FREE VIEWPOINT, IRIS AND FOCUS IMAGE GENERATION BY USING A THREE-DIMENSIONAL FILTERING BASED ON FREQUENCY ANALYSIS OF BLURS

Kazuya Kodama, Hiroshi Mo

National Institute of Informatics,
Research Organization of
Information and Systems
2–1–2 Hitotsubashi, Chiyoda-ku,
Tokyo 101–8430, Japan
{kazuya, mo}@nii.ac.jp

Akira Kubota

Interdisciplinary Graduate School of
Science and Engineering,
Tokyo Institute of Technology
4259–G2–31 Nagatsuta, Midori-ku,
Yokohama 226–8502, Japan
kubota@ip.titech.ac.jp

ABSTRACT

This paper describes a method of image generation based on transformation integrating certain sequences of multiple differently focused images. First, we assume that a scene is defocused by a geometrical blurring model. Then we combine spatial frequencies of the scene and the sequence with a 3-D convolution filter that expresses how the scene is defocused on the sequence. The filter can be represented with a linear combination of ray-sets through each point of the lens. Based on the relation, in the 3-D frequency domain we extract each ray-set from the filter as certain frequency components and merge them to reconstruct various filters that can generate images with different viewpoints and blurs.

1. INTRODUCTION

In order to generate a certain image by using multiple differently focused images, conventional methods usually analyze each acquired image independently and merge them into a desired image. For example, detection of focused regions on each image is widely adopted for reconstruction of an all-in-focus image and so on[1, 2, 3]. However, the estimation results are not accurate enough to use as depth information for generating different views with good quality. On the other hand, our previously proposed methods[4, 5] just apply linear filters to each image and merge them for image generation without depth estimation. Unfortunately, the method is not easy to extend for merging very many images because it is almost impossible to design such complicated filters robustly.

In this paper, we propose a novel method for image generation by integrating multiple differently focused images as structured three-dimensional information. The integrated image sequence is analyzed in the frequency domain and transformed to desired images directly without any depth estimation. A three-dimensional filter derived from our geometrical blurring model combines spatial information of the scene and the image sequence with a space-invariant equation using a convolution. By transforming the equation into the frequency domain, we analyze preserved frequency components of the scene on the sequence. Then, we design linear filters that transform the components into various images with different viewpoints and blurs. Some experiments of image generation utilizing synthesized images are shown.

2. GEOMETRICAL TRANSFORMATION BASED ON OUR BLURRING MODEL

Blurs of acquired images are determined by relative locations between traces of a ray-set that comes from a certain point p in the scene and imaging planes, which are orthogonal to z–axis, as shown in Fig.1(a). Here, we express a certain ray before refraction as follows:

$$l_p : (x, y) = z/f \cdot (c, d) + (s, t) \quad (z > 0) \ , \tag{1}$$

where f denotes the focal length of the lens. The ray l_p goes through the point p_1:$(c + s, d + t, f)$ and then it is refracted at the point p_2:$(s, t, 0)$ on the lens plane. We can regard (c, d) as parameters that express the direction of the ray l_p. If the focal point p' is determined by the paraxial theory of Gauss, the refracted ray $l_{p'}$ can be expressed as follows:

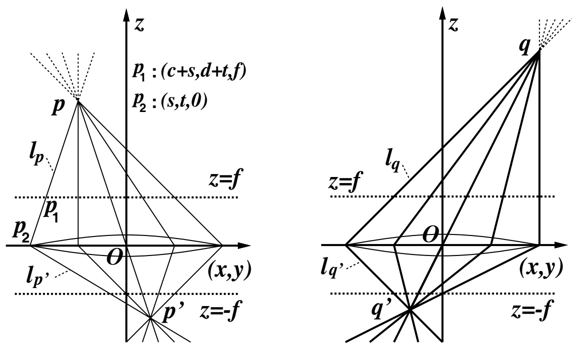
$$l_{p'} : (x, y) = z/f \cdot (c + s, d + t) + (s, t) \quad (z < 0) \ . \tag{2}$$

