A Thin Lens Based Camera Model for Depth Estimation from Defocus and Translation by Zooming

Masashi Baba, Naoki Asada, Ai Oda, Tsuyoshi Migita
Department of Intelligent Systems
Hiroshima City University
{baba,asada,ai,migita}@cv.its.hiroshima-cu.ac.jp

Abstract

Depth recovery is a central concern in computer vision, and many methods were proposed for the monocular depth estimation by zooming as well as focusing and irising. In the past, there are two distinct approaches in depth by zooming; one is from motion parallax along the optical axis using a pinhole camera model, and the other from defocus using a thin lens camera model. This paper presents a new camera model that accounts for both effects of defocus and lens center translation by zooming. We first discuss the optical properties of zoom lenses, then present a thin lens based camera model that describes the mutual relationship between zoom, focus and iris parameters. Using this model with calibration results, we have performed some experiments with real images and evaluated the accuracy of the depth information recovered from defocus and lens center translation. Experimental results have demonstrated the validity of our camera model and also shown its applicability to the depth estimation from defocus and translation by zooming.

1 Introduction

Zoom lens systems are widely used in various types of cameras because of the imaging flexibility and versatility; that is, we can take images with arbitrary magnification, focus and brightness conditions by controlling zoom, focus and iris settings. In computer vision, however, since such a zoom lens system has complex optical properties and thus it is difficult to model it, there are few works to deal with the zoom lens effects totally. The purpose of this study is to develop a new camera model that describes the mutual relationship between zoom, focus and iris parameters and to show its applicability to the depth estimation by zooming.

Depth recovery is a central concern in computer vision, and many methods were proposed for the depth estimation by focusing[1, 2, 3, 4] and irising[5, 6, 7]. In these studies, while the camera models and lens parameters are different, all of the methods are established on a common principle; the image blur is observed when the imaging system has a finite lens aperture and the degree of defocusing depends upon the distance between the lens and the object, therefore the depth information is derived from the image blur.

On the other hand, some researchers are interested in a zooming effect that is the lens center translation along the optical axis of the lens, and they have developed a method called depth by zooming[8, 9, 10]. This approach is based on the motion parallax using a pinhole camera model, which means that the image blur caused by zooming is ignored.

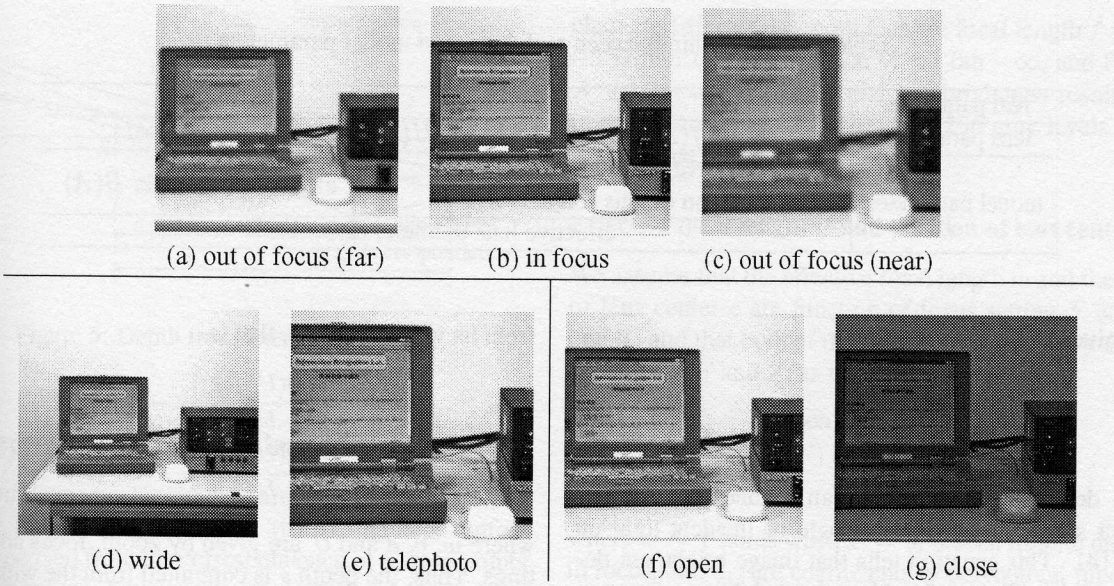
Recently, we have reported another approach to the depth recovery by zooming [11] in which the depth information is estimated from defocus by zooming. The study developed a thin lens based camera model that describes the defocusing by zooming, focusing and irising, yet the lens center translation is not considered.

