

Fusing Depth from Defocus and Stereo with Coded Apertures

Yuichi Takeda
Osaka University

Shinsaku Hiura
Hiroshima City University

Kosuke Sato
Osaka University

Abstract

In this paper we propose a novel depth measurement method by fusing depth from defocus (DFD) and stereo. One of the problems of passive stereo method is the difficulty of finding correct correspondence between images when an object has a repetitive pattern or edges parallel to the epipolar line. On the other hand, the accuracy of DFD method is inherently limited by the effective diameter of the lens. Therefore, we propose the fusion of stereo method and DFD by giving different focus distances for left and right cameras of a stereo camera with coded apertures. Two types of depth cues, defocus and disparity, are naturally integrated by the magnification and phase shift of a single point spread function (PSF) per camera. In this paper we give the proof of the proportional relationship between the diameter of defocus and disparity which makes the calibration easy. We also show the outstanding performance of our method which has both advantages of two depth cues through simulation and actual experiments.

1. Introduction

Stereo method is a most popular technique to acquire a depth map of a scene passively, i.e. without projecting light on that scene. This method only requires just two ordinary cameras, and recent progresses in computing performance have reduced the time for finding point correspondences. Use of light projection for the distant outdoor scene is not effective because of the daylight, so this technique will also be indispensable in the future.

Finding correct correspondence is the most important step of the stereo method. An object that has a texture with repetitive pattern or edges parallel to an epipolar line tends to cause incorrect correspondences. There are, however, passive techniques to measure depth besides stereo imaging, such as depth from defocus (DFD), that utilize changes in the imaging characteristics of lenses depending on the distance to an object. Unlike stereo imaging, the performance of these techniques is not influenced by the direction of edges. A problem with these techniques, however, is the less accuracy with a distant object because the size of de-

focus (the diameter of the circle of confusion) is limited by the effective diameter of the lens. Therefore, in this paper, we propose a novel depth measurement technique that combines stereo imaging and DFD. Stereo imaging allows customization of the baseline length, which ensures the accuracy of depth. This approach is combined with the effect of DFD to improve the robustness of measurement. In addition, coded apertures are incorporated to optimize the blurring phenomenon of lenses. Improvement in the accuracy and robustness of measurement is confirmed using simulation. Experiments using an actual equipment shows the result of depth measurement and deblurring for a real scene.

2. Related work

Stereo imaging is a typical passive depth measurement technique that has often been studied and used. However, using 2 cameras causes an ambiguous correspondence between images when an object has edges parallel to an epipolar line or a repetitive texture. Thus, a technique has been proposed[8] to facilitate finding correct correspondence by using 3 or more cameras with baselines of multiple lengths and directions. However, additional cameras mean increased cost. In addition, having a short distance from the camera to the object causes a problem of shallow depth of field. In other words, defocus is regarded as an undesired phenomenon and the aperture of the lens should be stopped down for sufficiently deep depth of field.

On the other hand, techniques using the blurring caused by lenses have been proposed as a passive depth measurement method. While depth from focusing physically performs focusing, DFD analyzes the amount of blurring in images taken using a fixed lens and has been actively studied[3, 5, 7]. DFD uses an two-dimensional aperture, so there is no performance dependence on the direction of edges as stereo imaging has. However, DFD uses blurring from a single lens, so accuracy cannot be ensured when using a wide-angle lens or imaging a distant object. This is because the effective length of the baseline is constrained by the effective size of the aperture[11].

A dynamic scene is not as easy to measure with DFD as it is with stereo imaging. To determine the depth of a scene without a priori knowledge of that scene, more than 2

images with different optical parameters, such as focus distance, must be obtained. This is because a single image does not resolve ambiguity of whether an object with a detailed texture appears blurred or whether an object with a blurred texture is positioned in focus. Thus, several techniques that features prisms or mirrors that split light into beams striking multiple image sensors has been proposed[2, 3, 9]. These techniques allow imaging in real-time, but its setup is more complicated than that for stereo imaging, which simply requires multiple cameras. The technique also requires special optics, making it costly. A technique that changes aperture geometry[4, 6, 12, 15] cannot image a dynamic scene.

Techniques to change the geometry of the lens aperture are closely related to a technique involving what are known as coded apertures[3, 5, 14, 16]. Takeda *et al.* proposed a technique to improve the accuracy of depth estimation in shallow depth of field by combining coded apertures with stereo imaging[13]. However, the technique clearly did not use the blurring phenomenon of a lens as a cue to estimate distance. In contrast, we propose a technique combining stereo imaging and cues of depth estimation based on the same principle as DFD. This is accomplished by giving different focus distances for 2 cameras of a stereo camera.

