SIMPLE AND FAST ALL-IN-FOCUS IMAGE RECONSTRUCTION BASED ON THREE-DIMENSIONAL/TWO-DIMENSIONAL TRANSFORM AND FILTERING

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 deals with all-in-focus image reconstruction by merging multiple differently focused images. We previously proposed a method of generating an all-in-focus image from multi-focus imaging sequences based on a 3-D filtering. In this paper, we first combine the sequence into a 2-D image. Just by applying a 2-D filter to the image, we realize fast reconstruction of all-in-focus images robustly. In order to reduce the cost of image acquisition, the optimal number of multiple differently focused images is also discussed. In addition, we introduce a simple estimation method utilizing the 2-D filter for the parameter of 3-D blurs. We show experimental results of fast all-in-focus image reconstruction by using synthetic and real images.

Index Terms— Image restoration, image reconstruction

1. INTRODUCTION

In order to generate a certain image such as an all-in-focus image by using multiple differently focused images, conventional methods[1, 2, 3, 4] usually analyze each acquired image independently and merge them into a desired image. These methods are not easy to extend for merging very many images. Therefore, we deal with various methods of integrating the multiple differently focused images into desired images directly[5, 6].

We previously proposed a method of all-in-focus image reconstruction by applying a 3-D filter to the multi-focus imaging sequences[6]. However, the cost of such a 3-D filtering is not inexpensive, if the number of multiple differently focused images or the image size increases. In this paper, we derive a novel 2-D relation between the multi-focus imaging sequence and the desired image. Based on the relation, simple and fast all-in-focus image reconstruction is realized robustly even just by using a 2-D filter instead of the previous 3-D one.

In addition, in order to reduce the cost of image acquisition, the optimal number of original multiple differently focused images is discussed. We also introduce a simple estimation method utilizing the 2-D filter for the parameter of 3-D blurs.

We show experimental results of simple and optimal all-in-focus image reconstruction by using synthetic and real images. The novel method realizes both high-speed and robustness.

2. SCENE ANALYSIS USING A MULTI-FOCUS IMAGING SEQUENCE

We previously derived the relation between spatial information of a scene and a sequence of multiple differently focused images on certain imaging planes $P_0 \sim P_{N-1}$ by using a convolution of a 3-D blurring filter as shown in Fig.1 under a geometrical blurring model[6]. Let the filter be denoted by $h(x, y, z; r)$, the relation can be expressed as follows:

$$g(x, y, z) = h(x, y, z; r) * f(x, y, z) \ , \tag{1}$$

where $f(x, y, z)$ corresponds to the scene through a geometrical transformation and $g(x, y, z)$ consists of the acquired multi-focus imaging sequence after image size correction, respectively. The filter $h(x, y, z; r)$ expresses how the scene is defocused in the sequence, where r corresponds to the radius of the iris. In the frequency domain, the convolution is transformed as follows:

$$G(u, v, w) = H(u, v, w; r)F(u, v, w) \ . \tag{2}$$

By analyzing $H(u, v, w; r)$, we can know how the sequence $g(x, y, z)$ preserves spatial frequencies of the scene $f(x, y, z)$.

3. FAST ALL-IN-FOCUS IMAGE GENERATION

Let 1-D and 2-D Gaussian blurs with the variance of σ^2 be denoted by $p(x; \sigma)$ and $p(x, y; \sigma)$, respectively (we define

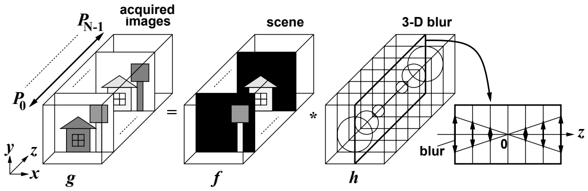
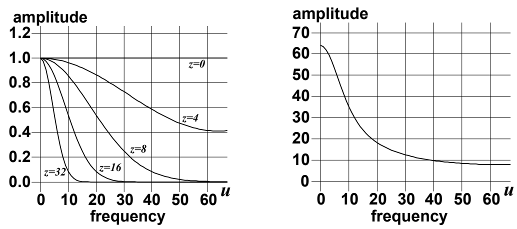


Fig. 1. A 3-D blur combines the scene and acquired images.