In this paper, we propose a new camera model of zoom, focus and iris parameters that accounts for the image blur as well as the lens center translation. In the following section, we first discuss the optical properties of zoom lens parameters, then present a thin lens based camera model. Using this model with calibration parameters, we will show some experimental results with real images and accuracy evaluation of the depth from defocus and translation by zooming.

2 Camera Model

Zoom lens systems have three parameters, zoom, focus and iris, and those imaging effects are observed as follows. **Zoom** allows to take images with arbitrary magnification of a scene without changing the focused distance and image brightness, see Fig. 1(d)(e). **Focus** enables to take focused images of an object at any distance with varying the image magnification slightly without changing the image brightness, see Fig. 1(a)(b)(c). **Iris** is capable of controlling the image brightness with changing the depth of field, see Fig. 1(f)(g).

Table 1 summarizes these imaging effects in terms of three parameters. The image magnification is main and side effects of zoom and focus, respectively. This suggests that



(a) out of focus (far) (b) in focus (c) out of focus (near)

(d) wide (e) telephoto (f) open (g) close

Figure 1: Imaging effects by focusing in (a)(b)(c), by zooming in (d)(e), and by irising in (f)(g).

Table 1: Imaging effects of zoom, focus and iris parameters.

	magnification	focusing	brightness
zoom	○	△	△
focus	△	○	△
iris	×	△	○

○:main effect, △: side effect, ×:no effect

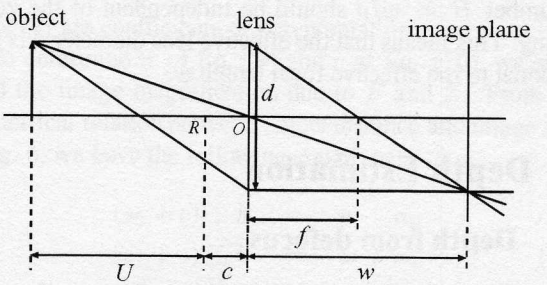


Figure 2: Camera model.

the magnification factor should be a function of zoom and focus parameters. In addition, zoom and focus modify the magnification without changing the image brightness. This implies that the aperture opening should be a function not only of iris but of zoom and focus parameters.

Based on the consideration mentioned above, We have built a new camera model of zoom, focus and iris parameters by using a thin lens imaging system, as shown in Fig. 2. In literature, the thick lens model is often used to model the zoom lens systems[12, 13]. However, both thin and thick lens models have the same capability to represent the optical properties if the lens thickness has no physical meaning. We thus take the simpler model to analyze the zoom lens systems in our study.

To describe the properties of the camera model, we introduce some notations for real, lens, and model parameters. The **real parameters** denote the actual value of zoom, focus, and iris settings such as a voltage of servo control. We define Z from 25(wide) to 75(telephoto), F from 0(infinite) to 50(nearest), and I from 25(open) to 75(close). The **lens parameters** represent the optical characteristics of the lens. f is a focal length that is a function of Z, U is a focused distance from the lens that is a function of F, and A is a F-number of the lens that is a function of I. The **model parameters** are the effective values that describes the relationship between zoom, focus, and iris. w is an effective focal length that is defined as a distance between the lens and the image plane. d is an effective lens diameter, and B denotes an effective F-number that is defined by w/d which is proportional to A. In addition, we have to introduce another effective parameter that represents the lens center position on the optical axis of the lens. c is defined as a length between R and the lens center O . Here, R denotes the reference point to the real world coordinates. We assign the frontal position of the actual zoom lens to R in our study. Table 2 shows the relationship between real, lens, and model parameters. Using these parameters, the properties of zoom lens can be represented.

In these model parameters, the effective lens diameter d is a function of the effective focal length w and the F-number A. The irradiance E of an infinitesimal area on the image

Table 2: Relationship between real, lens, and model parameters.

real parameter	zoom Z	focus F	iris I
lens parameter	focal length $f(Z)$	focused distance $U(F)$	F-number $A(I)$
model parameter	effective focal length $w(Z,F)$ position of lens center $c(Z,F)$		effective F-number $B(A)$
	effective lens diameter $d(w,A)$		

plane is given by

$$E = L\frac{\pi}{4}\left(\frac{d}{w}\right)^2 \cos^4\theta \tag{1}$$

where L denotes the radiance of an infinitesimal area on the object surface, and θ is an angle of incident light on the lens[14]. This equation tells that image brightness depends upon the effective focal length w which is a function of the focal length f that relies on the zoom setting Z. From the properties of zoom stated in section 2, however, zoom does not affect the image brightness. Thus, the effective F-number $B = w/d$ should be independent of the zoom setting. This means that the effective lens diameter d is proportional to the effective focal length w.