There are some previous works that combine DFD and stereo. For example, Rajagopalan *et al.* used 4 images based on the combinations of 2 viewpoints and 2 focus distances, and applied conventional stereo and DFD algorithms to the pairs of same focus distances and viewpoints, respectively[10]. Therefore, by their method, it is impossible to handle 2 images with both disparity and different focus distance. The other work by Gheta *et al.* uses simple focus measure for each image[1]. Contrary to their works, our method only use 2 input images, and make it possible to avoid some failure cases of simple stereo without any drawbacks in capturing time and device cost. Our technique also eliminated difficulties on obtaining multiple images with different optical parameters from the same point of view, which was a flaw of conventional DFD.

3. Fusing DFD and stereo

This section describes a proposed technique to create a depth map of a scene and blur-free (extended depth-of-field) images. This technique combines the 3 elements of stereo imaging, DFD, and coded apertures via a single point spread function (PSF).

3.1. Summary of the technique

The system proposed in this paper is shown in Figure 1. Like normal stereo imaging, 2 cameras are positioned in parallel, but the distance between the lens and image sensor is changed so that they differ for the 2 cameras. In addition, each lens is equipped with a coded aperture mask that optimizes the geometry of blurred images of a point light

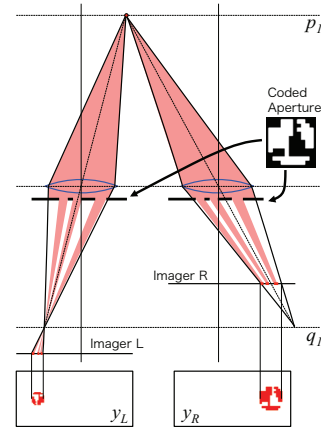


Figure 1. Optical diagram of proposed system.

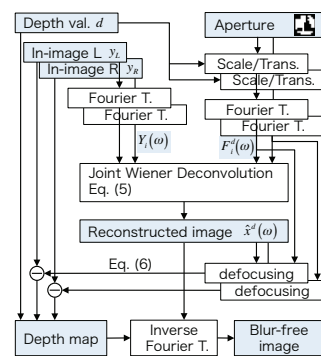


Figure 2. Flowchart of depth estimation and deblurring.

source, i.e. a PSF is optimized. The geometry of the coded aperture mask used is described in Section 3.5.

The steps in estimation of a scene’s depth and reconstruction of blur-free images are shown in Figure 2. Normal stereo imaging calculates a binocular disparity by comparing templates sliding on the image. In contrast, our technique prepares a PSF, which includes both blurring and binocular disparity as values corresponding to the depth. Then blur-free images are created assuming an object is at that distance. Using this PSF and blur-free images, images that are blurred again are compared to input images. If the assumed distance is correct, re-blurred images and input images are sure to correspond in terms of both the amount of defocus and binocular disparity. An incorrect distance will lead to error in both the reconstruction of blur-free images and reproduction of blurring and disparity, so image correspondence will decrease. Consequently, the depth of a scene can be determined using information on both binocular disparity and defocus.

As Schechner *et al.* noted[11], larger diameter of the lens aperture will result in accurate estimate of depth as a stereo camera with a longer baseline. However, increases in the numerical aperture are limited by lens design. Focal length

is determined by the size of the image sensor and angle of view of the scene being imaged, so the diameter of the lens aperture itself is limited by the size of the image sensor. The proposed technique allows customization of the length of the baseline between 2 cameras, so even the depth of a distant object can be estimated with a high resolution. However, depth cannot be estimated with stereo imaging when there are no changes in image information in the direction of an epipolar line. Contrary, Like DFD[11], our technique uses changes in images as a result of blurring, which allows depth to be estimated. However, in this case, the accuracy of depth estimation will be equivalent to that of DFD.

Having left and right cameras with different focus distances is an another advantage for the creation of blur-free images. Typically, the accuracy of reconstruction of the original image increases with less blurring, regardless of whether or not a coded aperture is used. That accuracy decreases with deviation of object distance from the focus distance. With our technique, the left and right cameras have different focus distances, and images from those cameras compensate for one another. This should increase the accuracy of reconstruction of the original image in comparison to stereo imaging with coded apertures[13].

