Image-Based Refocusing by 3D Filtering

Akira Kubota[1], Kazuya Kodama[2], and Yoshinori Hatori[1]

[1] Interdisciplinary Graduate School of Science and Technology,
Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama 226-8502, Japan
[2] Research Organization of Information and Systems, National Institute of Informatics,
Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
kubota@ip.titech.ac.jp, kazuya@nii.ac.jp, hatori@ip.titech.ac.jp

Abstract. This paper presents a novel spatial-invariant filtering method for rendering focus effects without aliasing artifacts from undersampled light fields. The presented method does not require any scene analysis such as depth estimation and feature matching. First, we generate a series of images focused on multiple depths by using the conventional synthetic aperture reconstruction method and treat them as a 3D image. Second we convert it to the alias-free 3D image. This paper shows this conversion can be achieved simply by a 3D filtering in the frequency domain. The proposed filter can also produce depth-of-field effects.

Keywords: Refocus, Light Field, 3D Filtering, Aliasing.

1 Introduction

Using acquired multiview images or light fields, synthetically post-producing 3D effects such as disparity and focus has attracted attention recently [2,8]. If dense light fields data is available, these effects can be easily produced with high quality by light field rendering (LFR) method [6,4]. LFR method essentially performs resampling the acquired light fields, independent of the scene complexity.

This paper addresses an image-based refocusing problem in the case when input light fields data was undersampled. In this case, applying LFR method to the undersampled data introduces aliasing or ghosting artifacts in the rendered image; hence the scene analysis such as depth estimation is needed to reduce the artifacts. The objective of this paper is to present a novel spatial-invariant filter that can produce both focal depth and depth-of-field effects with less aliasing artifacts, requiring no scene analysis.

1.1 Problem Description of Image-Based Refocusing

Consider a XYZ world coordinate system. We use multiview images $f_{(i,j)}(x, y)$ captured with a 2D array of pin-hole cameras on the XY plane as an input light field data. The coordinate (x, y) denotes the image coordinate in every image and $(i, j) \in \mathbb{Z}^2$ represents the camera positions on the XY plane (both distance between cameras and focal length are normalized to be 1 for simpler notation). The scene is assumed to exist in the depth range of $Z_{\min} \leq Z \leq Z_{\max}$.

D. Mery and L. Rueda (Eds.): PSIVT 2007, LNCS 4872, pp. 385–398, 2007.

The goal of image-based refocusing in this paper is to reconstruct the image $f_{(0,0)}(x, y;\ Z, R)$, that is, the image at the origin of the XY plane focused at depth Z with an aperture of radius R. Since the target camera position is fixed at the origin, we simply represent the desired image as $f(x, y;\ Z, R)$.

1.2 Synthetic Aperture Reconstruction

Refocusing effects can be generated by taking the weighted average of the multiview images that are properly shifted. This shifting-and-averaging method, called synthetic aperture reconstruction (SAR) method, was presented by Haeberli [5] and firstly applied to real scenes by Isaksen et. al [4].

In the SAR method, as illustrated in fig. 1 (a) where parameters y and Y are fixed, let the refocused image be $g(x, y;\ Z, R)$, it is synthesized by

$$g(x, y;\ Z, R) = \sum_{i,j} w_{(i,j)} f_{(i,j)}(x - i/Z, y - j/Z), \tag{1}$$

where $w_{(i,j)}$ is the weighting value for the multiview image $f_{(i,j)}$ and is a function of the aperture radius R:

$$w_{(i,j)}(R) = \frac{1}{\pi R^2} e^{-(i^2+j^2)/R^2}. \tag{2}$$

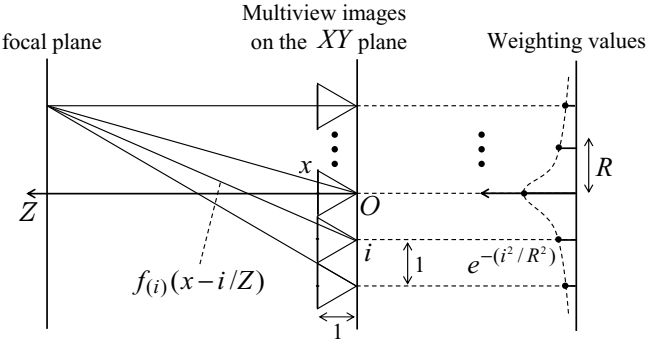
This weighting values are sampled from the point spread function (PSF) that we desire on the refocused image. In this paper, we use a Gaussian-type PSF function shown in equation (2), because it produces visibly natural blur effects. The shift amounts, i/Z and j/Z, correspond to respectively horizontal and vertical disparities between g (or $f_{(0,0)}$) and $f_{(i,j)}$ with respect to the focal depth Z.

