Depth Measurement by the Multi-Focus Camera

Shinsaku Hiura and Takashi Matsuyama
Department of Intelligence Science and Technology
Graduate School of Informatics, Kyoto University
Sakyo, Kyoto, 606-8501 JAPAN

Abstract

In this paper, we first introduce the multi-focus camera, a new image sensor used for depth from defocus (DFD) range measurement. It can capture three images with different focus values simultaneously. We then propose two different depth measurement methods using the camera. The first method, an augmented version of the one proposed in [3], employs a non-iterative optimization process to compute depth values on edge points. The second one incorporates a coded aperture with the camera and applies model-based pattern matching to estimate depth values of textured surfaces. Here we propose two types of coded apertures and corresponding analysis algorithms: 1D Fourier analysis to acquire a depth map and a blur-free image from three defocused images taken with a pair of pinholes, and 2D convolution based model matching for the fast and precise depth measurement using a coded aperture with four pinholes. Experimental results showed that the multi-focus camera works well as a practical DFD range sensor and that the coded apertures much improve its range estimation capability for real world scenes.

1 Introduction

Depth from defocus(DFD)[1][2][3][4] is a depth measurement method with various advantages: 1) passive, 2) single lens optics, and 3) no correspondence problem. These inherent characteristics will allow the realization of compact and portable range sensors for real world scenes.

From a practical point of view, however, DFD requires unrealistic fine tuning of optical systems and their precise modeling, since DFD analyzes very delicate blurring phenomena. For example, Nayar's system[5] requires 1/10 pixel alignment between multiple defocused images. Such precise physical adjustments make it difficult to develop practical DFD range sensors and moreover, the sophisticated modeling degrades the robustness of the measurement.

Our idea to overcome these difficulties and to develop practically usable DFD range sensors is as follows. While almost all DFD methods so far proposed focused on how accurately we can model *given* blurring phenomena, we should *design* them to realize robust and reliable DFD range sensors. That is, if we can easily design blurring phenomena, more robust and accurate DFD methods can be realized by using prominent characteristics of designed blurring patterns.

In this paper, we propose two novel ideas for the practical DFD range sensor development. First we introduce the *multi-focus camera*, a new image sensor for the DFD range measurement, which can capture three images with different focus values simultaneously. It is fabricated from an ordinary 3CCD video camera and outputs three defocused images to standard RGB video channels respectively. This allows the real time depth measurement by ordinary video image processors. In order to test its practical utilities, we augmented the DFD method proposed in [3]. It employs a non-iterative optimization process to compute depth values on edge points from three defocused images. Experimental results demonstrated practical utilities of the camera and the algorithm.

In the latter half of the paper, we incorporate a coded aperture with the multi-focus camera to design blurring patterns suitable for the DFD range measurement. Here we show two types of aperture designs and propose model-based pattern matching methods to measure depth values of textured surfaces. The first one uses an aperture with a pair of pinholes. This aperture design allows us to use 1D FFT for the fast depth measurement. We developed Fourier domain algorithms for the precise depth measurement and the blur-free image reconstruction. The second method employs an aperture with four pinholes and estimates a depth map by a pattern matching method based on the commutative law of the 2D convolution. In both methods, the depth information can be measured reliably due to the fact that structured blurring patterns preserve information in high spatial frequencies. Experimental results demonstrated their effectiveness.

2 Multi-Focus Camera

One important requirement to realize practical DFD range sensors is to develop compact and portable optical systems which can capture multiple images with different focus values simultaneously.

To satisfy this requirement, we reformed an ordinary 3CCD color video camera in the following two points:

1. The dichroic prism block is re-coated so that each CCD obtains just $\frac{1}{3}$ of the incoming light in all spectral bands.
2. One of the three CCDs is shifted about 1mm forward along the optical axis, and another about 1mm backward (Figure 1 Left). These spatial shifts of CCDs make captured images defocused. Note that

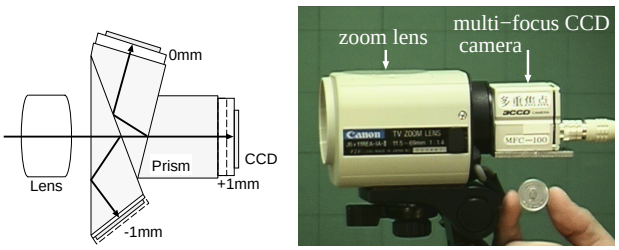

Figure 1: Multi-focus camera

the degree of the shift, i.e. about 1mm, was determined based on extensive experiments to model characteristics of the optical system as well as to design measurable depth ranges. Note also that all CCDs are tightly glued to the prism to make the sensor durable in practical use. Since the gluing process degrades the alignment between captured defocused images, several different types of calibrations should be conducted to model the camera[7].