For aligning the center of blurs on acquired images, we apply following spatial transform T' to the refracted ray-set:

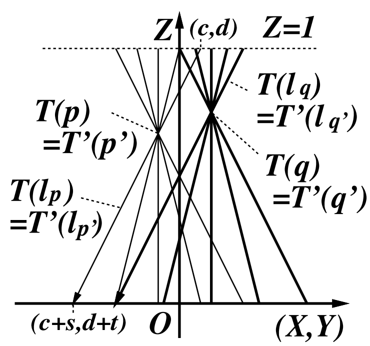
$$T' : (X, Y, Z) = (xf/z, yf/z, -f/z) \quad (z \le -f) \ . \tag{3}$$

The space after the transform T' is denoted by S. Straightness of lines in the original space is preserved through transform T' into the space S. For example, the ray $l_{p'}$ of Eq.(2) is transformed to the ray $T'(l_{p'})$ as follows:

$$(X, Y) = (c, d) + (1 - Z) \cdot (s, t) \quad (0 \le Z \le 1) \ . \tag{4}$$



(a) A ray-set that is focused on p' or q'



(b) The space S after transform (Z–axis is expanded)

Fig. 1. The structure of ray-sets that compose blurs.

First, it is clear that the ray $l_{p'}$, $l_{q'}$ focused on p', q' is transformed to the ray $T'(l_{p'}), T'(l_{q'})$ going through $T'(p'), T'(q')$, respectively. Concerning Eq.(4), we notice that the direction of a certain ray after transform is determined by only (s, t) that is the point on the lens plane where the ray is refracted. The direction does not change by (c, d) that is the original direction of the ray before refraction. Therefore, the rays refracted at a certain point $(s, t, 0)$ on the lens plane are transformed to the rays that are parallel to each other such as $T'(l_{p'}) \parallel T'(l_{q'})$. In the space S after transform T' all of the ray-sets has the same structure as shown in Fig.1(b). The structure does not depend on where the ray-set is focused.

3. SCENE ANALYSIS USING A MULTI-FOCUS IMAGING SEQUENCE

By setting imaging planes at $z = -f/(z_0 + nT_z)$, where z_0, T_z is constant and $n = 0, \ldots, N-1$, we can obtain the information that would correspond to ray-sets acquired on imaging planes $P_0 \sim P_{N-1}$, located at regular distances along Z–axis in the space S. We define $g(X, Y, Z)$ as the acquired three-dimensional information. By the way, the ray before refraction l_p is transformed to the ray $T(l_p)$, which is

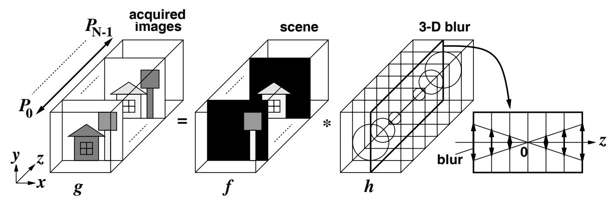


Fig. 2. A three-dimensional blur combines the scene and the acquired images.

also expressed by Eq.(4), by spatial transform T as follows:

$$T : (X, Y, Z) = (xf/z, yf/z, 1 - f/z) \quad (z \geq f) \ . \tag{5}$$

Therefore, the space S after transform T or T' also contains the information of ray-sets in the scene except occluded regions. That is, by extracting the intensity given by each ray-set just where it is focused and calling it $f(X, Y, Z)$, it expresses a three-dimensional information of the scene well (see Fig.2).