3.2. Expression of binocular disparity with a PSF

Blurred images can be represented by a convolution of blur-free original image and PSF. Images will also have the binocular disparity corresponding to the distance to an object. The images input by the left and right cameras of a system as shown in Figure 1 can be derived as shown below based on a common original image $x(u, v)$ where $*$ represents convolution.

$$y_i(u, v) = p_i^d(u, v) * x(u - d, v) \quad (1)$$

Here, $i = \{L, R\}$ represents the left and right cameras, and an epipolar line is assumed parallel to the horizontal axis u of the image. $p_i^d(u, v)$ and d represent the PSF and disparity of each camera with respect to depth. When we assume the origin of the disparity set at left image, the disparity is expressed by only $d = d_R$ and $d_L = 0$. PSF $p_i^d(u, v)$ is a homothetic shape of the aperture, though its ratio (scaling factor) is described later in Section 3.4. When this equation is Fourier-transformed, convolution is converted into a product, and translation of the function is expressed as the product of exponential functions, yielding:

$$Y_i(\omega) = F_i^d(\omega) \cdot X(\omega) \quad (2)$$

$$F_i^d(\omega) = P_i^d(\omega) e^{-j\omega_u d} \quad (3)$$

where $\omega = (\omega_u, \omega_v)$ is a spatial frequency vector corresponding to the horizontal and vertical axes u, v of the original image. $Y_i(\omega)$, $P_i^d(\omega)$ and $X(\omega)$ are the Fourier transforms of $y_i(u, v)$, $p_i^d(u, v)$ and $x(u, v)$, respectively. As

seen in Eq. (3), translation of the image is combined with the Fourier transform of the PSF. It allows calculation in the spatial frequency domain akin to analysis of blurring (used in techniques like DFD) when there is no binocular disparity.

3.3. Depth estimation and deblurring

Using the optical transfer function (OTF, i.e. Fourier transformation of the PSF) as was determined in the previous section results in the expression for the spatial frequency domain in Eq. (2). When the distance to an object is given, estimated values $\hat{X}_i^d(\omega)$ for the Fourier transform of the original image can be determined by the following equation of Wiener Deconvolution:

$$\hat{X}_i^d(\omega) = \frac{\bar{F}_i^d(\omega) \cdot Y_i(\omega)}{|F_i^d(\omega)|^2 + |C|^2} \quad (4)$$

where $\bar{F}_i^d(\omega)$ represents the complex conjugate of $F_i^d(\omega)$. C is a constant to prevent divergence in the solution when $|F_i^d(\omega)|$ is small. Typically, the inverse of the SN ratio of the input image is used for C .

The system proposed in this paper simultaneously yields 2 images, $Y_L(\omega)$ and $Y_R(\omega)$. Either can be used individually to reconstruct the original image, but information lost from one is likely to remain in the other since the zeros of the OTF are in different locations because the size of the PSF differs in the two images. Reconstruction of an original image from 2 images as proposed by Zhou *et al.* [15] is applicable to our technique. In other words, a reconstructed image when d is assumed to be the disparity will be estimated as

$$\hat{X}^d(\omega) = \frac{\sum_{i=L,R} \bar{F}_i^d(\omega) \cdot Y_i(\omega)}{\sum_{i=L,R} |F_i^d(\omega)|^2 + |C|^2} \quad (5)$$

then a blurred image is recreated based on this reconstructed image, and the disparity is determined in comparison to the input image,

$$\hat{d} = \arg \min_d \sum_{i=L,R} \left| y_i - \mathcal{F}^{-1}(F_i^d(\omega) \cdot \hat{X}^d(\omega)) \right|^2 \quad (6)$$

where \mathcal{F}^{-1} represents an inverse Fourier transform. If the assumed disparity d differs from the actual disparity, reconstruction of the original image in Eq. (5) will fail. Moreover, error will appear in both the size of the PSF and disparity of the blurred image, so the residual in Eq. (6) will increase. As a result, the extent of disparity and blurring is handled in an integrated manner and the depth of the scene is estimated.

In actuality, the depth of a scene is not a single value but instead differs at points, so Eq. (6) is solved for each small

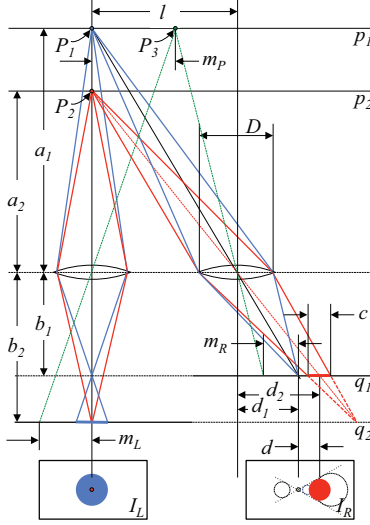


Figure 3. Relationships between circle of confusion and disparity.

