_1 and h_2 are space-invariant blur functions, which are assumed to be the Gaussian function.

For the modeling of the intermediate view focused on both objects from arbitrary position s, which is the continuous light field along the s axis, we have to consider two things; (1) how to model the parallax and (2) how to model occluded background texture. To model the parallax, we use a sum of the shifted foreground and background textures. When the textures at s_i are used, it is modeled by

$$\begin{aligned} &f(s_i, s, u, v) = \\ &\quad f_1(s_i, u - u_1(s_i, s), v) + f_2(s_i, u - u_2(s_i, s), v), \end{aligned} \tag{2}$$

where u_1 and u_2 are the shift amounts. They are determined by the position s and the disparity of the foreground and background between the adjacent cameras, d_1 and d_2, respectively, as follows:

$$u_1(s_i, s) = \frac{(s_i - s)}{\Delta s} d_1, \quad u_2(s_i, s) = \frac{(s_i - s)}{\Delta s} d_2. \tag{3}$$

Once the parallax is modeled, to create an intermediate view, we can simply interpolate between two neighboring images, with appropriately shifted textures. Doing so also has the effect of fill in occluded background in either one of two images. For $s_i < s < s_{i+1}$, the intermediate view that we interpolate is modeled by

$$\begin{aligned} &g(s, u, v) = \\ &\quad \alpha f(s_i, s, u, v) + (1 - \alpha) f(s_{i+1}, s, u, v), \end{aligned} \tag{4}$$

where α is the blending value that is $(s_{i+1} - s)/\Delta s$.

2.2. View interpolation with linear filters

We can take Fourier transform (FT) of the models with respect to (u, v) in eq. (1) - (4), since they are expressed by linear operations. The FT of the acquired image and the intermediate view can be represented, respectively, as follows:

$$\begin{cases} G_1(s_i, \xi, \eta) = F_1(s_i, \xi, \eta) + H_2(\xi, \eta) F_2(s_i, \xi, \eta) \\ G_2(s_i, \xi, \eta) = H_1(\xi, \eta) F_1(s_i, \xi, \eta) + F_2(s_i, \xi, \eta) \end{cases} \tag{5}$$

and

$$\begin{aligned} &G(s, \xi, \eta) = \\ &\quad \alpha F(s_i, s, \xi, \eta) + (1 - \alpha) F(s_{i+1}, s, \xi, \eta), \end{aligned} \tag{6}$$

where

$$\begin{aligned} F(s_i, s, \xi, \eta) &= e^{-j2\pi u_1(s_i, s)\xi} F_1(s_i, \xi, \eta) \\ &\quad + e^{-j2\pi u_2(s_i, s)\xi} F_2(s_i, \xi, \eta). \end{aligned} \tag{7}$$

The capital letter function is the 2-D Fourier transform of the corresponding small letter function and (ξ, η) indicates horizontal and vertical frequencies.

For each s_i, given H_1 and H_2, eliminating $F_1(s_i)$ and $F_2(s_i)$ from eq.(5) and (7) yields the following sum-of-products formula [6]:

$$\begin{aligned} F(s_i, s, \xi, \eta) &= K_1(s_i, \xi, \eta) G_1(s_i, \xi, \eta) \\ &\quad + K_2(s_i, \xi, \eta) G_2(s_i, \xi, \eta) \end{aligned} \tag{8}$$

where $K_1(s_i)$ and $K_2(s_i)$ can be considered as the frequency characteristics of the linear filters that are applied to $G_1(s_i, \xi, \eta)$ and $G_1(s_i, \xi, \eta)$ in the frequency domain, respectively. For the case of $(\xi, \eta) \neq (0, 0)$, $K_1(s_i)$ and $K_2(s_i)$ are given by

$$\begin{cases} K_1(s_i) = \dfrac{e^{-j2\pi u_1(s_i, s)\xi} - e^{-j2\pi u_2(s_i, s)\xi} H_1}{1 - H_1 H_2} \\ K_2(s_i) = \dfrac{e^{-j2\pi u_2(s_i, s)\xi} - e^{-j2\pi u_1(s_i, s)\xi} H_2}{1 - H_1 H_2} \end{cases}. \tag{9}$$

At DC, since the denominator, i.e. $1 - H_1 H2$, equals 0, the limit value of eq. (9) to the DC diverge, which results in visual artifact (undesirable luminance variation [6]) on the interpolated view.