Here, we notice that concerning the structured ray-set in the space S, each ray comes from an identical point in the scene. And so we assume that the rays give the same intensity on imaging planes, respectively. Under the assumption, the intensity given by a certain ray-set on acquired images $g(X, Y, Z)$ equals to what is produced from the intensity where the ray-set is focused by applying a 3-D blur. The 3-D blur is determined by the density of the ray-set in the space S, independent of where the ray-sets are focused. Therefore, we can define $h(X, Y, Z)$ as the 3-D blur, which gives an impulse of δ function at the origin as a focal point (see Fig.2).

For simplicity, we rewrite the coordinates of the space S from (X, Y, Z) to (x, y, z). Then, the multi-focus imaging sequence $g(x, y, z)$ composed of acquired images can be expressed by the combination of the three-dimensional information $f(x, y, z)$ of the scene and the 3-D blurring function $h(x, y, z)$ with a convolution as follows:

$$g(x, y, z) = h(x, y, z) * f(x, y, z) \ . \tag{6}$$

4. FREQUENCY ANALYSIS OF THREE-DIMENSIONAL BLURS

The relation derived in the previous section based on a convolution as shown in Fig.2 is a space-invariant equation. In the frequency domain, the convolution is transformed as follows:

$$G(u, v, w) = H(u, v, w)F(u, v, w) \ . \tag{7}$$

Therefore, by analyzing characteristics of $H(u, v, w)$, we are able to know how the multi-focus imaging sequence $g(x, y, z)$ preserves spatial frequency components of the scene $f(x, y, z)$.

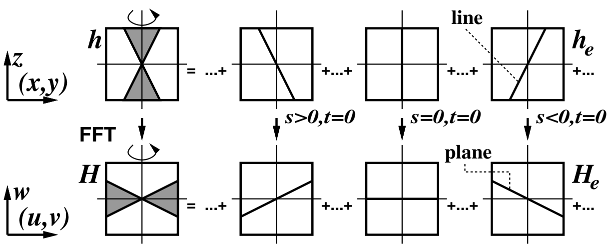
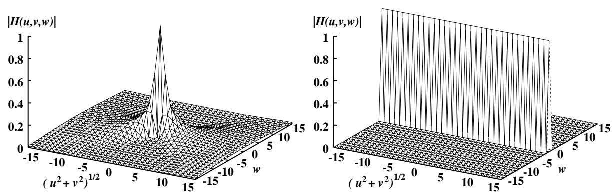


Fig. 3. Feature analysis of three-dimensional blurs based on a decomposition into each ray-set through (s, t).



(a) $L = 1.0$ (b) $L = 0.0$

Fig. 4. Characteristics of three-dimensional blurring filters $h(x, y, z)$ in the frequency domain.

For clarifying the characteristics of the three-dimensional filter $h(x, y, z)$, we decompose it into filters that correspond to each ray-set through (s, t) on the lens plane as shown in Fig.3. These filters are expressed as follows:

$$h_e(x, y, z; s, t) = \delta(x + sz, y + tz) \ . \tag{8}$$

In the frequency domain, it is transformed as follows:

$$H_e(u, v, w; s, t) = \delta(w - (su + tv)) \ . \tag{9}$$

It has impulse components on the plane of $w - (su + tv) = 0$. $H(u, v, w)$ is a linear combination of these planes.

Here, we assume that each ray-set from a certain point in the scene gives Gaussian blurs on imaging planes[4]. Then, the 3-D blurring filter can be expressed as follows:

$$h(x, y, z) = b(x, y; L|z|/\sqrt{2}) \ , \tag{10}$$

and it is transformed into the frequency domain as follows:

$$H(u, v, w) = b(w; L(u^2 + v^2)^{1/2}/\sqrt{2}) \ , \tag{11}$$

where L denotes the radius of the lens and $b(x; \sigma)$, $b(x, y; \sigma)$ denotes 1-D,2-D Gaussian function with the variance of σ^2 (we define $\lim_{\sigma \to 0} b = \delta$), respectively. We show characteristics of $H(u, v, w)$ under the condition $L = 1.0$ or $L = 0.0$ in Fig.4. For simplicity, they are calculated in a discrete manner and normalized. The filter of $L = 0.0$ corresponds to an ideal pin-hole camera, that is, a ray-set going through just the center of the lens. It generates all of acquired images as the same all-in-focus image.