3 Depth Estimation

3.1 Depth from defocus

Fig. 3 illustrates that the width b of blurred image of a point object P is a function of model parameters d, w, c, lens one U, and object distance s. From the geometric relation of these parameters, b is obtained by

$$b = wd\left|\frac{1}{s+c} - \frac{1}{U+c}\right| = \frac{w^2}{B}\left|\frac{1}{s+c} - \frac{1}{U+c}\right| \tag{2}$$

where $B = w/d$.

Using this equation, we have the depth s of an object from its width of blur b, that is,

$$s = \begin{cases} \dfrac{wd(U+c)}{wd+b(U+c)} - c & (s < U) \\ \dfrac{wd(U+c)}{wd-b(U+c)} - c & (s > U) \end{cases} \tag{3}$$

where w, c, d and U are given by zoom, focus and iris settings. Thus, the depth s is computed from the width of blur b at any setting of the three real parameters.

From a set of multiple images taken at different parameters, we can have reliable depth information. Based on the least squares criterion, the minimization of the following error function yields the optimal depth estimation s.

$$Error = \sum_n \left(w_n d_n \left|\frac{1}{s+c_n} - \frac{1}{U_n+c_n}\right| - b_n\right)^2 \tag{4}$$

where the suffix n denotes image number.

Here, it should be noted how to measure the width of blur b. We have the brightness function from the convolution of a step edge with a certain size of circle of confusion, as shown in Fig.4 by

$$g(x) = \frac{1}{2} + \frac{1}{\pi}\left(x\sqrt{1-x^2} + \arcsin x\right), \tag{5}$$

where the edge position is at $x = 0$ and the brightness range is normalized between 0 and 1. The width of the blurred edge is determined by fitting the brightness function to the profile across the blurred edge extracted from a images.

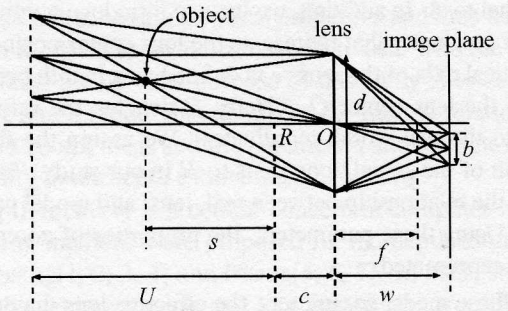


Figure 3: Width of blur b of a point object.

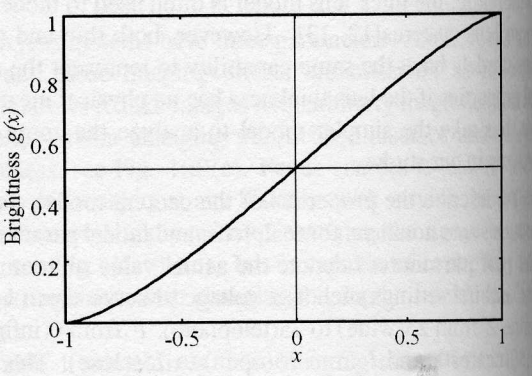


Figure 4: Theoretical function of blur.

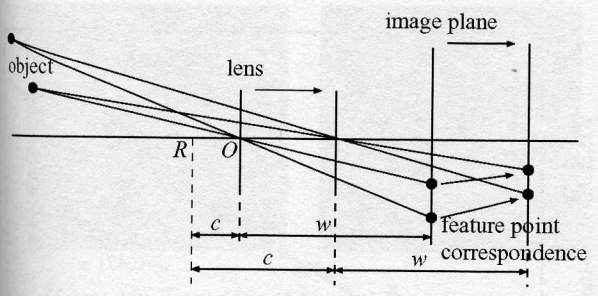


Figure 5: Depth from lens translation.

3.2 Depth from lens translation

Depth from lens translation is a particular case of "stereo" or "shape from motion" problem; that is, the depth information is derived from the correspondence of feature points detected in images. Figure 5 illustrates the lens center translation along the optical axis of the lens and the correspondence of the projected object points between two images.