To avoid this problem, we propose a frequency dependent shifting method that is designed to have the shift amount gradually decreased to zero at DC, shown in Fig. 2. In other words, we do not shift or interpolate DC component of the intermediate view. This is reasonable from the fact that the low frequency components including DC do not cause much visual artifacts in the quality of the image, even if they are not shifted.

The amount of the frequency dependent shifting that we use for u_1 is

$$u_1'(s_i, s, \xi) = \begin{cases} \dfrac{u_1(s_i, s)}{2} \left\{ 1 - \cos\left(\dfrac{\pi \xi}{\xi_{th}}\right) \right\}, & \xi \leq \xi_{th} \\ u_1(s_i, s), & \xi > \xi_{th} \end{cases} \tag{10}$$

$u_2'(s_i, s, \xi)$ is similar to the above. By using those frequency dependent shifting, the limit of eq. (9) to the DC converges to real number:

$$\begin{cases} \lim_{\xi, \eta \to 0} K_1(s_i, \xi, \eta) = R_1^2 / (R_1^2 + R_2^2) \\ \lim_{\xi, \eta \to 0} K_2(s_i, \xi, \eta) = R_2^2 / (R_1^2 + R_2^2) \end{cases}, \tag{11}$$

where R_1 and R_2 are the blur radii of the blur function h_1 and h_2. As a result, $K_1(s_i)$ and $K_2(s_i)$ can be designed for i = 0, 1,...,N, and then $F(s_i, s)$ can be generated. Therefore, $g(s, u, v)$ can be interpolated simply by linear filtering of the acquired images.

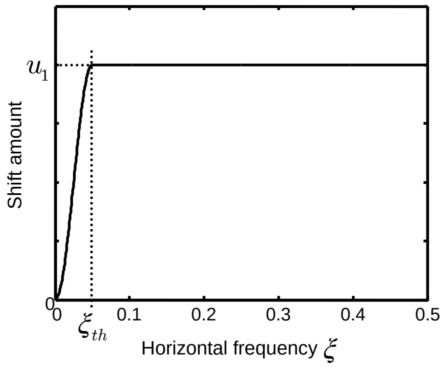


Fig. 2. Frequency dependent shifting method

3. Experimental results on real images

Experiments were performed using real images (480 × 280 [pixels]) acquired with 4 cameras by interval of Δs = 5 [mm], as shown in Fig. 3. Image registration has been done between two differently focused images at the same camera position to correct the difference of magnification due to focusing. The scene consists of the foreground (plant and globe) and background (bear toy and magazines) objects located approximately at z_1 = 400 and z_2 = 800 [mm], respectively. We measured disparities between adjacent cameras by searching the minimum of $|g_1(s_i, u, v)$ - $g_1(s_{i+1}, u - d_1, v)|^2$ for d_1 and $|g_2(s_i, u, v)$ - $g_2(s_{i+1}, u - d_2, v)|^2$ for d_2, They were estimated as d_1 = 15.8 and d_2 = 7.9 [pixels]. The blur radiuses were estimated by a method in [7] as R_1 =2.1 and R_2 = 2.8 [pixels]. Note that all these preprocessing steps do not involve estimating the depth map of the scene, which is one of the advantages offered the proposed technique.

The number of intermediate views needed to be interpolated for anti-aliased LFR can be given by the depth range of the scene according to the Plenoptic sampling theory [2]. From this theory, the sampling interval of s, say δs, in $g(s, u, v)$ requires to satisfy the condition that the difference of disparities of the foreground and background object between adjacent sampled views must be less than 2 [pixels], which leads to the condition $\delta s < 1.2$ [mm] in our experimental setting. In our experiment, to render the view in higher quality, we oversampled by generating 31 intermediate views with interval of 0.5 [mm]. ξ_{th} was set to 0.05 for frequency dependent shifting. Figure 4 compares the epipolar-plane images (EPIs) at v = 220 [pixels] of the original acquired images and the interpolated views. It shows that the intermediate pixel values were interpolated well according to their corresponding depth.