In the resultant image g, objects on the focal plane appear sharp, while other objects not on the focal plane appear blurred. Changing the aperture radius R produces depth-of-field effects on the image. In addition, the SAR method can be efficiently performed in the 4D Fourier domain by a 2D slice operation [3]

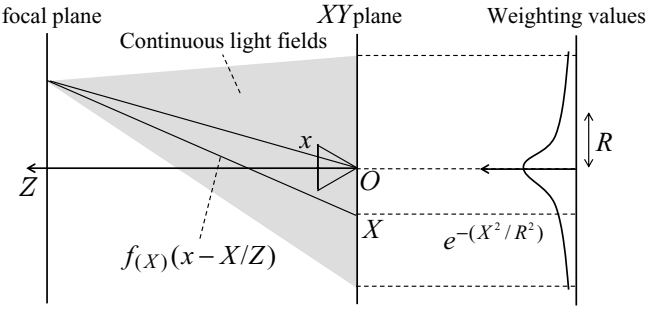
1.3 Motivation

The SAR method requires a large number of images taken with densely arranged cameras. With sparsely spaced cameras, in the resulting image, the blurred regions suffer from aliasing or ghosting artifacts. The blur effects rendered in these regions are not the desired effects but unnatural blur effects due to sparsely sampled PSF (see fig. 1(a)).

Our goal in this paper is to reduce aliasing artifacts in the blurred regions, improving the quality of synthetic aperture images. If continuous light fields $f_{(X,Y)}(x, y)$ are available (see fig 1(b)), the desired refocused image can be generated as an integration of all light rays with appropriate weights of Gaussian function:



(a) for multivew images as sparsely sampled light fields.



(b) for continuous light fields.

Fig. 1. Refocusing by the SAR method to reconstruct an image focused at depth Z with Gaussian aperture of radius R. (a) The refocused image is synthesized as a weighted average of shifted multiview images based on the focal plane. The weighting values are sampled from Gaussian point spread function. (b) The desired refocused image is synthesized by a weighted integration of corresponding light rays.

$$f(x,y;\ Z,R) = \iint w_{(X,Y)} f_{(X,Y)}(x-X/Z, y-Y/Z)\,dX\,dY,$$

$$\text{where}\quad w_{(X,Y)} = \frac{1}{(\pi R^2)} e^{-(X^2+Y^2)/R^2}; \quad (3)$$

hence the goal is to reconstruct f from g.

This is however generally a difficult problem: one has to identify the aliased regions in g and change them into those with desired blur effects. One reasonable way is to estimate missing light rays densely and then apply the SAR method to the obtained dense light fields. This approach essentially requires estimating the scene structure to some extent, according to plenoptic sampling theory [7]. For instance, Georgeiv et al. [8] have used a morphing technique based on feature correspondences to fill all necessary light fields; as a result, they successfully produced focus effects with less aliasing artifacts.

In this paper, in contrast to computer vision based approaches, we present a novel reconstruction method that does not require any scene analysis. Our idea is that we treat a series of images refocused on multiple depths by the conventional SAR method as a 3D image and convert it to the alias-free 3D image. This conversion can be achieved simply by a 3D filtering in the frequency domain; hence the computation is constant independent of the scene complexity. This approach can cope with rendering depth-of-field effects as well. Related to our method, Stewart [1] presented a novel reconstruction filter that can reduce artifacts by blurring aliased regions much more; however it cannot handle rendering effects of focal depth and depth-of-focus.

2 The Proposed Image-Based Refocusing Method

We derive a novel reconstruction filter that converts g into f. In our method, instead of g in (1), we use the following image series generated by SAR method with the circular aperture of fixed radius $R_{\max}$.

$$g(x, y;\ Z, R_{\max}) = \frac{1}{N} \sum_{i,j \in A_{\max}} f_{(i,j)}(x - i/Z, y - j/Z), \tag{4}$$

where $A_{\max}$ denotes the aperture region defined as a set $\{(i, j) | \sqrt{i^2 + j^2} \leq R_{\max}\}$, and N is the number of cameras inside the aperture.