(a) $h(x, y, z)$ (b) $c(x, y)$

Fig. 2. Characteristics of $h(x, y, z)$ and $c(x, y)$ in the frequency domain ($r = 0.2$ and $N_x = N_y = 128, N = 64$).

$\lim_{\sigma \to 0} p = \delta$). Here, we replace a geometrical blur, the radius of which is R, with a Gaussian blur of $\sigma = R/\sqrt{2}$ [1]. Then, the 3-D blur can be expressed as follows:

$$h(x, y, z; r) = p(x, y; r|z|/\sqrt{2}) \ , \tag{3}$$

$$H(u, v, w; r) = Np(w; r|s|/\sqrt{2}) \ , \tag{4}$$

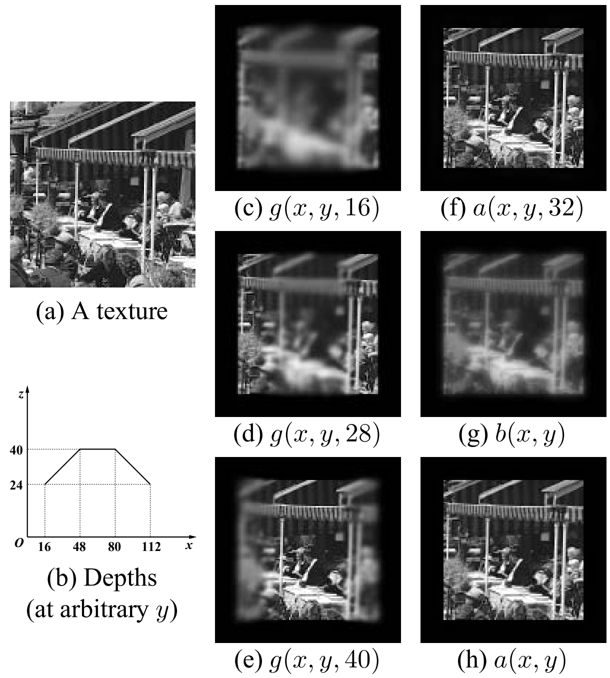
where $s^2 = {K_x}^2 u^2 + {K_y}^2 v^2$, if the corrected size of images is denoted by (N_x, N_y) and $(K_x, K_y) = (N/N_x, N/N_y)$. For simplicity, we rewrite $H(r) = H(u, v, w; r)$. We previously proposed a method of generating an all-in-focus image $a(x, y, z)$ (any z will do) from the multi-focus imaging sequence $g(x, y, z)$ without any scene estimation of $f(x, y, z)$ as follows:

$$\begin{aligned} A(u, v, w) &= H(0.0)F(u, v, w) \\ &= H(0.0)H^{-1}(r)G(u, v, w) \ , \end{aligned} \tag{5}$$

where $A(u, v, w)$ denotes $a(x, y, z)$ in the frequency domain. Interestingly, we can define $H(0.0)H^{-1}(r)$ above as a single filter that uniquely exists and remains robust for all (u, v, w). Therefore, all the frequency components of $a(x, y, z)$ are reconstructed from the multi-focus imaging sequence $g(x, y, z)$ by the 3-D filtering of Eq.(5).

The method obtains good results very robustly in comparison with ordinary all-in-focus image reconstruction from a single degraded image by a 2-D deconvolution. However, the cost of such a 3-D filtering is not inexpensive, if the number of multiple differently focused images or the image size increases. In order to reduce the cost, here, we introduce a novel method of replacing the 3-D filtering of Eq.(5) with a certain 2-D one. First, we define 2-D information as follows:

$$a(x, y) = \int f dz \ , \quad b(x, y) = \int g dz \ , \quad c(x, y) = \int h dz \ . \tag{6}$$



(a) A texture
(b) Depths (at arbitrary y)
(c) $g(x, y, 16)$
(d) $g(x, y, 28)$
(e) $g(x, y, 40)$
(f) $a(x, y, 32)$
(g) $b(x, y)$
(h) $a(x, y)$

Fig. 3. Simulations of all-in-focus image reconstruction: (a),(b) the assumed scene, (c)∼(e) a synthetic multi-focus imaging sequence, (f)∼(h) reconstruction results of 3-D and 2-D filtering methods.