In general, the pth object point $\boldsymbol{s}_p = (X_p, Y_p, Z_p)^T$ represented in the world coordinates is perspectively projected on the fth image coordinates as $\boldsymbol{u}_{fp} = (x_{fp}, y_{fp})^T$.

$$\boldsymbol{u}_{fp} = \mathcal{P}(M_f \boldsymbol{s}_p + \boldsymbol{t}_f) \tag{6}$$

where M_f and $\boldsymbol{t}_f$ denote the rotation matrix and translation vector respectively, and $\mathcal{P}$ is defined as a perspective projection operator with effective focal length w

$$\mathcal{P}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{w}{z}\begin{bmatrix} x \\ y \end{bmatrix}. \tag{7}$$

Depth recovery, or 3D shape reconstruction in other words, is performed by using the least squares criterion by minimizing the following error function.

$$Error = \sum_{f,p} ||\tilde{\boldsymbol{u}}_{fp} - \boldsymbol{u}_{fp}||^2 \tag{8}$$

where $\tilde{\boldsymbol{u}}_{fp}$ represents the pth feature point detected in the fth image and $\boldsymbol{u}_{fp}$ is the feature point computed from the estimated 3D object point.

In order to reduce the computational complexity of Eq. (8), we convert the images to be taken by a camera that has a fixed w in Eq. (7). This means that the image size is first modified then the depth estimation is performed.

4 Experiments

4.1 Camera Calibration

We used a video camera SONY XC-007 with a zoom lens CANON J16×9.5B4RAS to take real images. The optical characteristics of the zoom lens are focal length f = 9.5mm – 152mm, focused distance U = 1.0m – ∞, and F-number A = 1.8 – close. The spatial and brightness resolutions of images were 512×480 pixels and 256 gray levels for each RGB signal.

Effective focal length w and position of lens center c

We assume that the effective focal length w and the position of lens center c are function of focus setting F and zoom one Z, and that both of them are represented by polynomial function of F and Z, as follows,

$$w(Z,F) = \sum_{i=0}^{I}\sum_{j=0}^{J} a_{ij} F^i Z^j \tag{9}$$

where $w = f$ holds when $F = 0$, that is, infinite distance is in focus, and a_{ij} are coefficients of polynomial function.

$$c(Z,F) = \sum_{k=0}^{K}\sum_{l=0}^{L} b_{kl} F^k Z^l \tag{10}$$

where b_{kl} are coefficients of polynomial function.

To determine w of Eq. (9) and c of Eq. (10), we evaluated the image magnification due to F and Z. From the geometrical relation between object distance and image size in Fig. 6, we have the following equations.

$$\begin{aligned} (s_0 + c) : h &= w : n_0 \\ (s_1 + c) : h &= w : n_1 \end{aligned} \tag{11}$$

$$\begin{bmatrix} \frac{h}{n_0} & -1 \\ \frac{h}{n_1} & -1 \end{bmatrix}\begin{bmatrix} w \\ c \end{bmatrix} = \begin{bmatrix} s_0 \\ s_1 \end{bmatrix} \tag{12}$$

where n_0 and n_1 denote the size of the target on the image plane when the target is set at the distance s_0 and s_1 from the reference point respectively, and h is the width of the target.

From the images taken at F = 5 to 50 and Z = 25 to 75 at intervals of 5, we have determined $w(Z,F)$ and $c(Z,F)$ with the 4th order polynomials of Z by means of the least squares method. The calibration results of lens position

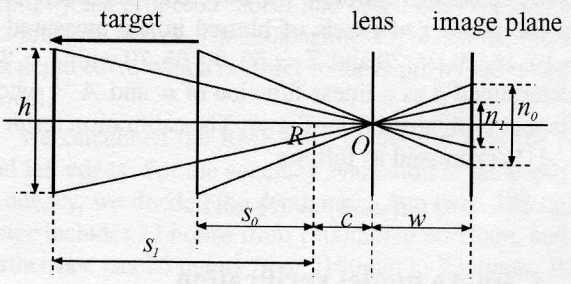


Figure 6: Calibration of w and c.

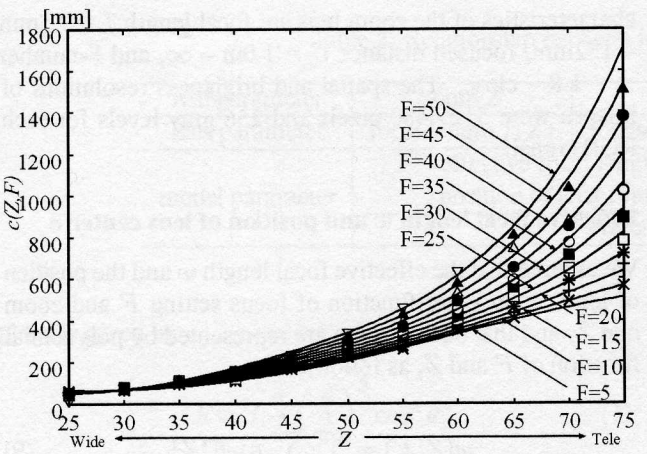


Figure 7: Position of lens center c with regression curves.