The right picture in Figure 1 shows an overview of the developed multi-focus camera with a zoom lens. Since the camera outputs three defocused images to standard RGB video channels respectively, the real time depth measurement can be realized.

3 Edge and Depth from Focus by the Multi-Focus Camera

In order to test practical utilities of the multi-focus camera, we augmented the depth measurement method proposed in [3]. It employs a non-iterative optimization process to compute depth values on edge points from three defocused images. Due to space limitations, we will give just an outline of the method. (See [7] for details.)

3.1 Blurring Characteristics in the Spatio-Focal Image

Suppose we want to measure the depth of a locally straight step edge on an object surface located at u from the lens, whose focal length is f (Figure 2). Let w denote a focus position (i.e. location of the image plane) and v the focused image plane position satisfying $\frac{1}{f} = \frac{1}{u} + \frac{1}{v}$. Let I_1 and I_2 ($I_1 > I_2$) denote observed intensity values of the two regions delineated by the edge respectively.

A set of multi-focus images can be modeled as a sequence of images taken at different image plane positions ws, which define a three dimensional images in the $x-y-w$ coordinate system (Figure 2), where $x-y$ stands for the image coordinate system. In this 3D space, we define the plane ϕ which includes an edge point E on the object surface, the optical center O, and the line perpendicular to the edge. We also define the plane μ which includes the optical axis and shares the lens diameter with ϕ. The *spatio-focal image* is the orthogonal projection of the image (i.e. intensity distribution) on ϕ onto the plane μ.

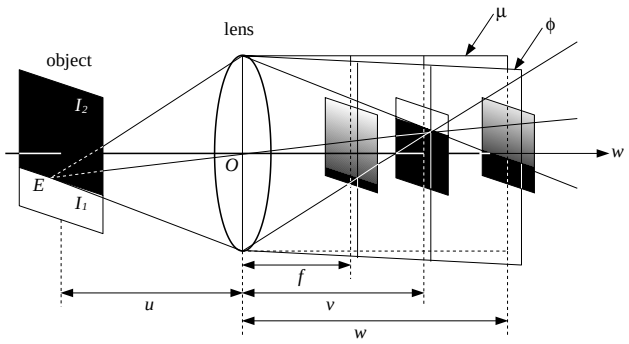

Figure 2: Multi-focus image space of a linear step edge

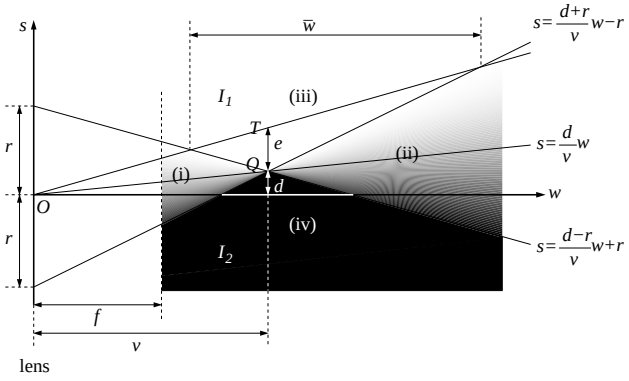

Figure 3: Spatio-focal image

Assuming an isotropic blur kernel, the spatio-focal image is divided into four areas by the pair of straight lines as shown in Figure 3, where d denotes the distance between the focused edge point and the optical axis, and r the radius of the iris. Suppose we transform the spatio-focal image so that the main ray (i.e. line connecting the centers of blurred images of the edge point) becomes parallel to the optical axis. We call the transformed image the *aligned spatio-focal image* and can prove that its intensity distribution satisfies the following symmetric properties: iso-intensity lines of $I_1 - \delta$ and $I_2 + \delta$ are symmetric with respect to the horizontal and vertical lines passing through the focused edge point.