2.1 3D Image Formation and Its Modeling

By changing the focal depth Z, we synthesize the refocused image sequence and treat it as a 3D image. Introducing the parameter z that is inversely proportional to Z, we represent $g(x, y;\ Z, R_{\max})$ in (4) and $f(x, y;\ Z, R)$ in (3) as 3D image $g(x, y, z;\ R_{\max})$ and $f(x, y, z;\ \ R)$, respectively. The synthesis process of these 3D images can be modeled by

$$g(x, y, z;\ R_{\max}) = h(x, y, z) * s(x, y, z) \tag{5}$$

$$f(x, y, z;\ R) = b(x, y, z) * s(x, y, z), \tag{6}$$

where $h(x, y, z)$ and $b(x, y, z)$ are the 3D PSF, $s(x, y, z)$ is the color intensity distribution in the 3D scene, and $*$ denotes a 3D convolution operation. These model are derived in similar manner to the observation model of multi-focus images in microscopic imaging [9]. Note that we assume here that z ranges $(-\infty, \infty)$ and effects of occlusions and lighting are negligible.

The 3D PSF h is given by

$$h(x, y, z) = \frac{1}{N} \sum_{i,j \in A_{\max}} \delta(x - iz, y - jz), \tag{7}$$

where δ is Dirac delta function. Unlike the case of microscopic imaging, the PSF does not need to be estimated; it can be computed correctly based only on the

camera arrangement set. The 3D PSF b is the desired PSF and can be ideally modeled by

$$b(x,y,z) = \frac{1}{\pi R^2} \iint e^{-(X^2+Y^2)/R^2} \delta(x - zX, y - zY) dX dY. \tag{8}$$

The scene information s can be defined as

$$s(x,y,z) \doteq f_{(0,0)}(x,y) \cdot \delta(d_{(0,0)}(x,y) - 1/z) \tag{9}$$

where $d_{(0,0)}(x,y)$ is the depth map from the origin. The scene information is a stack of 2D textures at depth Z visible form the origin. Neither the scene information nor the depth map are known.

2.2 Reconstruction Filter

Taking 3D Fourier transform in both sides in equations (5) and (6) yields respectively

$$G(u,v,w;\ R_{\max}) = H(u,v,w)S(u,v,w), \tag{10}$$

$$F(u,v,w;\ R) = B(u,v,w)S(u,v,w). \tag{11}$$

where capital-letter functions denote the Fourier transform of the corresponding small-letter functions and (u,v,w) represents the frequency domain counterparts of the spatial variables (x,y,z). Both spectra $H(u,v,w)$ and $B(u,v,w)$ are calculated as real functions.

By eliminating the unknown S from the above equations, we derive the following relationship if $H(u,v,w)$ is not zero.

$$F(u,v,w;\ R) = \frac{B(u,v,w)}{H(u,v,w)} \cdot G(u,v,w;\ R_{\max}). \tag{12}$$

To avoid error amplification due to division by the value of $H(u,v,w)$ close to zero, we employ Wiener type regularization to stably reconstruct the desired 3D image

$$\hat{F}(u,v,w;\ R) = \frac{B(u,v,w)H(u,v,w)}{H^2(u,v,w) + \gamma} \cdot G(u,v,w;\ R_{\max}), \tag{13}$$

where γ is a positive constant. By taking 3D inverse Fourier transform of $\hat{F}$, we can finally obtain the desired refocused 3D image $\hat{f}$.

This suggests that image-based refocusing can be achieved simply by this linear and spatially invariant filtering. The coefficient of G is the reconstruction filter. It consists of known PSFs independent of the scene; hence no scene analysis such as depth estimation is required in our method.

3 Experimental Results

3.1 Rendering Algorithm

Let us first here give the algorithm of our method when applying to digital dataset of multiview images, as follows:

1. synthesizing the 3D image based on the discrete version of eq. (4);
2. computing the reconstruction filter with the aperture radius R we desire;
3. taking 3D Fourier transform of both g and the filter, and compute $\hat{F}$ by eq. (13);
4. finally we take inverse 3D Fourier transform of the $\hat{F}$ to reconstruct the refocused images $\hat{f}$.