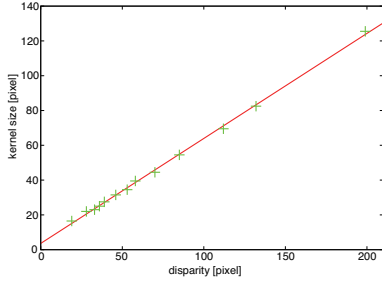


Figure 4. An example of the relationship between disparity and circle of confusion.

region in the proximity of points in the image, creating a depth map. Finally, the reconstructed image $\mathcal{F}^{-1}(\hat{X}^d(\omega))$ based on distances to individually determined points is pieced together to yield a blur-free image.

3.4. Relationship between binocular disparity and the diameter of the circle of confusion

As described in Section 3.2, PSF $p_i^d(u, v)$ is a scaled shape of the aperture, and its size (the diameter of the circle of confusion) changes in accordance with the distance to the object. Past studies on extended depth-of-field (EDOF) and DFD used only the extent of blurring as a depth cue[3, 4, 5, 14, 15]. In contrast, our study determined depth using both the extent of blurring and disparity in stereo vision, so the relationship between the two must be identified by calibration. Here, the relationship between the diameter of the circle of confusion and binocular disparity was derived and it was found to be linear even if the focused distances of 2 cameras are different. This makes calibration easy.

Figure 3 depicts lenses with the same aperture diameter D and parallel optical axes in a stereo camera consisting of

2 cameras with differing focus distances. The camera on the right I_R with image plane q_1 focuses on plane p_1 while the camera on the left I_L with image plane q_2 focuses on plane p_2 . The positions of the 4 planes, p_1 , p_2 , q_1 and q_2 , are expressed by an imaging formula as follows:

$$\frac{1}{a_1} + \frac{1}{b_1} = \frac{1}{a_2} + \frac{1}{b_2} = \frac{1}{f} \quad (7)$$

where f is the focal length of the lens. For simplicity's sake, the following discussion considers an instance where point light sources P_1 and P_2 are both on the optical axis of the left camera I_L .

Disparity will now be considered. Here, disparity is assumed to be 0 when the distance to the object is infinite. The difference of disparity d in disparity d_1 with regard to the in-focus point P_1 and disparity d_2 with regard to out-of-focus point P_2 can be determined by:

$$d = d_2 - d_1 = \left(\frac{b_1}{a_2} - \frac{b_1}{a_1} \right) l \quad (8)$$

where l is the baseline length. When point P_2 on object p_2 is imaged by the image sensor on plane q_1 , the diameter of the circle of confusion c is determined by

$$c = \frac{b_2 - b_1}{b_2} D \quad (9)$$

using lens aperture diameter D . Based on this equation and imaging formula in Eq. (7), the ratio of disparity d and the diameter of the circle of confusion c when blurring occurs during focusing is:

$$\frac{d}{c} = \frac{b_1 b_2}{b_2 - b_1} \left(\frac{1}{a_2} - \frac{1}{a_1} \right) \frac{l}{D} = \frac{l}{D} \quad (10)$$

so variables related to the distance to the object can be completely eliminated. Accordingly, the diameter of the circle of confusion c and disparity d are proportional. The ratio of the two is the ratio of the lens aperture diameter D to the baseline length l . As shown in the right image I_R in Figure 3, the circle of confusion shifts and changes size in response to changes in distance, and the envelope of the circles of confusion forms 2 straight lines that intersect at the focusing point.