Then, we obtain the following relation from Eq.(1):

$$b(x, y) = c(x, y) * a(x, y) \ , \tag{7}$$

$$B(u, v) = C(u, v)A(u, v) \ , \tag{8}$$

where, for example, $B(u, v)$ denotes $b(x, y)$ in the frequency domain. We notice that $a(x, y)$ equals an all-in-focus image. We show the characteristics of $h(x, y, z)$ and $c(x, y)$ in the frequency domain for $r = 0.2$ and $N_x = N_y = 128, N = 64$ in Fig.2. Since $C(u, v)$ is always larger than 1 as shown in Fig.2(b), we can reconstruct an all-in-focus image $a(x, y)$ robustly as follows:

$$A(u, v) = C^{-1}(u, v)B(u, v) \ , \tag{9}$$

where $C^{-1}(u, v)$ uniquely exists and remains robust for all (u, v). The cost of such a 2-D filtering is much smaller than the cost of the previous 3-D one.

4. SIMULATIONS USING SYNTHETIC IMAGES

We assume a scene that has a certain texture and various depths as shown in Fig.3(a) and (b) in order to synthesize a multi-focus imaging sequence $g(x, y, z)$ in Fig.3(c)-(e), where the center regions in acquired images are far from a viewpoint. The sequence consists of 64 images having 128×128 pixels and it is structured with $r = 0.2$.

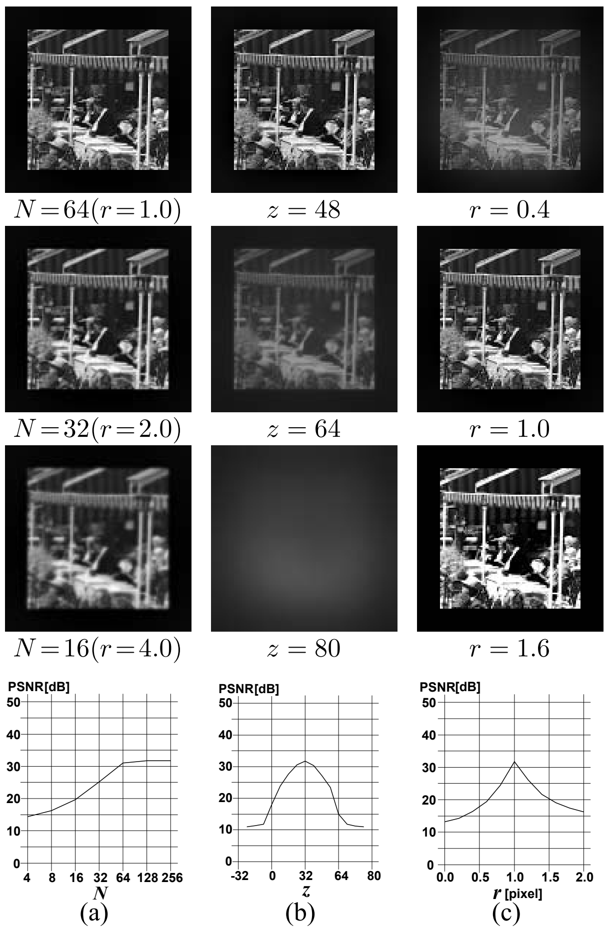
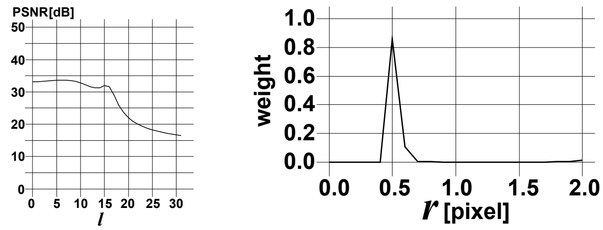


Fig. 4. Some conditions achieving good quality of reconstructed all-in-focus images: (a) the number of acquired images, (b) depths where the scene exists, (c) difference between estimated r and real $r = 1.0$.