3.2 Edge Detection and Depth Measurement Algorithms

Give a set of three defocused images taken by the multi-focus camera, we first extract edge points. Here w_1, w_2, and w_3 denote three focus positions at which the defocused images are taken.

1. Adjust the size of each observed image depending on w_i.
2. Compute the intensity variance at each pixel position using the three normalized images.

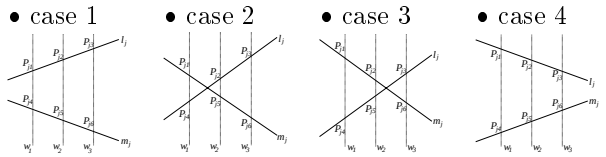

Figure 4: Relations about w_1, w_2, w_3 and v

3. Extract as edge points those pixels with locally minimum intensity variances.

Note that since all observed images are out of focus, we cannot use ordinary edge detectors.

Then, compute the spatio-focal image for each edge point. There are four possibilities about where the focused edge point is located in the spatio-focal image: each case in Figure 4 illustrates three vertical lines denoting the observed intensity profiles at w_1, w_2, and w_3 and iso-intensity lines of $I_1 - \delta$ and $I_2 + \delta$.

We introduce case classification variables σ_1, σ_2, and σ_3, which are defined as follows:
1. When $v_0 < w_1$, $\sigma_1 = \sigma_2 = \sigma_3 = 1$.
2. When $w_1 < v_0 < w_2$, $\sigma_1 = -1$œB!œ(B$\sigma_2 = \sigma_3 = 1$.
3. When $w_2 < v_0 < w_3$, $\sigma_1 = \sigma_2 = -1$œB!œ(B$\sigma_3 = 1$.
4. When $w_3 < v_0$, $\sigma_1 = \sigma_2 = \sigma_3 = -1$.

To realize real-time depth computation, we introduce the following two assumptions:
1. The intensity profile at w_i changes linearly in the range delimited by the pair of iso-intensity lines.
2. The gradient (a_i) and the bias (b_i) of the linear intensity profile at w_i can be estimated from the observed image at w_i.

With these assumptions, the focused position v can be calculated analytically by LMS method:

$$v = \frac{\sum_{i=1}^{3} \sigma_i b_i w_i \sum_{i=1}^{3} b_i^2 w_i - \sum_{i=1}^{3} \sigma_i b_i \sum_{i=1}^{3} b_i^2 w_i^2}{\sum_{i=1}^{3} b_i^2 \sum_{i=1}^{3} \sigma_i b_i w_i - \sum_{i=1}^{3} \sigma_i b_i \sum_{i=1}^{3} b_i^2 w_i} \qquad (1)$$

The case classification and the depth (i.e. u) computation are done as follows:
1. Compute the residual of LMS method for each case.
2. Select the case that gives the least residual.
3. Assign those values to σ_1, σ_2, and σ_3 that specify the selected case and compute v by equation 1.
4. Finally, calculate the depth of the edge point u from v using $\frac{1}{f} = \frac{1}{u} + \frac{1}{v}$.

3.3 Experimental Results

Here we show two experimental results to test the performance of the multi-focus camera and the depth measurement algorithm. We used Canon J6x11REA-IA-II zoom lens, whose focal length is set to 69mm (Figure 1).

The first experiment is for testing the accuracy of the depth measurement. The object is a flat checker pattern whose grid size is 40mm (Figure 5 Left) and

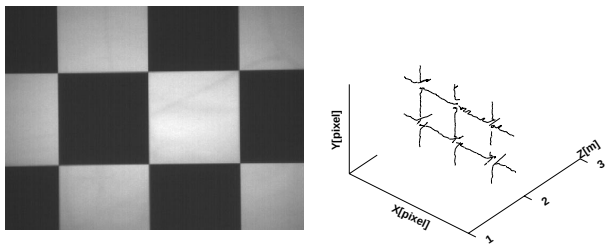

Figure 5: Depth measurement of a checker pattern plane

Table 1: Statistics of the depth measurement for the checker pattern in Figure 11

Real Distance	1900 mm
Mean of measured depth values	1891.4mm
Std. Dev. of measured depth values	24.8mm
Std. Dev. of measured depth values excluding corner points	14.9mm
(percentile of the outliers)	15.8%

is located at 1900mm from the camera. The right 3D plot in Figure 5 and Table 1 summarize the experimental result. Except for the corner points, where the linear edge assumption is not satisfied and hence the accuracy decreases, the accurate depth measurement is attained; the standard deviation of measured depth values is sufficiently small. The computation time is about 0.45 seconds by DEC Alpha 400MHz workstation without special hardware.