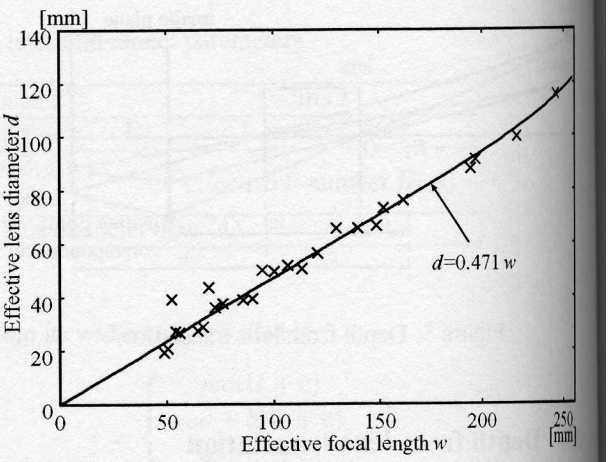


Figure 8: Effective lens diameter d.

with regression curves is shown in Fig. 7, and the function $c(Z, F)$ is represented as follows.

$$c(Z,F) = 199.01 - 4.93F + 8.09 \times 10^{-1}F^2 + (-5.88 - 3.96 \times 10^{-1}F - 6.74 \times 10^{-2}F^2)Z + (-1.52 \times 10^{-1} + 4.17 \times 10^{-2}F + 2.03 \times 10^{-3}F^2)Z^2 + (8.726 \times 10^{-3} - 9.80 \times 10^{-4}F - 2.70 \times 10^{-5}F^2)Z^3 + (-6.56 \times 10^{-5} + 7.05 \times 10^{-6}F + 1.42 \times 10^{-8}F^2)Z^4$$

Effective lens diameter d

The effective lens diameter d and the pixel size β are determined from image blur. Equation (2) tells that d is a function of w, c, U, s and b.

$$d = \frac{b}{w}\left|\frac{1}{s+c} - \frac{1}{U+c}\right|^{-1} \quad (13)$$

Equation (2) also shows that b is a quadric function of w. This consequently leads a simple relation that d is proportional to w.

We used a step edge target and put the target at 1980mm from the camera to calibrate the effective lens diameter d. From the number of pixels of blurred image measured at $F = 10, 15, 20, 25, 30$ and $Z = 55, 60, 65, 70$ with $I = 30$, we determined d as a linear function of w and A^{-1} whose coefficient is obtained from Fig. 8. The calibration result of $d(w, A)$ is expressed as follows.

$$d(w, A) = 1.14wA^{-1}$$

4.2 Camera model verification

A planar target having a step edge was placed at 1008mm, 1508mm, 2008mm, 2508mm and 3008mm from the camera position as shown in Fig. 9. The width b of the blurred edge was measured from each image taken at Z=55,60,65,70,75 and F=10,15,20,25,30,35,40,45,50 with I=40. Fig. 10 shows examples of the blurred step edge images, and Fig. 11 shows the brightness profiles along the horizontal lines around the center of the images with the measured width of them by fitting the theoretical brightness function.

Fig. 12 shows the estimated depth s from the image blur b of multi-zoom images at each F and target position. The solid lines show the depth estimated by the proposed camera model, and the dotted ones show the depth estimated by the previous camera model in [11]. From Figure 12, we find that the depth accuracy is improved by using the proposed camera model in comparison with the previous one.

The accuracy evaluation of average depth, error, and standard deviation for each depth are shown in Table 3. Error and standard deviation increase in the farther distance in both camera models. This degradation is the common ob-

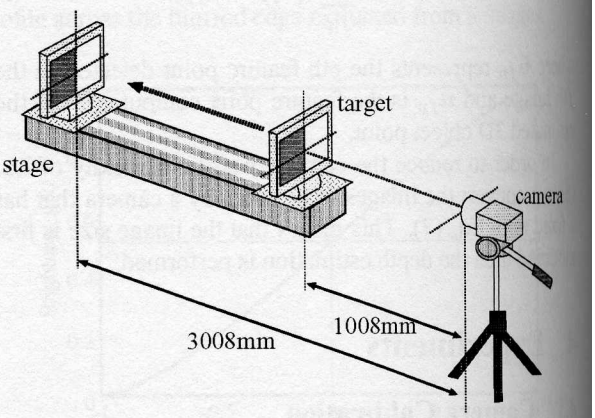


Figure 9: Environment to estimate the depth from defocus.