In the step 1, assume L focal planes discretely located at Z_l $(l = 1, ..., L)$ from near to far, we synthesize L images focused on these focal depths. The 3D image g is formed as a set of the synthesized images with paremerar z_l, the inverse of Z_l.

Each depth Z_l is determined to be

$$Z_l = \left[\frac{1}{Z_L} - \left(\frac{1}{Z_1} - \frac{1}{Z_L} \right) \frac{l-1}{L-1} \right]^{-1}, \tag{14}$$

such that the disparities i/Z_l and j/Z_l (i.e., iz_l and jz_l) can be changed with equal interval. The number L should be given so that the interval be less than 1 pixels, as suggested by plenoptic sampling theory.

The depth range where we arrange the focal planes was determined such that it satisfies both conditions: $1/Z_1 - 1/Z_L = 2(1/Z_{\min} - 1/Z_{\max})$ and $(1/Z_1 + 1/Z_L)/2 = (1/Z_{\min} + 1/Z_{\max})/2$, which impose the range be two times wider than that of the scene in the inverse dimension z, setting the center of both the ranges to be the same.

In the step 2, let the focal plane range $1/Z_1 - 1/Z_L$ be $\bar{z}$, we set the support range in z for both PSFs h and b to be $-\bar{z}/2 \leq z \leq \bar{z}/2$ (see fig. 7 (a)). That is, the range in z should be the same $\bar{z}$ for all 3D data, g, f, h and b, which ensures the DC energy conservation of $G(0,0,0) = F(0,0,0) = H(0,0,0) = B(0,0,0)$.

3.2 Results for a Synthetic Scene

The proposed method was tested using synthetic 9x9 multiview images (640x480 pixels each) for a synthetic scene. The scene structure and camera array setting are illustrated as a top view (from Y axis) in Fig. 2. The scene consists of slant a wall of wood texture and a boll of checker pattern. They exist in the depth range of 30–50. Cameras are regularly arranged in 2D lattice with equal space of 1 and the horizontal field of view is set to be 50 degree for all the cameras.

Figure 3 shows the reconstructed images that were refocused by our method and those by the conventional SAR method. In this simulation, we fixed R at 2 and used the following parameters: $L = 16$, $R_{\max}$=4 and $\gamma = 0.1$. In these results, from the top, the focal depth were 33.9, 37.4, 41.8, 50.6 and 58.9, which

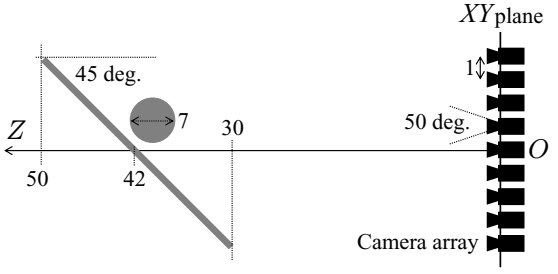


Fig. 2. Synthetic scene and camera array setting in our simulation

corresponds to Z_4, Z_6, Z_8, Z_{11}, and Z_{13}. In the results by our method, blur effects were rendered without artifacts for all the cases, whereas in those by the conventional method, blurred regions appear ghosted in the latter two cases—especially artifacts are much visible in the checker pattern regions.