Since the multi-focus camera is nothing but a camera, we can incorporate various types of lens systems as well as adjust focus and iris values depending on depth measurement requirements. This means that we can realize a very wide range of depth measurement using the camera. We conducted several experiments to demonstrate its wide range depth measurement capability. In these experiments we used as an object a plane with a single straight step edge. Figure 6 illustrates the accuracy and measurable range of depth using the following camera parameters: focal length: 69mm, iris: open, focus value: the central CCD is focused at about 350cm. Figure 7 illustrates how the measurable depth range changes depending on the focus value of the lens while keeping focal length: 69mm, iris: open. The horizontal axis denotes the focus value and the shaded area stands for the depth range measurable with reasonable accuracy (within 5% relative errors). We also conducted similar experiments with a 235mm zoom lens, which showed that we can measure the depth from about 5m to 20m.

4 Coded Aperture for DFD

The coded aperture[9][10] was originally proposed for special imaging devices such as X-ray telescopes. Here we show that it can improve the performance of

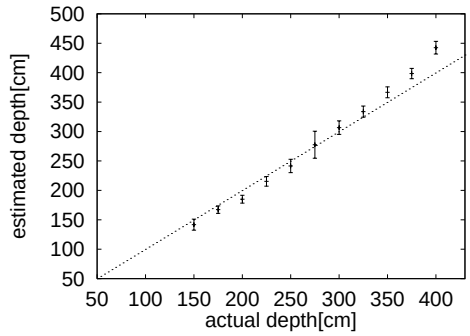

Figure 6: Measurable depth range with fixed camera parameters

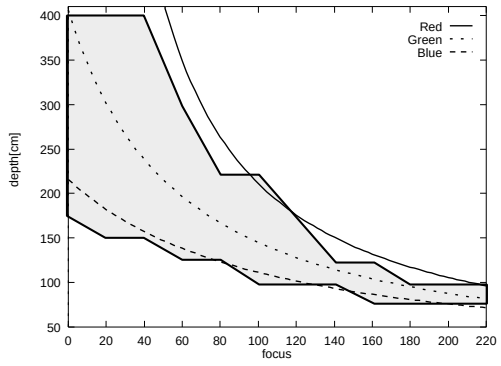

Figure 7: Measurable depth range vs focus value

the DFD range measurement.

4.1 Designing Blurring Patterns by Coded Apertures

Most studies about DFD[1][2][5][4][6] start with the assumption of the circular aperture, because ordinary photographic lenses have a circular iris for isotropic smooth blurring. The blurring effect caused by the circular aperture works as a low-pass filter whose cut-off frequency depends on the radius of the circle. This effect is so delicate that DFD methods using the circular iris require both carefully tuned optical systems and the very precise modeling of blurring phenomena to accurately analyze defocused images. For example, Watanabe[6] prepared defocused images without noise by averaging 256 images, and Nayar[5] referred even to diffraction effects of the optics. Such naive characteristics of the isotropic blurring make it difficult to develop robust DFD range sensors for real world scenes.

Our idea to develop robust DFD methods is to improve characteristics of blurring by incorporating a coded aperture. That is, by structuring the aperture shape, we can *design* blurring characteristics suitable

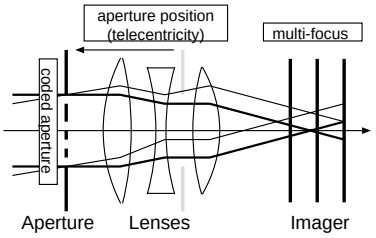

Figure 8: Multi-focus camera with a coded aperture

for the DFD range measurement. For example, the aperture pattern can be designed to work as high-pass or band-bass filters which preserve the useful information for the depth measurement. We believe that the essence of DFD does not rest in the precise modeling of blurring phenomena but in the design of useful blurring phenomena.