(a) Reducing acquired images (b) Estimation of r

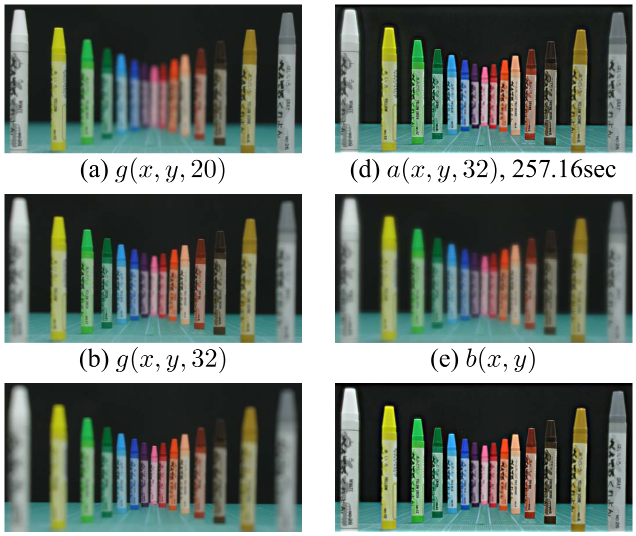
Fig. 5. Estimation of the optimal image acquisition and the 3-D blur parameter r.

We reconstruct an all-in-focus image in Fig.3(f) by applying the previous 3-D filter to the sequence as Eq.(5). Based on the novel 2-D filtering of Eq.(9), we can also reconstruct an all-in-focus image in Fig.3(h) after combining the multi-focus imaging sequence $g(x, y, z)$ into a degraded image $b(x, y)$ in Fig.3(g). We notice that the degraded image has fine quantization step size of 64 times.

In order to estimate the quality of reconstructed images, we compare them with the original texture to calculate PSNR. Actually, 64×64 pixles in the center are used for the comparison. By using Pentium-III 866MHz on IBM ThinkPad X23, the previous 3-D filtering spends 7.80 seconds with given $H(0.0)H^{-1}(r)$ to gain 36.77dB, and the novel 2-D filtering spends 0.21 seconds with given $C^{-1}(u, v)$ to gain 34.86dB. The results show that the 2-D method realizes all-in-focus image reconstruction much faster than the previous one and gains almost the same quality without any visible artifacts.

Next, we discuss the optimal number of original acquired multiple differently focused images for good quality of reconstructed images. Here, the 3-D filter is used to estimate the robustness of our approach itself. Even if the iris is fixed, when we decrease the number of acquired images by setting image planes in Fig.1 sparsely, r becomes larger. The quality variation of reconstructed images against the number of acquired images N is shown in Fig.4(a). We assume that the iris is fixed where $r = 1.0$ for $N = 64$. We notice that generated images are not clearly focused against smaller N, that is, larger r. We should set image planes in Fig.1 more closely in order to prevent r from getting too large. At the same time, as shown in Fig.4(a), where $N > 64$, that is, $r < 1.0$, the improvement of the quality is very little. Therefore, the number of acquired images N should be determined where r is close to 1.0. In Fig.4(b), we show the quality variation against depths where the scene exists. In this simulation, we set $N = 64, r = 1.0$ and assume that the scene exists at a constant depth z. In case of $z = 32$, which is the center of focused depths, the best quality is obtained. Reconstructed images of $z = 32 - \alpha$ and $z = 32 + \alpha$ have the same quality. In case of $|\alpha| > 32$, the depth is not in focus on any acquired image, and so reconstructed images are not meaningful. On the other hand, the depth of $|\alpha| < 16$ can be clearly reconstructed enough. Moreover, we notice that if the scene exists at $z \geq l$, $g(x, y, z)$ of $z < l$ can be predicted well from $g(x, y, l)$ by applying an ordinary 2-D blurring filter. That is to say, if the scene exists at $l \leq z \leq m$, we need to acquire just $g(x, y, z)$ of $l \leq z \leq m$ in Fig.1. The other images can be synthesized from $g(x, y, l)$ and $g(x, y, m)$. In Fig.5(a), by setting $N = 64$, $r = 0.2$ and assuming that the scene exists at $16 \leq z \leq 48$ like Fig.3(b), we show the quality variation of reconstructed images against the number of synthesized images from $g(x, y, l)$ and $g(x, y, 63 - l)$ instead of the original acquired images. It is notable that even just acquired images that are in focus on $16 \leq z \leq 48$ ($l = 16$) can realize good quality of reconstruction. The other images can be synthesized from the acquired images well for integration and reconstruction and do not need to be actually acquired.