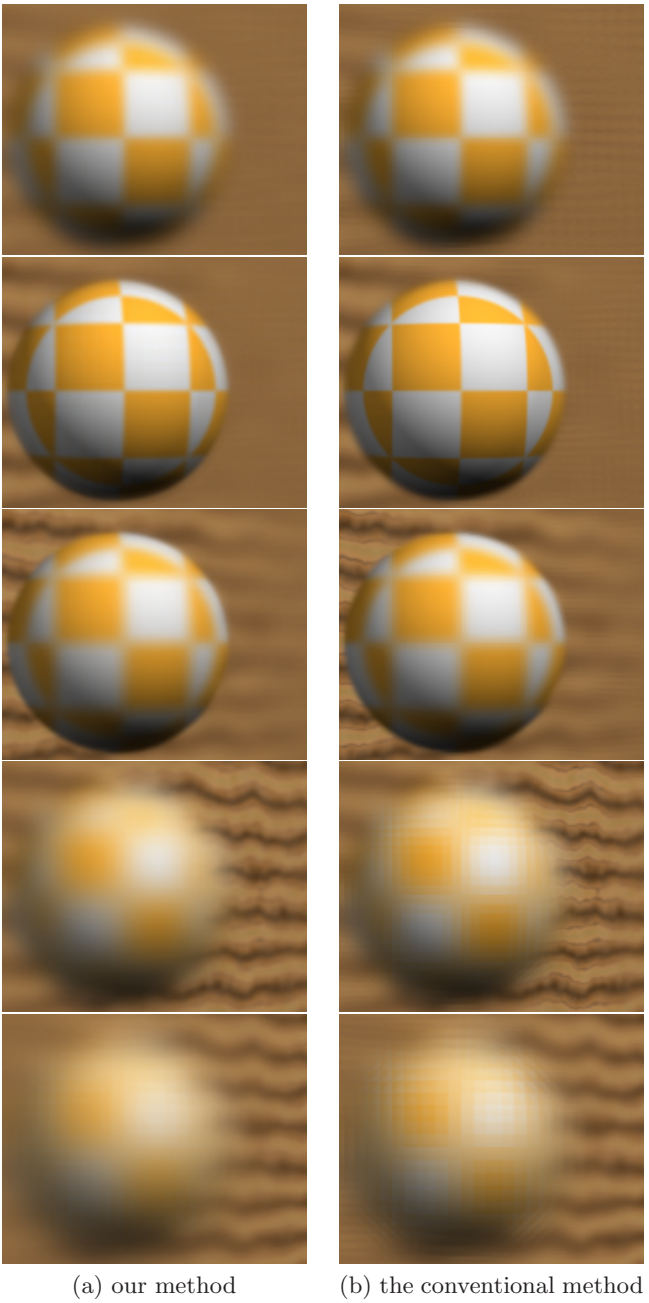
Rendering depth-of-field effects are demonstrated in fig. 4, where the focal depth was fixed at $Z_{11} = 50.6$ and the aperture radius R was varied from 0.5 to 2.5 with increments of 0.5. In the case of $R = 0.5$, the image refocused by the conventional method (the top image in fig. 4(b)) was focused in every depth region and the desired blur effects were not correctly rendered. This is another drawback of the conventional method. In contrast, our method allows rendering smaller size blur effects, as shown in the top image in fig. 4(a). In the other cases, it can be clearly seen that our method produced alias-free blur effects with high quality; while the conventional method introduced aliasing artifacts in the blurred regions.

3.3 Results for a Real Scene

We used 81 real images captured with a 9x9 camera array, which are provided from "The Multiview Image Database," courtesy of the University of Tsukuba, Japan. Image resolution is 320x240 pixels (down sampled from the original resolution of 640x480) and the distance between cameras is 20 [mm] in both the horizontal and vertical directions. The scene contains an object ("Santa Claus doll") in the depth range of 590–800 [mm], which is the target depth range in this experiment. In this experiment, we used the following parameters: $L = 16$, $R_{\max}$=4x20 [mm] and $\gamma = 0.1$.