4.2 Multi-Focus Camera with a Coded Aperture

Figure 8 illustrates the Multi-Focus camera with a coded aperture. The aperture is placed at the front-focal plane of the lens to realize the telecentric optics; telecentric optics eliminates the image magnification caused by defocusing[5][6].

In this optical system, observed multi-focus images $i_m(x,y)(m = 1, \cdots, N)$ can be described by the following convolution:

$$i_m(x,y) = \frac{1}{k_m^2} a(\frac{x}{k_m}, \frac{y}{k_m}) * s(x,y), \tag{2}$$

where $a(x,y)$ denotes a blur kernel, $s(x,y)$ a scene image without blur, and $x-y$ stands for the coordinate system parallel to the image plane. Basically, the blur kernel is defined as a scaled pattern of the aperture. Using the Gaussian lens law, its scale parameter k_m can be computed by

$$k_m = \frac{v - w_m}{f}. \tag{3}$$

This equation implies that the scale parameter k_m changes depending on the focused position v. When the image plane is just in focus (i.e. $v = w_m$), the blur kernel is reduced to the delta function.

If we can estimate the scale parameter k_m, the object depth and the original blur-free image $s(x,y)$ will be easily recovered. With the multi-focus camera, we can acquire three images with different scale parameters. Thus, the problem to be solved is how we can design an aperture pattern which allows the stable estimation of k_m from $i_m(x,y)(m = 1,2,3)$. In what follows, we propose two types of coded apertures and corresponding analysis algorithms.

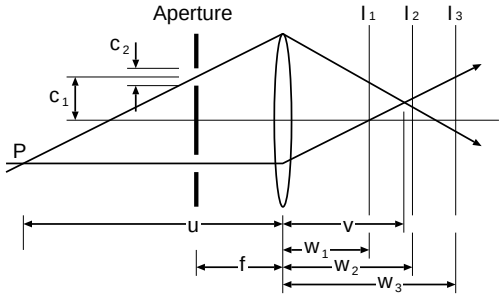

Figure 9: Multi-focus camera with a pair of pinholes aperture

5 Depth Measurement with a Pair of Pinholes Aperture

The analysis method of defocused images depends on the aperture design. Here we begin with a simple aperture pattern, a pair of pinholes. This aperture enables 1D Fourier analysis for the fast depth estimation and blur-free image reconstruction.

5.1 Blur Analysis by 1D FFT

Figure 9 illustrates the imaging process for the multi-focus camera with a pair of pinholes aperture.

In this case, the blur kernel $a(x, y)$ is described as follows:

$$a(x,y) = \frac{1}{2}(\delta(x - c_1, y) + \delta(x + c_1, y)), \qquad (4)$$

where $\delta(x, y)$ denotes Dirac's delta function and the vertical direction in Figure 9 represents the x axis. Applying 1D Fourier transform to equation (2), we have

$$\begin{aligned} I_m(s,y) &= \frac{1}{2}(e^{2\pi j c_1 k_m s} + e^{-2\pi j c_1 k_m s}) \cdot S(s,y) \\ &= A_m(s,v) \cdot S(s,y), \end{aligned} \qquad (5)$$

where s denotes the spatial frequency and $A_m(s, v)$ is defined by

$$A_m(s,v) = \cos(2\pi c_1 \frac{v - w_m}{f} s). \qquad (6)$$

We can see that the frequency response of the pair of pinholes aperture is represented by a cosine function in the real domain whose frequency varies with the focused position v.

Note that the aperture defined by equation(4) is the ideal one without any finite areas. In practical optical systems, however, the radius of each pinhole becomes finite to introduce enough light, which additionally causes a low-pass filter effect on observed images. To model this low-pass filter effect, a sinc function is incorporated into equation (6):

$$\tilde{A}_m(s,v) = A_m(s,v) \cdot \mathrm{sinc}(c_2 \frac{v - w_m}{f} s), \qquad (7)$$

where c_2 denotes the width of the aperture (Figure 9).