A novel view $I(x, y)$ is synthesized by light-field rendering as $I(x, y) = g(s, u, v)$. The parameters s, u and v are computed for each (x, y) based on the optimal depth that is given by $2(z_1^{-1} + z_2^{-1})^{-1}$ [2]. Figure 5 (a) and (b) show the synthesized views from z = -50 and 50 [mm], respectively. It can be seen that they were synthesized without visible artifact in the occluded background. Vertical distortion was slightly caused because of the synthesizing using 1-D array image inputs, but it does not affect the subjective quality of the novel view. Figure 5 (c) and (d) show the magnified images of the views from z = 20 synthesized using the interpolated 31 views and the original 4 images (here, all in-focus images generated at original position), respectively. The quality in (c) was much improved compared with (d) where blurring and ghosting artifacts appear due to the under-sampling of the light field data. The novel views synthesized from different directions were shown in Fig. 6, where aliasing artifact is successfully avoided.

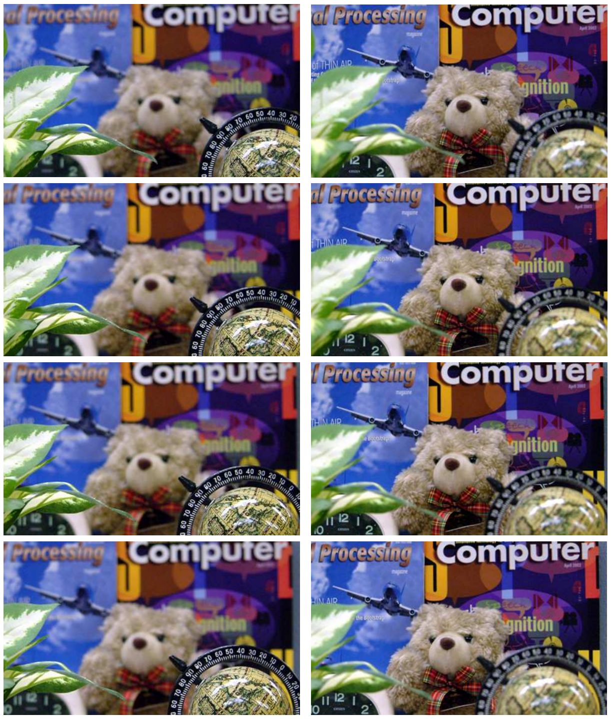
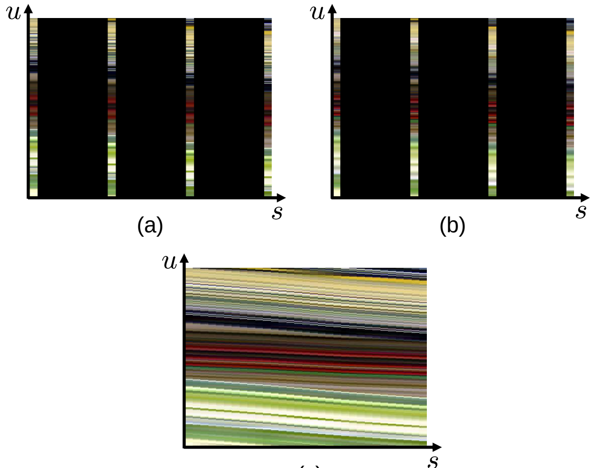


Fig. 3. Two sets of images taken from 4 horizontally aligned cameras at s_i = 0, 5, 10, 15 [mm] (from Top to Bottom), with different foci on the foreground (Left) and the background (Right)