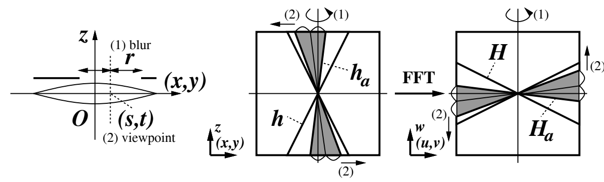


Fig. 5. A virtual lens and its three-dimensional filter.

5. IMAGE GENERATION BY USING A THREE-DIMENSIONAL FILTERING

Here, we would like to acquire images of the same scene by using a virtual lens with a certain aperture, which is different from the one actually used, as shown in Fig.5. By generalizing a linear combination of ray-sets in Fig.3, the filter corresponding to the virtual lens can be expressed as follows:

$$h_a(x, y, z) = b(x + sz, y + tz; r|z|/\sqrt{2}) \ , \tag{12}$$

and it is transformed into the frequency domain as follows:

$$\begin{aligned} &H_a(u, v, w; r, s, t) \\ &\qquad = b(w - (su + tv); r(u^2 + v^2)^{1/2}/\sqrt{2}) \ . \end{aligned} \tag{13}$$

We can use parameters (s, t) and r that determine virtual viewpoints and blurs. If the aperture area is limited in the original one, without artifacts introduced from the components where $|H(u, v, w)| \simeq 0$, we are able to generate an image sequence $a(x, y, z)$ acquired by the virtual lens as follows:

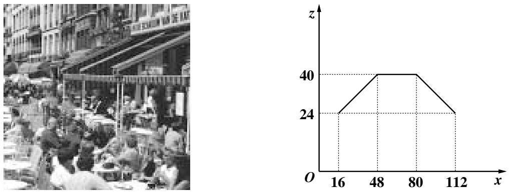
$$\begin{aligned} &A(u, v, w) \\ &\quad = H_a(u, v, w; r, s, t)F(u, v, w) \\ &\quad = H_a(u, v, w; r, s, t)H^{-1}(u, v, w)G(u, v, w) \ , \end{aligned} \tag{14}$$

where $A(u, v, w)$ denotes $a(x, y, z)$ in the frequency domain. We can consider $H_a(u, v, w; r, s, t)H^{-1}(u, v, w)$ as a single filter, and so the desired sequence $a(x, y, z)$ can be generated from the acquired sequence $g(x, y, z)$ directly by using three-dimensional filtering without any depth estimation.

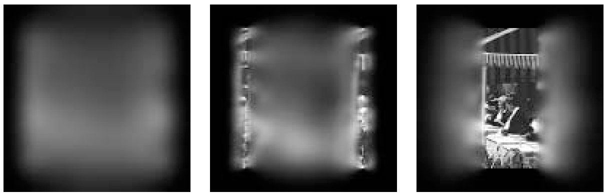
6. EXPERIMENTS

We assume that a scene has a certain texture and various depths as shown in Fig.6(a). The coordinates of assumed depths are in the space S. Then, we synthesize a multi-focus imaging sequence $g(x, y, z)$, which consists of 64 images as shown in Fig.6(b). Each image has 128×128 pixels and the sequence is structured with $L = 1.0$.

We can generate a sequence that consists of images whose blurs are suppressed by applying the above-mentioned single filter of $r = 0.2$ and $(s, t) = (0, 0)$ to the multi-focus imaging sequence $g(x, y, z)$ as shown in Fig.7. It is notable that the images are robustly generated and not interfered by frequency components where $|H(u, v, w)| \simeq 0$.