5.2 Depth Estimation by Frequency Response Matching

With the multi-focus camera, we have three ($N = 3$) defocused images:

$$I_1(s,y) = \tilde{A}_1(s,v) \cdot S(s,y), \qquad (8)$$

$$I_2(s,y) = \tilde{A}_2(s,v) \cdot S(s,y), \qquad (9)$$

$$I_3(s,y) = \tilde{A}_3(s,v) \cdot S(s,y). \qquad (10)$$

Dividing each side in a pair of above equations, we can eliminate the component of the original (blur-free) image $S(s, y)$. Consequently, we can estimate the focused position v by minimizing the following criterion function:

$$r(y,v) = \sum_{(m,n)\in P_3} \sum_s \left| \frac{I_m(s,y)}{I_n(s,y)} - \frac{\tilde{A}_m(s,v)}{\tilde{A}_n(s,v)} \right|, \qquad (11)$$

where P_3 denotes a set of the all combinations of neighboring image planes: $\{(1,2),(2,1),(2,3),(3,2)\}$. The first term in the right side of the above equation denotes observed data and the second the blurring model. In the optimization, a set of models with different v are prepared in advance. Then the one that minimizes equation (11) is selected, which determines the object depth v.

To examine the sensitivity of this model matching method, we compute the following similarity measure between the blurring models with different depth parameters v_1 and v_2:

$$sim(v_1,v_2) = \sum_{(m,n)\in P_3} \sum_s \left| \frac{\tilde{A}_m(s,v_1)}{\tilde{A}_n(s,v_1)} - \frac{\tilde{A}_m(s,v_2)}{\tilde{A}_n(s,v_2)} \right|, \qquad (12)$$

As is shown by the steep-walled valley in Figure 10, the accurate depth estimation can be attained by the proposed model matching.

5.3 Reconstruction of Blur-free Image

Now we have three defocused (i.e. degraded) images and their corresponding blurring (degradation) models have been determined. Using such redundant information, we can recover $S(s, y)$ stably by the following weighted inverse filter:

$$S(s,y) = \sum_{m=1}^{3} W_m \cdot \frac{I_m(s,y)}{\tilde{A}_m(s,v)}, \qquad (13)$$

where weight W_m is determined in proportion to $|\tilde{A}_m(s,v)|$ and inversely proportional to $|v - w_m|$. Note that this weighted inverse filter works well even if some $|\tilde{A}_m(s,v)|$ becomes zero at certain frequencies.

5.4 Experimental Results

Here we show experimental results of applying the above methods to a real world scene. Figure 11 shows

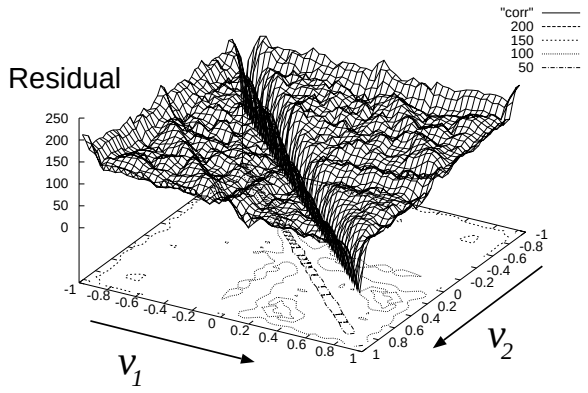

Figure 10: Residual of the frequency response with the different focus parameter v_1, v_2

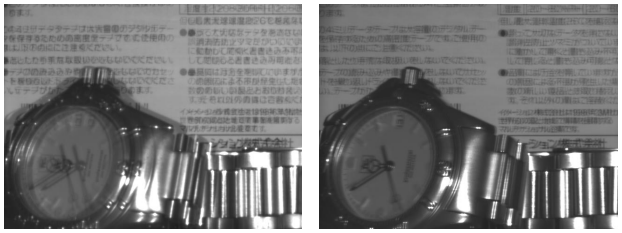

Figure 11: Input images

a pair of images (size: 640×480) taken by the multi-focus camera with an aperture using the following parameters: span between two pinholes: 20mm, radius of each pinhole: 2mm, focus value: the central CCD is focused at about 1.1m. The window size for 1D FFT is set to 128 pixels and neighboring 16 horizontal lines are averaged to eliminate noise. The window is slided vertically and horizontally every 16 pixels, and then the depth at the central position of each window is computed. In the optimization to measure the depth, the range of the focus position, i.e. that from w_1 to w_3 in Figure 9, is partitioned into 64 steps and the optimal focus position is selected as v.