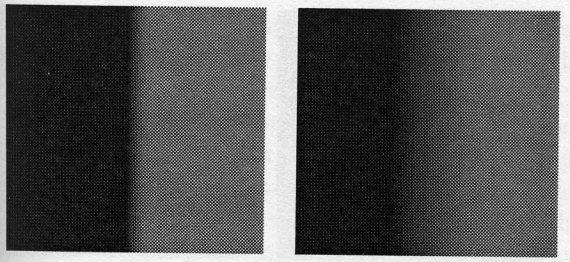


Figure 10: Examples of blurred step edge image.

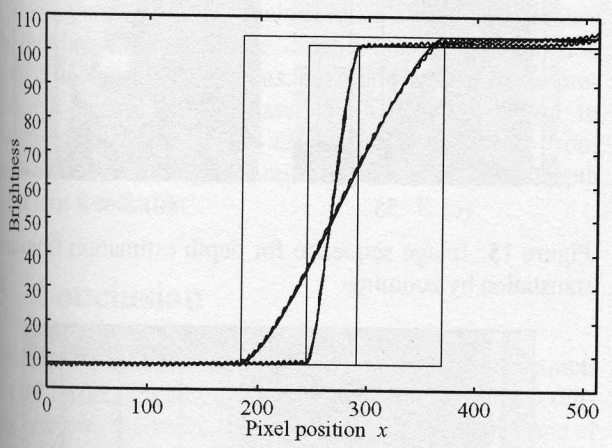


Figure 11: Brightness profile and width of blur.

servation in this type of depth recovery methods. From Fig. 12 and Table 3,we have the following findings.

- The absolute value of the depth error was very small around the focused distance at each focus setting for both models.
- The standard deviation increased in the farther distance for both models, and that of the proposed model was smaller than that of the other one.
- The error was relatively small in the nearer distance for both models, and that of the proposed model was smaller than that of the other one.

This comparative study of the improved and previous camera models strongly indicates that the lens center translation is important in modeling the zoom lens.

4.3 Depth from defocus

We performed the experiment for depth from defocus by using a depth varying target. We placed a toy on a desk and took an image sequence of zooming without changing focus and iris settings. The images were taken at Z=55,60,65,70 with F=24 and I=35, two of which are shown in Fig. 13.

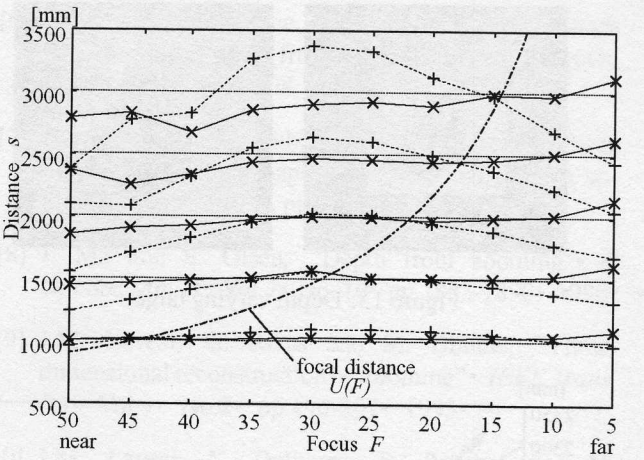


Figure 12: Accuracy evaluation of depth from defocus.

Table 3: Quantitative evaluation of depth from defocus.
(a) proposed camera model

truth	average	error	s.d.
1008	1035.0	27.0	22.3
1508	1518.7	10.7	45.6
2008	1980.4	-27.6	71.0
2508	2439.1	-68.9	93.5
3008	2899.4	-108.6	112.5

unit=[mm]

(b) camera model in [11]

truth	average	error	s.d.
1008	1044.8	36.8	48.4
1508	1421.3	-86.7	95.1
2008	1843.3	-164.7	147.7
2508	2348.8	-159.2	211.3
3008	2921.6	-86.4	338.0

unit=[mm]

The depth was computed from the blur along the left and right edge of the bright plane under the toy. The estimated depth is shown in Fig. 14, where the solid lines represent the depth by the proposed model and doted ones the depth by the previous model. The broken lines represent the theoretical depth curve which is closer to the depth by the proposed model.

We calculated the RMS error of the depth of the right and left edges. For the accuracy evaluation of the depth dependency, we divided the depth range into two. The nearer range includes 22 points from 1900mm to 2050mm, and the farther one has 26 points from 2160mm to 2260mm. RMS error of the nearer range was 4.60 and 3.91 for the farther one. These values were small enough to demonstrate the depth stability of this method.