Considered next is an instance where the object is not on the optical axis of the left camera I_L . Typically, in a normal stereo camera with parallel optical axes, the disparity on a plane perpendicular to the optical axis is constant. In contrast, cameras with different focus distances as were proposed in this paper have a slightly different angle of view. As a result, the disparity is no longer constant. An instance where point light source P_1 is shifted on plane p_1 with displacement m_P is shown in Figure 3 in green. Displacement

of the image of point P_3 with respect to the image of point P_1 is

$$m_L = \frac{b_2}{a_1} m_P, m_R = \frac{b_1}{a_1} m_P$$

for the left camera and right camera, and the ratio $\frac{m_L}{m_R}$ of the shifting of images is always $\frac{b_2}{b_1}$ and thus constant. Therefore, image I_L provides an image $\frac{b_2}{b_1}$ -fold larger than image I_R when blurring due to the finite lens aperture diameter D is ignored. This problem is remedied through image scaling. In other words, the image from the left camera I_L can be scaled using the ratio $\frac{b_2}{b_1}$ of the effective focal length. This step results in disparity with respect to a plane perpendicular to the optical axis that is constant, regardless of the distance to the object. Even if the object is not on the optical axis of the left camera I_L , the aforementioned linear relationship between disparity and the diameter of the circle of confusion is kept. Also apparent is the fact that the linear relationship between disparity and the diameter of the circle of confusion is retained for the scaled image from the left camera I_L as well. Another way to eliminate changes in image magnification due to changes in the focus distance is the use of rear-telecentric optics[7].

Conversion from the binocular disparity to the diameter of the circle of confusion is rather easy through use of the relationship as described thus far. Specifically, images of 2 bright spots placed in a scene at different distances can be used. The binocular disparity and size of the image of the spot are both determined, and then the relationship between binocular disparity and the diameter of the circle of confusion can be determined at other distances using linear interpolation. It can also be determined at a large number of distances to increase the accuracy of calibration and they should describe a straight line according to the least-squares method.

Measurements of the binocular disparity and diameter of the circle of confusion when the bright spot was actually placed at various distances are shown in Figure 4. The point light source in this experiment was a fiber optic light source, and the extent of blurring was determined manually. The green crosses in Figure 4 are measured points, and the straight red line is the line they describe according to the least-squares method. The lens is almost focused to infinity, so when binocular disparity is 0 the diameter of the circle of confusion will also be 0. The graph shows that the binocular disparity and diameter of the circle of confusion can easily be converted back and forth.

3.5. Shape of the aperture

The PSF for an image taken by a camera with coded aperture represents scaling of the aperture geometry in accordance with the depth. As indicated in a previous section, the scale of the PSF changes with binocular disparity in a linear fashion. Our method reconstructs a blur-free image,

as in Eq. (5), while estimating depth. More accurate reconstruction will result in more accurate estimation of depth. The Wiener Deconvolution has absolute values for the OTF in its denominator, so spatial frequency response of the PSF should be broad-banded. Thus, we utilize the broadband code proposed by Zhou *et al.* [16] (denoted here as Zhou's code) since it is suited to Wiener Deconvolution. This code assumes pink noise in the prior distribution of natural images. Patterns differ depending on the noise level σ in an input image. In this paper, we performed experiments using a code of $\sigma = 0.005$ by assuming quantization noise. The shape of used code is shown in Figure 1.

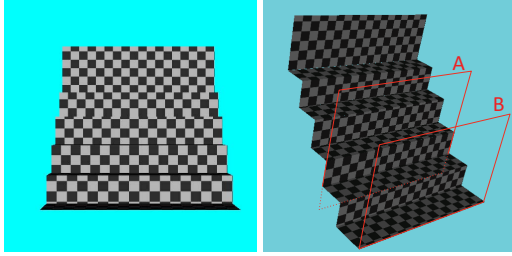
4. Simulation experiment

The depth estimation technique proposed in this paper combines coded stereo imaging and DFD. This section describes the results of a simulation that compared this technique to stereo method with coded apertures[13], DFD with different focus distance and a single coded aperture[3], and DFD with coded aperture pairs[15].

4.1. Generating images

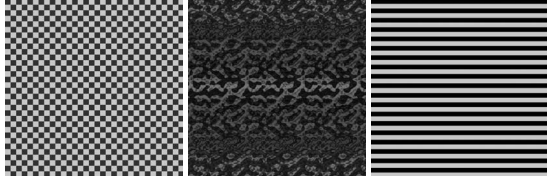
A scene with a stair-like shape was used in this experiment. The scene is shown in Figure 5. Images featuring the 3 textures of a checkerboard pattern, dappled pattern and horizontal stripes were created. The textures used are shown in Figure 6. Experiments were all done with grayscale images. The checkerboard pattern featured sharp, distinct edges in both the vertical and horizontal directions. In contrast, the dappled pattern featured changes in luminance in the image overall but weaker contrast than the checkerboard pattern. The horizontal striped pattern had only edges parallel to the epipolar line of stereo rig, but it was deliberately used to assess the performance of defocus depth cue. OpenGL was used for scene definition and rendering. Depth was retrieved from the OpenGL depth buffer. The corresponding disparity map for Figure 5(a) is shown in Figure 7(a).