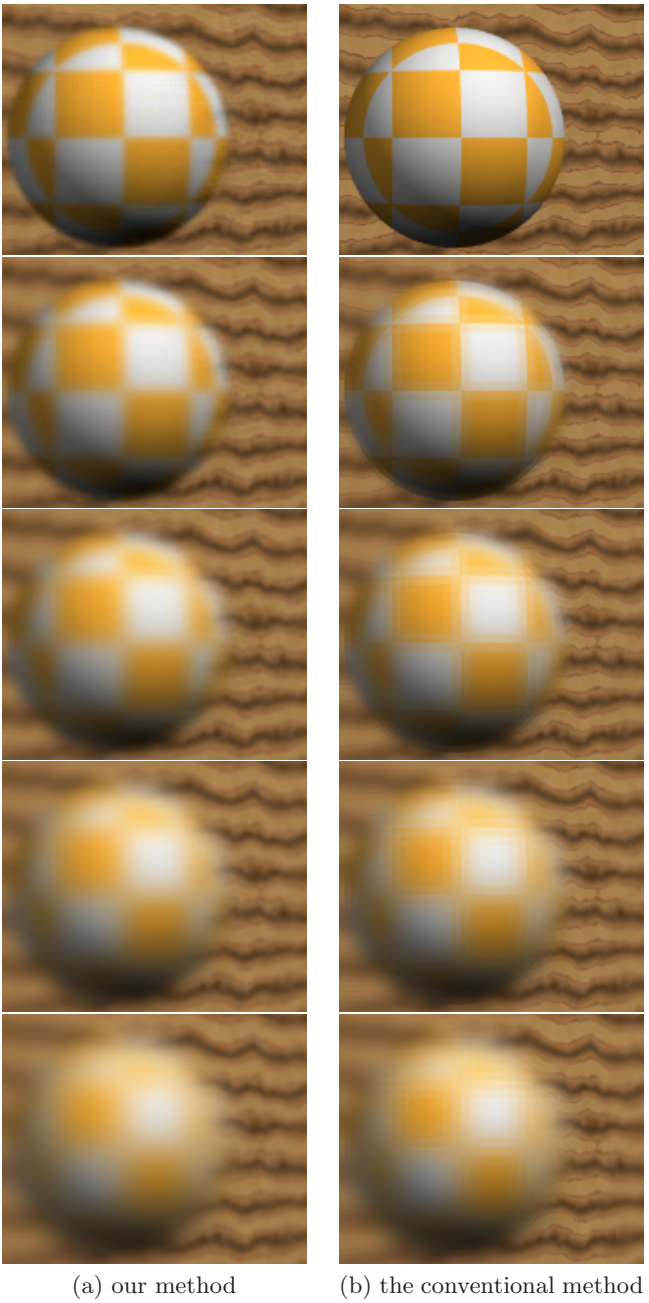
One problem arises from narrow field-of-view (which was 27.4 degree) of the camera used: it is too narrow to produce refocused images with enough field-of-view. To overcome this problem, we applied Neumann expansion to all the multiview image to virtually fill the pixel values outside the field-of-view with the pixel value at the nearest edge.

The reconstructed images are shown in figures 5 and 6. The former demonstrates effects of change of focal depth, the latter does those of depth-of-field. The results show that advantage of our method over the conventional one; the our method works well, significantly suppressing aliasing artifacts that were visible



(a) our method (b) the conventional method

Fig. 3. Refocusing results when the focal depth was varied from near to far (from the top image) with fixed aperture radius at $R = 2$



(a) our method (b) the conventional method

Fig. 4. Refocusing results demonstrating depth-of-field effects when the aperture radius R is varied from 0.5 to 2.5 (from the top image) with the focal depth fixed

in the images reconstructed by the conventional method. In the images of our method, some vertically scratched lines are slightly visible. This is due to the effect of Neumann expansion, not aliasing artifacts.

Note that noise amplification, which is generally a critical issue, does not occur as large as an general inverse filtering method may have, because noise on the multiview images are averaged out and reduced in the 3D image g, which is a well-known feature in synthetic aperture methods.

4 Discussion

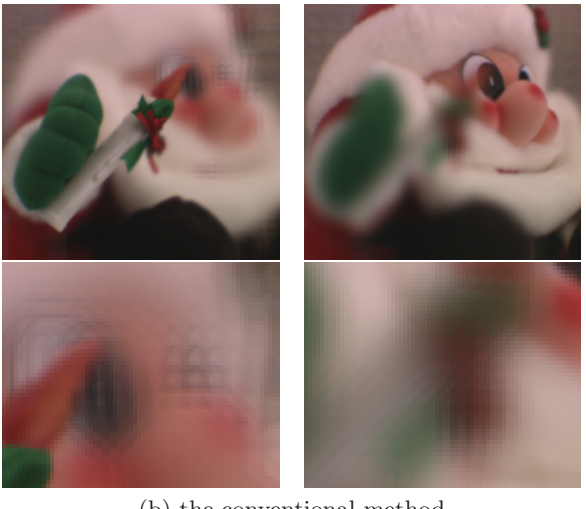
We have shown that we can effectively create alias-suppressed refocused images from those with aliasing generated by the conventional method. This conversion process was achieved by a spatially invariant filtering, not estimating any of scene information even for undersampled light field data. This section gives this reason in the frequency domain analysis.