The 3D profile in Figure 12 shows the computed depth map. The right pictures demonstrate the performance of the blur-free image restoration: Left: a part of an input image in Figure 11, Right: restored image.

6 Depth Measurement with a 2D Coded Aperture

Here we show another depth measurement algorithm which utilizes a convolution based model matching method. While the discussion here holds for any aperture pattern, the practical performance of the proposed method strongly depends on the anisotropy of the incorporated aperture.

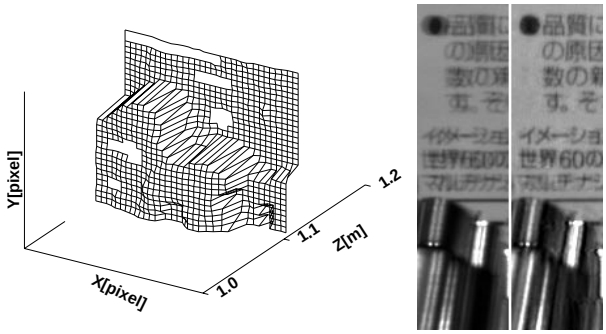

Figure 12: Depth map and restored blur-free image

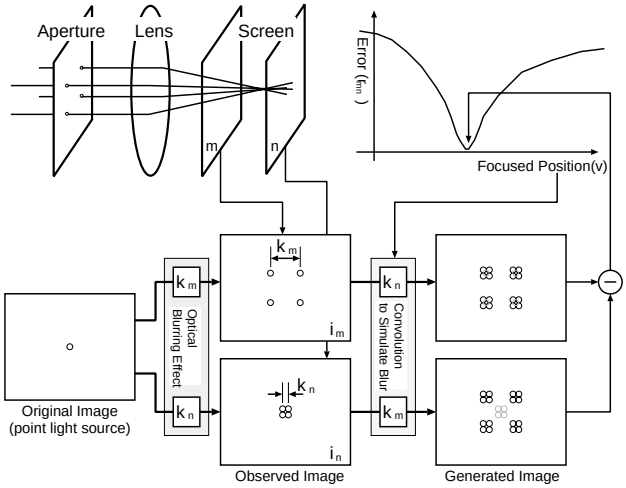

Figure 13: Depth analysis by convolution

6.1 Depth Analysis by Convolution

Using the commutative law of the 2D convolution, we can develop a much simpler depth measurement method without frequency domain analysis. Figure 13 illustrates the principle of the algorithm. Suppose we have a pair of defocused images $i_m(x, y)$ and $i_n(x, y)$ defined by equation (2). Let define the blur kernel with scale parameter k_m as follows:

$$a_{k_m}(x, y) = \frac{1}{k_m^2} a(\frac{x}{k_m}, \frac{y}{k_m}). \tag{14}$$

Then, we have the following equation based on the commutative law of the 2D convolution.

$$\begin{aligned} a_{k_m}(x, y) * i_n(x, y) &= a_{k_m}(x, y) * a_{k_n}(x, y) * s(x, y) \\ &= a_{k_n}(x, y) * i_m(x, y) \end{aligned} \tag{15}$$

Based on the above property, the focused position v can be calculated by minimizing the following criterion function:

$$\tilde{r}(v) = \sum_{(m,n)\in C_3} \sum_{(x,y)\in W} |a_{k_n} * i_m - a_{k_m} * i_n| , \quad (16)$$

where W denotes a 2D window for the depth estimation and C_3 a set of all possible pairs of input images. While in principle, this method can be used with any aperture, the sensitivity of the minimization varies with the aperture pattern. With the Gaussian blur kernel, for example, the stable depth measurement can hardly be realized, because images twice filtered by Gaussian contain almost no prominent features.