4. Conclusion

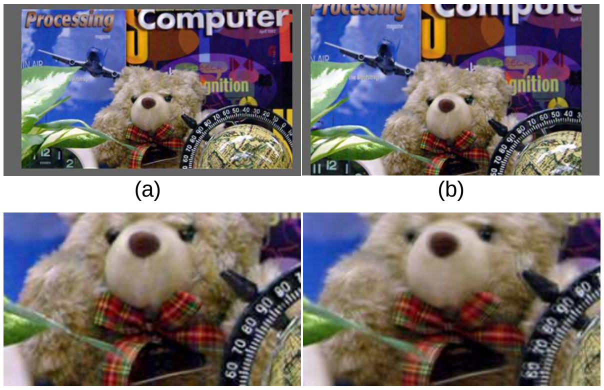
In this paper, we presented a view interpolation method for reconstructing dense light field from an image array with different focus on a two-depth scene. The view interpolation can be done by simple linear processing without estimating the depth map of the scene. By using the proposed method, we have created effectively a dense array of pin-hole cameras (i.e., all-focused images). Therefore, the final view result was reconstructed better than traditional method that uses a sparse array of cameras. The proposed method can be extended to 2-D image array data easily.

As future work, we extend our method to deal with more than two layers in the scene. We also may render the depth-of-field effect more effectively by rebinding the dense light rays using a wide-aperture filter [8].



(a) (b) (c)

Fig. 4. Comparison with EPIs at line v = 220 [pixles]. (a) and (b): the EPI of the original images focused on the foreground and background, i.e. $g_1(s_i, u, 220)$ and $g_2(s_i, u, 220)$. (c): The dense EPI interpolated from (a) and (b), i.e. $g(s, u, 220)$



(a) (b) (c) (d)

Fig. 5. Synthesized views from different depths with the horizontal position fixed at s = 7.5 [mm]. (a) and (b) are the view at z = -50, 50 [mm]. Bottom: the comparison between the magnified images of the views at z = 20 [mm]. (c) is synthesized using the interpolated views; (d) using only a set of all in-focus images at original position.

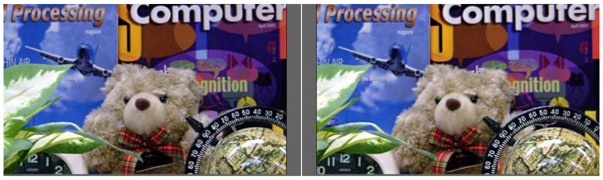


Fig. 6. Synthesized views from different view direction of 1.0 (Left) and -1.0 [deg.] (Right) with respect to z axis.

5. References

[1] H-Y. Shum, S. B. He, and S-C. Chan, “Survey of Image-Based Representations and Compression Techniques”, *IEEE Trans. on Circuits and Systems for Video Technology* Vol. 13, No. 11, pp. 1020-1037, 2003

[2] J-X. Chai, X. Tong, S.-C. Chany and H.-Y. Shum, “Plenoptic Sampling,” *proc. of SIGGRAPH2000*, pp. 307-318, 2000

[3] C. Zhang and T. Chen, “Spectral Analysis for Sampling Image-Based Rendering Data”, *IEEE Trans. on Circuits and Systems for Video Technology* Vol. 13, No. 11, pp. 1038-1050, 2003

[4] M. Levoy and P. Hanrahan, “Light field rendering”, *proc. of SIGGRAPH96*, pp. 31-42, 1996

[5] K. Aizawa, K. Kodama and A. Kubota, “Producing Object Based Special Effects by Fusing Multiple Differently Focused Images”, *IEEE Trans. on Circuits and System for Video Techinology*, Vol. 10, No. 2, pp. 323-330, 2000

[6] A. Kubota and K. Aizawa, “A novel Image-based rendering method by linear filtering of multiple focused images acquired by a camera array”, *proc. of ICIP03*, pp. 701-704, 2003

[7] A. Kubota, K. Kodama and K. Aizawa, “Registration and blur estimation methods for multiple differently focused images”, *proc. of ICIP99*, pp. 447-451, 1999

[8] A. Isaksen, L. McMillan and S. K. Gortler, “Dynamically Reparameterized Light Fields”, *proc. of SIGGRAPH2000*, pp. 297-306, 2001