Figure 7 (b) shows the reason schematically. We can consider that not only h but b is composed of a set of Delta functions (a line). The Fourier transform of Delta function is given by the plane perpendicular to the Delta functions; hence the Fourier transform of both PSFs, H and B, is composed of a set of the planes that are rotated along w axis and slanted according to the corresponding camera position (i, j). As a result, H and B become double-cones shape functions, illustrated in the fig. 7 (b), and the radial size is proportional to the size of $R_{\max}$ and R, respectively. We have set $R_{\max} \geq R$; therefore, it ideally holds for most regions in the frequency domain that B is always zero when H is zero, which leads to that B/H is stable. Note that due to the limited support range $\bar{z}$, each plane in H become "thick" because of the effect of production with the Sinc function, resulting in having more non-zero components. This property helps make B/H more stable. Mathematically precise arguments on this property is needed in our future work.

An alternative interpretation of our method is that by generating G, our method extract all the information needed for reconstructing F. It is not necessary at all to recover the whole 3D scene S, since our goal is to reconstruct F. The underlying idea of our method is the same with that in computerized tomography where the frequency components of the desired image are extracted by projections from many directions. In our method, by generating g using many multiview images at different positions, we extract the frequency components needed for reconstruct the desired 3D image f.

Figure 8 shows an example of numerically computed frequency characteristics H and B as a series of cross sections at different frequencies w. They were used for reconstructing the images in fig. 3 (a). The whiter regions indicate higher amplitude (black level represents zero). The whiter regions in B indicates the frequency components needed for reconstructing F and the white regions in H is the frequency components that can be extract from the 3D scene. Comparison between the both regions shows that the latter regions almost covers the former regions; hence stable reconstruction is possible.

(a) our method



(b) the conventional method

Fig. 5. Results for real multiview images. The left is focused on the near region; the right on the far region. The expanded images are also shown for comparison. The aperture radius set was R =1.5x20mm.

To extract the necessary information, we have to set $R_{\max}$ larger than R. It is true that the larger we set the aperture, the more stable and accurate we could obtain the scene information; however the larger aperture setting picks up the occluded regions much more often. This causes undesirable transparent effects at occluded boundaries, which are observed in the top image in fig. 5 (a) and

(a) our method (b) the conventional method

Fig. 6. Results for real multiview images when the aperture radius R is varied at 0.5, 1.0, 2.0 and 3.0 times of the distance between cameras. The focal depth was fixed at 590.

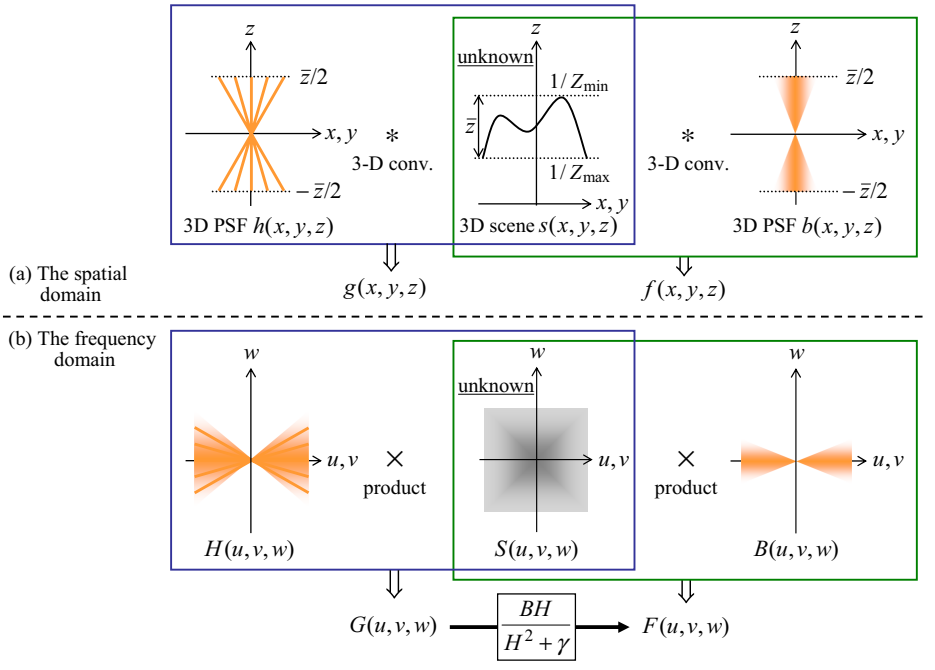


Fig. 7. The spatial and frequency analysis in our method. Image formation model in the x-z spatial domain and our reconstruction method in the u-w frequency domain.