6.2 Experimental Result

In the experiment, we used an aperture with four pinholes, which are aligned at the each corner of a square 15mm each side. The radius of each pinhole is 1.5mm. The upper two pictures in Figure 14 show input images (size: 640×480) taken by the multi-focus camera and the lower-left the overview of the scene, where a dry cell is placed in front of a rolled paper with random texture pattern. The central CCD is focused at the point between the cell and the paper. We used 16×16 window for the convolution, which is shifted vertically and horizontally every 4 pixels to compute the depth at the central position of the window. In the optimization to measure the depth, the range of the focus position is partitioned into 40 steps and the optimal focus position is selected as v. The lower-right picture in Figure 14 shows the computed depth map. It takes about 3 seconds to calculate depth using DEC Alpha 400MHz.

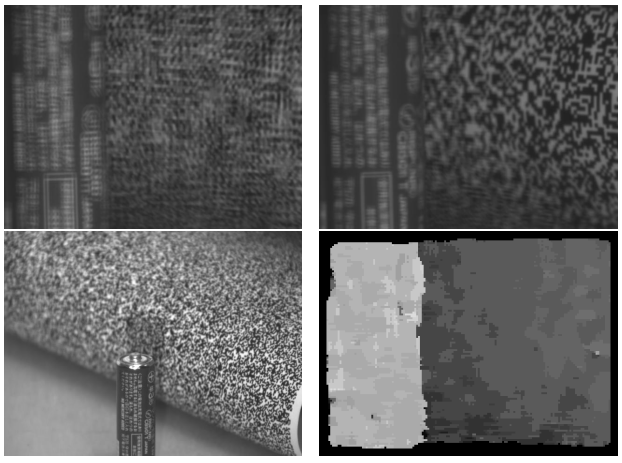

Figure 14: Input images (Upper), target scene (Lower-Left) and depth map (Lower-Right)

7 Conclusions

In this paper we proposed the multi-focus camera and two types of depth from defocus (DFD) algorithms using the camera. The multi-focus camera has many advantages as a practical range finder: compact, portable, passive, real time, wide measurable range, and single camera system without the correspondence problem. The augmented edge and depth from focus algorithm[3] worked very well to give about 0.7% RMS depth measurement accuracy.

In the latter part of the paper, we emphasized that the blurring pattern design is crucial to realize robust DFD range measurement and proposed as a novel range finder the multi-focus camera with a coded aperture. Two DFD methods have been proposed. They use the model-based pattern matching to estimate depth values of textured surfaces and restore a blur-free image from three defocused images taken by the multi-focus camera with a coded aperture. Experimental results demonstrated their effectiveness for real world scenes.

This work was supported by the Research for the Future Program of the Japan Society for the Promotion of Science (JSPS-RFTF96P00501) and Grant-in-Aid for Scientific Research ((A)(2)07558047).

References

[1] A.P.Pentland: A New Sense for Depth of Field, *IEEE Trans. PAMI*, Vol.9, No.4, pp.523-531 (1987)

[2] M.Subbarao and G.Surya: Depth from Defocus: A spatial domain approach, International Journal of Computer Vision, Vol.13, No.3, pp.271-294 (1994)

[3] N.Asada, H.Fujiwara and T.Matsuyama: Edge and Depth from Focus, *International Journal on Computer Vision*, Vol.26, No.2, pp.153-163, (1998)

[4] A.Pentland, T.Darrell, M.Turk and W.Huang: A Simple, Real-time Range Camera, Proc. of CVPR89, pp.256-261 (1989)

[5] S.K.Nayar, M.Watanabe, M.Noguchi: Real-time focus range finder, *Proc. of 5th Int. Conf. on Computer Vision*, pp.995-1001 (1995)

[6] M.Watanabe, S.K.Nayar: Minimal Operator Set for Passive Depth from Defocus, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp.431-438 (1996)

[7] G.Takemura: Real Time 3D Depth Measurement Using Multi-Focus Images, Master Thesis, Kyoto University, 1997 (in Japanese)

[8] E.Simoncelli and H.Farid: Direct Differential Range Estimation Using Optical Masks, Proc. of ECCV96, Vol.2, pp.82-93 (1996)

[9] M.Harwit, N.J.A.Sloane: Hadamard Transform Optics, Academic Press (1979)

[10] E.E.Fenimore, T.M.Cannon: Coded Aperture Imaging with Uniformly Redundant Arrays, Applied Optics, Vol.17, No.3, pp.337-347(1978)