Creation of blurred images in the simulation will now be described. The blurred image formed by a lens is equivalent to the average of translated images observed from every viewpoint in the lens aperture. Thus, the shape of the lens aperture was first specified and then images seen from various viewpoints in the aperture were created. Next, images multiplied by the transmittance $\{0, 1\}$ of the mask at individual points on the aperture were averaged to provide a blurred image. Images were translated so that the disparity at the focus distance would be 0, allowing adjustment of the focus distance. An example of the blurred images created is shown in Figure 7(b). For the precise simulation of blurring effect, we set 3600 viewpoints in an aperture, and final images were quantized to 8 bits and saved. image was 512×512 pixels in size.



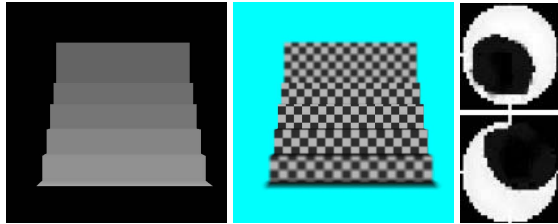
(a) From the front (b) From another viewpoint

Figure 5. Shape of the scene for simulation.



(a) Checkerboard (b) Dappled (c) Horizontal stripe

Figure 6. Textures for simulation.



(a) Ground truth disparity image (b) An example of generated images (c) Zhou's aperture pair

Figure 7. Images and apertures used for evaluation. (a) Ground truth of disparity for evaluation. (b) An example of generated image focused on the plane A. (c) Two aperture shapes used in Technique 4 : Coded Aperture Pair.

4.2. Compared methods

We compared 4 measurement methods for the evaluation experiment presented here. All methods use 2 input images each. The focus distance in each image was a combination of two planes, A and B, as shown in Figure 5(b). Plane A aims to minimize the amount of defocus in both of front and back of the plane because the depth of field in the back of a focused plane is deeper than front. Since the accuracy of depth estimation should be evaluated at not only between 2 focal planes but also outside, we set Plane B at the front of the stair-like scene. Detailed settings for each measurement methods are in following:

Technique 1: Proposed method The same coded aperture was used for the left and right cameras, but the focus distance for each lens differed. The left image was focused to the Plane B, and the right was to the Plane A. To incorporate disparity depth cue, the viewpoint

was shifted horizontally with just the baseline length. The algorithm was as described in the previous section.

Technique 2: Stereo imaging with coded apertures[13]

The same coded aperture was used for the left and right cameras, and the focus distance was also the same; both were focused on plane A. A block-matching algorithm with normalized correlation was used to find corresponding points. The length of baseline is same to the Technique 1.

Technique 3: DFD with a coded aperture[3]

Only the focus distance differed in the 2 images. The shape of coded aperture and viewpoint for the 2 images was the same, and the focus distance is set to plane B for input image 1 and to plane A for input image 2, just like in the Technique 1.

Technique 4: Coded aperture pair[15]

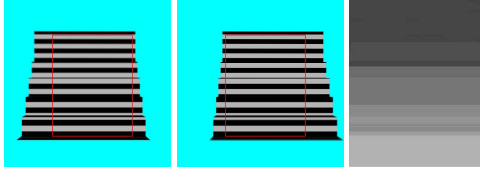
Distance measurement according to Zhou *et al.* using 2 images with different aperture shapes. The apertures used were shown in Figure 7(c). The focus distance for the lenses was the same, and both were focused on plane A.

The shape of the coded apertures for techniques 1 - 3 are Zhou's code shown in Figure 1 as described in Section 3.5. Disparity ranged upto 60 pixels because of the baseline length specified, and this coincides with the disparity search range. The aperture diameter was 1/3 of the baseline length.

4.3. Results

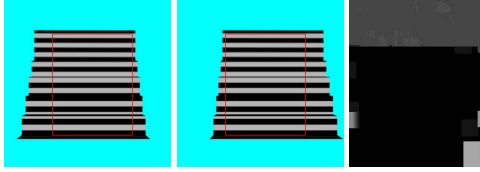
Two input images and an output disparity map (cropped area within the red box) for each technique are shown in 8-11 when the scene texture was a horizontal stripe pattern. Although DFD and aperture pair do not have disparity, we converted the estimated depth to a corresponding disparity value for comparison. In disparity maps, brighter areas represent a larger disparity. The averages of absolute error of disparity using 4 methods and 3 textures are summarized in Figure 12. Distance was not defined in the area outside the steps, so error was calculated for the area within the red boxes in Figures 8-11.