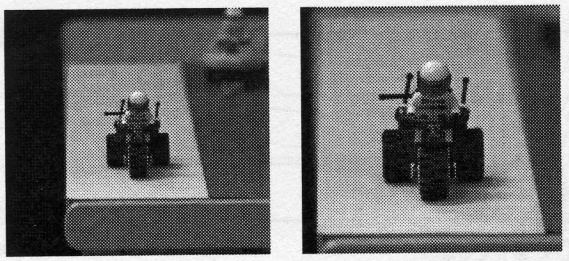


Figure 13: Depth varying target.

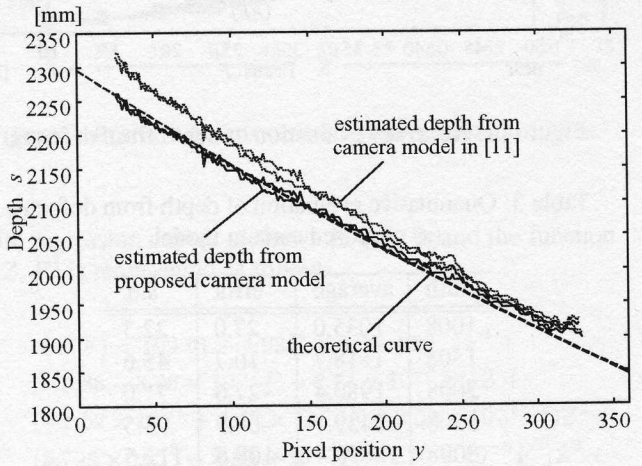


Figure 14: Estimation Result for depth varying target.

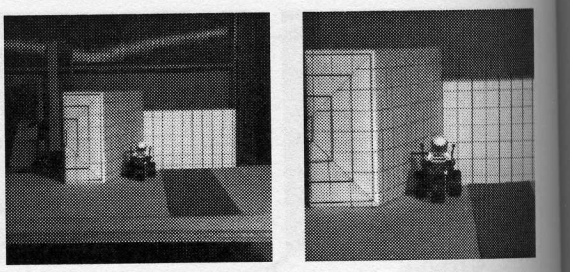
4.4 Depth from lens translation

Two boxes having the texture of the structured pattern were used for depth from lens translation experiment. The pattern was 15mm × 15mm square and the distance was around 1m from the camera position. To eliminate the defocusing effect the images were taken by closing the iris opening. Fig.15 shows the images taken by changing the zoom settings from Z = 35 to 75 at intervals of 5, while focus and iris parameters were fixed at F=46 and I=68. The feature points were detected manually from multi-zoom image sequence, and those drawn on the image of Z=55 were shown in Fig.16.

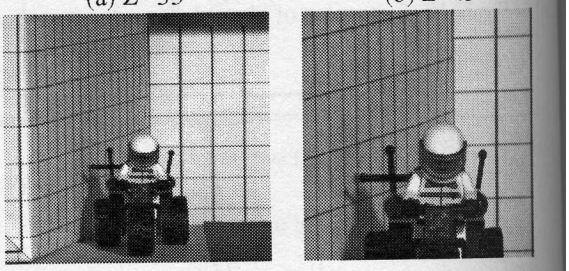
The recovered 3D positions of the feature points were shown in Fig. 17. The left plane was recovered successfully, but the right one was not. The average length between adjacent points on the left plane was 15.16mm which is very close to the truth. The RMS error of the depth on the left plane was 5.74mm, and that on the right one was 53.36mm.

4.5 Discussion

"Depth from lens translation by zooming" is a particular case of the stereo method. Camera moves along the optical axis and depth is recovered based on the principle of



(a) Z=35 (b) Z=45



(c) Z=55 (d) Z=65

Figure 15: Image sequence for depth estimation from lens translation by zooming.

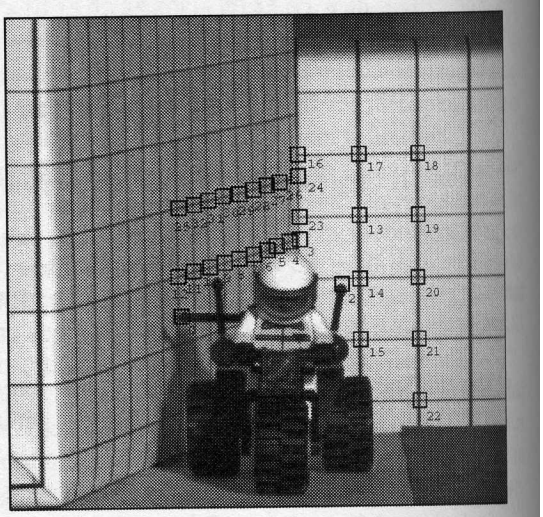


Figure 16: Feature points for depth estimation from translation.