(a) A texture and assumed depths for synthesis of (b)



(b) A synthesized multi-focus imaging sequence
($z = 16, 28, 40$)

Fig. 6. A texture and depths (at arbitrary y) of the scene and synthesized images.

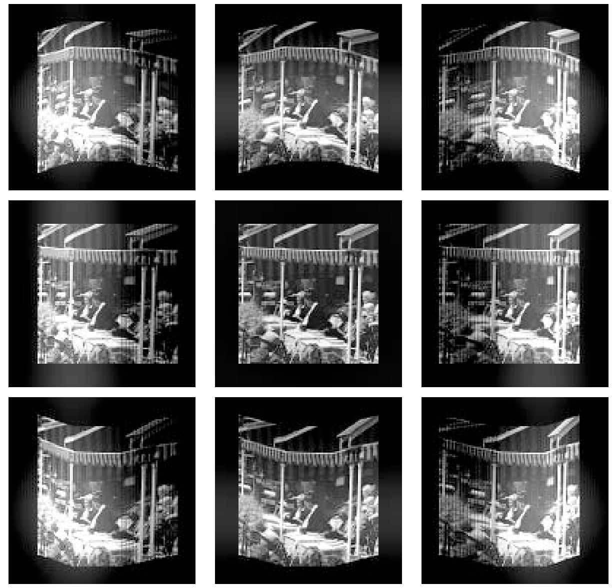
Fig. 7. Generated images ($z = 16, 28, 40$) with suppressed blurs by using a three-dimensional filtering.

We also generate free viewpoint images by applying filters of various (s, t) to the original sequence $g(x, y, z)$ as shown in Fig.8. In Fig.8(a), we use the filter of $r = 0.0$ and various (s, t) and show just $a(x, y, 32)$ of each generated sequence, and in Fig.8(b), a certain (s, t) and $r = 0.2$ are adopted. The results indicate that we success in giving various disparities, although we need some correction for the intensity of regions where some duplications or lacks of depths occur.

7. CONCLUSION

In this paper, we proposed a novel method of image generation integrating multiple differently focused images. Based on the spatial frequency analysis of a structured multi-focus imaging sequence by using a three-dimensional filter, free viewpoint, iris and focus images are directly generated from the sequence without any depth estimation.

In the future, we would like to analyze actual blurs of ordinary lenses in comparison of our blurring model for applying this method to real images. We have already obtained preliminary results about all-in-focus images of real images[6].



(a) Free viewpoint images
(an ordinary all-in-focus image is at the center)

(b) Images with a certain viewpoint and suppressed blurs
($z = 16, 28, 40$)

Fig. 8. Free viewpoint, iris and focus image generation by using a three-dimensional filtering.

8. REFERENCES

[1] P.J.Burt, “A Gradient Pyramid Basis for Pattern-Selective Image Fusion,” Proc. SID, pp.467-470, 1992

[2] P.J.Burt, and R.J.Kolczynski, “Enhanced Image Capture Through Fusion,” Proc. 4th ICCV, pp.173-182, 1993

[3] M.Subbarao, T.-C.Wei, and G.Surya, “Focused Image Recovery from Two Defocused Images Recorded with Different Camera Settings,” IEEE *Trans. on Image Processing*, Vol.4, No.12, pp.1613-1627, 1995

[4] K.Kodama, K.Aizawa, and M.Hatori, “Generation of arbitrarily focused images by using multiple differently focused images,” SPIE *Journal of Electronic Imaging*, Vol.7, No.1, pp.138-144, 1998

[5] A.Kubota, and K.Aizawa, “A novel image-based rendering method by linear filtering of multiple focused images acquired by a camera array,” Proc. ICIP, Vol.III, pp.701-704, 2003

[6] K.Kodama, H.Mo, and A.Kubota, “All-in-Focus Image Generation by Merging Multiple Differently Focused Images in Three-Dimensional Frequency Domain,” Pacific-Rim Conference on Multimedia 2005, Part I, pp.303-314, 2005