With the checkerboard pattern and dappled pattern, depth was estimated with higher accuracy using the proposed technique and coded aperture stereo than when using DFD and aperture pair. This is because a long baseline in stereo imaging helps to increase the resolution of depth estimation. With the horizontal striped pattern, however, stereo imaging resulted in an extremely large error as shown in Figure 9(c). This is, as was mentioned earlier, because no depth cues were provided since the texture of the scene was parallel to an epipolar line. In contrast, the proposed technique yielded a depth with the same principle as DFD. With a horizontal striped pattern, the aperture pair technique of



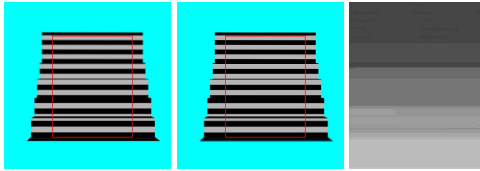
(a) Left input (b) Right input (c) Disparity

Figure 8. Proposed method.



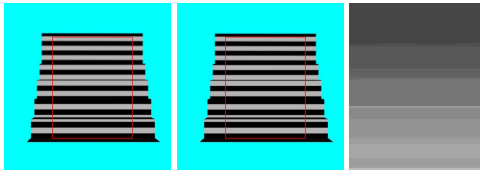
(a) Left input (b) Right input (c) Disparity

Figure 9. Coded Aperture Stereo.



(a) Input 1 (b) Input 2 (c) Disparity

Figure 10. Depth from Defocus.



(a) Input 1 (b) Input 2 (c) Disparity

Figure 11. Zhou's aperture pair method[15].

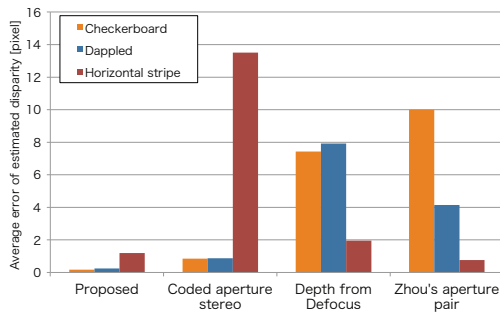


Figure 12. Average error of estimated disparity[μm].

Zhou *et al.* provides the best performance since their technique has similar effect of stereo imaging with a baseline perpendicular to the pattern. This is because the aperture shapes (Figure 7(c)) are decentered to the top or bottom of the aperture, so the images obtained are similar to images captured at the center of gravity of the aperture opening.



Figure 13. Lens with coded aperture and camera on a slide stage.



(a) Left input image (b) Right input image

Figure 14. Inputs image for the real experiment.



(a) Depth map (b) Deblurred image

Figure 15. Estimated disparity map and deblurred image.



(a) Close-up of red box in Figure 14(a) (b) Close-up of red box in Figure 14(b) (c) Close-up of red box in Figure 15(b)



(d) Close-up of blue box in Figure 14(a) (e) Close-up of blue box in Figure 14(b) (f) Close-up of blue box in Figure 15(b)

Figure 16. Close-ups from input and deblurred images.

5. Experiment with an actual equipment

5.1. Experimental setup

The camera used was a Nikon D200, and 2 lenses (Cosina Carl Zeiss Makro-Planar T* 2/50mm ZF.2) were used. Focus distances of two lenses are fixed to 2m and 5m. These distances correspond to the frontmost object and background of the scene. Input images are downsampled to 968×648 . In this experiment, one camera was mounted on a sliding stage instead of using 2 cameras as shown in Figure 13. Two images were taken by shifting the camera to the left and right with interchanging lenses. The baseline length (the displacement of the camera) was 14mm to

clearly show the effect of DFD and deblurring instead of stereo. The shape of the coded aperture was Zhou’s code of $\sigma = 0.005$ like that in the experiment in the previous section, so the code was printed on opaque paper, cut out, and attached to the rear of the lens. The setting of aperture blade of the lens was $F=2.0$ (fully open) to prevent vignetting, and we conformed no vignetting of the coded aperture occurs at the edges of the image. The relationship between the extent of disparity and PSF was calibrated using the method in Section 3.4. With the lenses used in this experiment, the effective focal length and the angle of view changes as a result of changes in the focus position. Thus, correction of the input images was done by image magnification as described in Section 3.4. The coded aperture in place on the lens and the system that was actually used are both shown in Figure 13.