By the way, if we do not know camera parameters, we must estimate r from acquired images themselves. The quality variation of reconstruction against difference between es-



(a) $g(x,y,20)$ (d) $a(x,y,32)$, 257.16sec

(b) $g(x,y,32)$ (e) $b(x,y)$

(c) $g(x,y,44)$ (f) $a(x,y)$, 7.05sec

Fig. 6. Reconstruction of an all-in-focus image from real images: (a)∼(c) an acquired multi-focus imaging sequence, (d)∼(f) results of 3-D and 2-D filtering methods.

timated r and real $r = 1.0$ is shown in Fig.4(c). They have insufficient or too strong contrasts against smaller r or larger r, respectively. If the estimation error is smaller than about 20%, they almost preserve the quality. In the previous paper, for the estimation of r, we apply the 3-D filter $H(0.0)H^{-1}(r)$ of various r to the sequence and generate images that would be all-in-focus images as Fig.3(f) when r is correct. In this paper, we use the novel 2-D filter $C^{-1}(u,v)$ of various r instead of the 3-D filter for fast estimation. One of these generated images, that is denoted by $\bar{a}(r)$, is compared with the original sequence (we rewrite $g_z(x,y) = g(x,y,z)$). First, at each pixel appropriate r is estimated as follows:

$$\bar{r}(x,y) = \max_z \arg\min_r \left| \frac{\sigma^2_{g_z}(x,y) - \sigma^2_{\bar{a}(r)}(x,y)}{\sigma^2_{g_z}(x,y) + \sigma^2_{\bar{a}(r)}(x,y)} \right| , \quad (10)$$

where $\sigma^2_I(x,y)$ means the variance of intensity in the regions of 5×5 pixels around (x,y) of the image I. Then, we estimate final r using the weighted mean of them as follows:

$$r = \sum_{x,y} \sigma^2_{\bar{a}(\bar{r})}(x,y)\bar{r}(x,y) \Bigg/ \sum_{x,y} \sigma^2_{\bar{a}(\bar{r})}(x,y) \ . \quad (11)$$

For the simulation, we apply this estimation method to the sequence of $r = 0.5$. We show the weight against r in Fig.5(b). Finally, $r = 0.55$ is obtained and its error is smaller than 20%.

5. EXPERIMENTS USING REAL IMAGES

We experiment using a multi-focus imaging sequence of real images as shown in Fig.6(a)-(c). Each image has 512×256 pixels and the sequence consists of 64 images. Based on our previously proposed method[5], we correct the size of actually acquired images to be fit to the condition described in Sect.2. The 3-D blur parameter of the corrected sequence is estimated as $r = 0.36$. We reconstruct an all-in-focus image in Fig.6(d) by applying the previous 3-D filter to the sequence.

By integrating the sequence $g(x,y,z)$ into a degraded image $b(x,y)$ in Fig.6(e), we also reconstruct an all-in-focus image in Fig.6(f) with the novel 2-D filter. Actually, we use subsampled 16 images from $g(x,y,z)$ of $16 \le z < 48$ with 16 predicted images ($l = 8$) and set $r = 0.72$ for the 2-D method. The previous method spends 257.16 seconds and the novel one spends 7.05 seconds. The novel one realizes all-in-focus image reconstruction much faster without awful artifacts.

6. CONCLUSION

In this paper, we proposed a simple and fast reconstruction method of all-in-focus images using a novel 2-D filter from acquired multi-focus imaging sequences. The optimal number of acquired images that reduces the cost of image acquisition and preserves the quality of reconstruction is also discussed. In addition, the 2-D filter can estimate the 3-D blur of the sequence much faster than the 3-D method.

In the future, we would like to extend the method for fast reconstruction of free viewpoint images[6]. The hardware implementation to realize our proposed method with real-time speed for all-in-focus video[3] will be also considered.

7. REFERENCES

[1] M.Subbarao, N.B.Agarwal, and G.Surya, "Application of Spatial-Domain Convolution/Deconvolution Transform for Determining Distance from Image Defocus," Computer Vision Laboratory, Stony Brook, Tech.Report 92.01.18, 1992

[2] 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

[3] K.Ohba, et al., "Microscopic vision system with all-in-focus and depth images," Machine Vision and Applications, Vol.15, No.2, pp.55-62, 2003

[4] H.A.Eltoukhy and S.Kavusi, "A Computationally Efficient Algorithm for Multi-Focus Image Reconstruction," Proc. SPIE Electronic Imaging, pp.332-341, 2003

[5] 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

[6] K.Kodama, H.Mo, and A.Kubota, "Free Viewpoint, Iris and Focus Image Generation by Using a Three-Dimensional Filtering based on Frequency Analysis of Blurs," Proc. ICASSP, Vol.II, pp.625-628, 2006