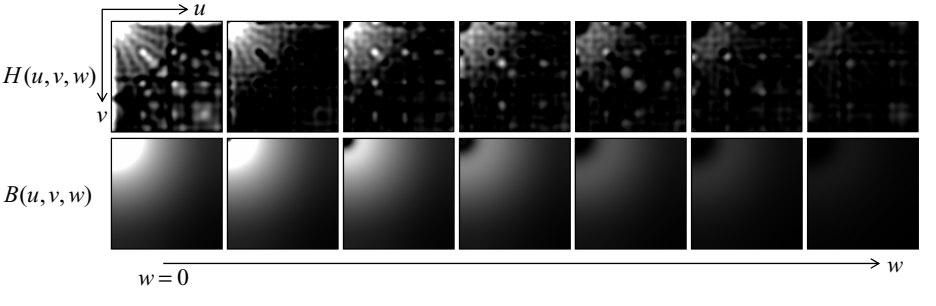


Fig. 8. Computed frequency characteristic of point spread functions

more visible in the top image in fig. 6 (a). We can not completely avoid this effect, because we do not estimate depth in our method. This is an disadvantage of our approach.

Another disadvantage is that the method requires computation much more than the conventional SAR method. This is mainly due to 3D filtering process.

One limitation of our approach is that the virtual view position must be on the XY plane. This is because shift amounts in eq. (4) depend on image coordinate; hence we cannot use Fourier transform in the image model. To

handle spatially varying shift amounts, we will have to consider to use some suitable orthogonal transformation such as wavelet transform.

5 Conclusions and Future Work

What we have shown in this paper is summarized as follows: by arranging an array of cameras or micro lens such that the reconstruction filter B/H be stable, alias-less image refocusing is possible through spatially invariant filtering without analyzing any scene information, even if the acquired light field data was undersampled in a sense of Plenoptic sampling. We believe that our method can be applied to the data acquired with integral camera and produce synthetic refocused images with higher quality.

The underlying methodology is similar to the idea used in computerized tomography; as future work, we could adopt some sophisticated methods that have been developed in that field, taking into account regularization, to enhance the quality more even for the case of less density of light field data.

References

1. Stewart, J., Yu, J., Gortler, S.J., McMillan, L.: A new reconstruction filter for undersampled light fields. In: Eurographics Symposium on Rendering 2003, EGSR 2003, pp. 150–156 (2003)
2. Ng, R., Levoy, M., Bredif, M., Duval, G., Horowitz, M., Hanrahan, P.: Light Field Photography with Hand-held Plenoptic Camera. Stanford Tech Report CTSR,2005-02(2005)
3. Ng, R.: Fourier slice photography. SIGGRAPH 2005, 735–744 (2005)
4. Isaksen, A., McMillan, L., Gortler, S.J.: Dynamically reparameterized light fields. SIGGRAPH 2000, 297–306 (2000)
5. Haeberli, P.E., Akeley, K.: The accumulation buffer: Hardware support for high-quality rendering. SIGGRAPH 1990, 309–318 (1990)
6. Levoy, M., Hanrahan, P.: Light field rendering. SIGGRAPH 1996, 31–42 (1996)
7. Chai, J.-X., Tong, X., Chany, S.-C., Shum, H.-Y.: Plenoptic sampling. SIGGRAPH 2000, 307–318 (2000)
8. Georgeiv, T., Zheng, K.C., Curless, B., Salesin, D., Nayar, S., Intwala, C.: Spatio-Angular Resolution Tradeoff in Integral Photography. In: Eurographics Symposium on Rendering, EGSR 2006, pp. 263–272 (2006)
9. Castleman, K.R.: Digital image processing, pp. 566–569. Prentice Hall, Englewood Cliffs (1996)
10. Sarder, P., Nehorai, A.: Deconvolution method for 3-D fluorescence microscopy images. Signal processing magazine 23(3), 32–45 (2006)