5.2. Results

The input images obtained by the system are shown in Figure 14. The disparity map and an image that has been deblurred (blur-free) are shown in Figure 15. Figure 16 shows magnified images of the red boxes and blue boxes in Figure 14 and Figure 15(b). Based on the disparity map, depth has been estimated correctly for the most part with evident texture. Compared to Figure 14, the blur-free image features clarity at most distances. The area inside the red box is blurred in the left input image (Figure 14(a)) but clear in the right (Figure 14(b)). Thus, the output result is an image with the same level of clarity. Letters within the blue box are blurry in both input images, but a blur-free image has been reconstructed to the extent that letters are readily legible.

6. Conclusion

This paper has proposed a technique combining stereo imaging and DFD with coded apertures. This is achieved by utilizing 2 cameras with different focus distances like a stereo camera and by expressing binocular disparity and defocus as a single PSF. Experimental results indicated that the technique was able to determine the distance to object with a texture parallel to an epipolar line, which is something stereo imaging could not accomplish. The technique also provided higher depth resolution because it utilized a longer effective baseline length than DFD. In addition, experiments using the proposed setup facilitated estimation of the depth of an actual scene and creation of blur-free images.

One topic for the future is improved reconstruction of the original image. This study used Wiener Deconvolution with no prior of the original image. A number of effective techniques regarding the removal of blurring have been proposed with priors of the edge intensity distribution[5] or sparseness regularization. Other topics for the future are uti-

lizing different aperture shapes for both lenses in the device proposed in this manuscript and performing experiments using 3 or more cameras.

Acknowledgements This work was supported by Grant-in-Aid for Scientific Research (No. B:21300067) and Grant-in-Aid for Scientific Research on Innovative Areas ”Shitsukan” (No. 22135003) from MEXT, Japan.

References

- [1] I. Gheta, C. Frese, M. Heizmann, and J. Beyerer. A new approach for estimating depth by fusing stereo and defocus information. *Gesellschaft fur Informatik e.V Notes in Informatics*, P-109:26–31, 2007.
- [2] P. Green, W. Sun, W. Matusik, and F. Durand. Multi-aperture photography. *ACM Trans. Graph.*, 26(3), 2007.
- [3] S. Hiura and T. Matsuyama. Depth measurement by the multi-focus camera. In *CVPR’98*, pages 953–959, 1998.
- [4] A. Levin. Analyzing depth from coded aperture sets. In *ECCV2010*, pages 214–227, 2010.
- [5] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.*, 26(3), 2007.
- [6] C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen. Programmable aperture photography: multiplexed light field acquisition. *ACM Trans. Graph.*, 27(3):55:1–55:10, 2008.
- [7] S. Nayar, M. Watanabe, and M. Noguchi. Real-time focus range sensor. In *ICCV95*, pages 995–1001, 1995.
- [8] M. Okutomi and T. Kanade. A multiple-baseline stereo. In *CVPR’91*, pages 63–69, 1991.
- [9] A. P. Pentland. A new sense for depth of field. *IEEE Trans. PAMI*, PAMI-9(4):523–531, July 1987.
- [10] A. N. Rajagopalan, S. Chaudhuri, and U. Mudenagudi. Depth estimation and image restoration using defocused stereo pairs. *IEEE Trans. PAMI*, 26(11):1521–1525, 2004.
- [11] Y. Y. Schechner and N. Kiryati. Depth from defocus vs. stereo: How different really are they? *Int. J. Comput. Vision*, 39:141–162, 2000.
- [12] G. Surya and M. Subbarao. Depth from defocus by changing camera aperture: a spatial domain approach. In *CVPR’93*, pages 61–67, 1993.
- [13] Y. Takeda, S. Hiura, and K. Sato. Coded aperture stereo - for extension of depth of field and refocusing. In *Int. Conf. on Computer Vision Theory and Applications*, pages 103–111, 2012.
- [14] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph.*, 26, 2007.
- [15] C. Zhou, S. Lin, and S. Nayar. Coded aperture pairs for depth from defocus. In *ICCV2009*, pages 325–332, 2009.
- [16] C. Zhou and S. Nayar. What are good apertures for defocus deblurring? In *Int. Conf. on Computational Photography*, pages 1–8, 2009.