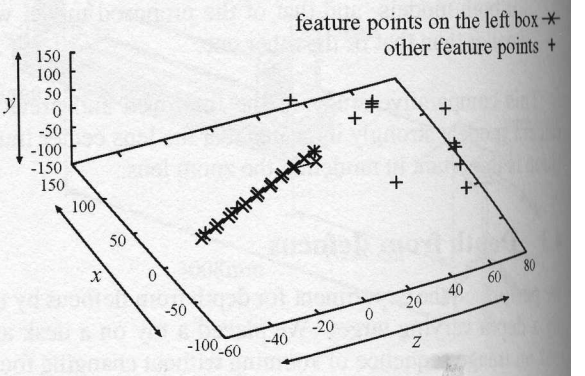


Figure 17: Recovered 3D positions of the feature points.

triangulation. However, the accuracy of this type of stereo is lower than that of ordinary binocular stereo because the forward translation makes smaller disparity.

Moreover, the stereo by forward translation has a serious problem; that is, the depth can not be recovered at the image center because no disparity is observed. And the disparities depends upon the pixel position of the image; that is, the depth from marginal pixels is more accurate than the pixels around the image center. However, the marginal pixels become out of view by changing the zoom parameter from wide to telephoto.

Note that real images include more or less blur because the lens has a finite aperture. Such a blurring phenomenon is used in depth from defocus, but that deteriorates the precision of the feature point based depth recovery method. In our experiments with RMS error evaluation, "Depth from defocus" is promising and relatively accurate than "Depth from lens translation".

5 Conclusion

This paper has presented a novel camera model that accounts for the image blur and the lens center translation by zooming. We have discussed the zoom lens properties in terms of zoom, focus and iris parameters, then proposed a thin lens based camera model that describes the relationship between three parameters. The proposed model has an important parameter that represents the lens center translation. Therefore, our model takes account for the external parameters, i.e. pose and position calibration, as well as the internal parameters, i.e. zoom, focus and iris calibration. The experimental results have demonstrated the validity of our camera model and also shown its applicability to the depth estimation from defocus as well as the lens translation by zooming. In future, we plan to develop a unified method to recover depth information from defocus and translation by zooming.

References

[1] E. Krotkov, "Focusing," *IJCV*, vol.1, no.3, pp.223-237, 1987.

[2] M. Subbrarao, "Parallel depth recovery by changing camera parameters," *Proc. ICCV*, pp.149-155, 1988.

[3] H.N. Nair and C.V. Stewart,"Robust focus ranging," *Proc. CVPR*, 1992, 309-314.

[4] N. Asada, H. Fujiwara and T. Matsuyama, "Edge and depth from focus,"*IJCV*, Vol.26, No.2, 1998, 153-163.

[5] A. Pentland, "A new sense for depth of field," *IEEE Trans. PAMI*, Vol.9, No.4, pp.523-531, 1987.

[6] J. Ens and P. Lawrence, "An investigation of methods for determining depth from focus," *IEEE Trans. PAMI*, Vol.15, No.2, pp97-108, 1993.

[7] G. Surya and M. Subbarao, "Depth from defocus by changing camera aperture: A spatial domain approach," *Proc. CVPR*, pp.61-67, 1993.

[8] J. Ma and S. Olsen, "Depth from zooming"• *J. Opt.Soc. Am. A*• Vol.7• No. 10• pp.1883-1890• 1990• •

[9] J.M. Lavest, G. Rives and M. Dhome, "Three-dimensional reconstruction by zooming"• *IEEE Trans. RA*• Vol.9• No.2• pp.196-207• 1993• •

[10] J.M. Lavest, C. Delherm, B. Peuchot and M. Dhome,"Implicit Reconstruction by zooming," *CVIU*, Vol.66, No.3, pp.301-315, 1997.

[11] N. Asada, M. Baba, and A. Oda, "Depth from blur by zooming," *Proc. Vision Interface*, pp.165-172, 2001.

[12] R. Kingslake, *Optical System Design* (Orlando: Academic Press, Inc., 1983).

[13] K. Tarabanis, R.Y. Tsai and D.S. Goodman, "Calibration of a computer controlled robotic vision sensor with a zoom lens," *CVGIP*, Vol.59, No.2, pp.226-241, 1994.

[14] B.K.P. Horn, *Robot vision* (Cambridge: The MIT Press, 1986).