A Defocus Technique based Depth from Lens Translation using Sequential SVD Factorization

Jong-Il Kim, Hyun-Sik Ahn, Gu-Min Jeong, and Do-Hyun Kim

Department of Electrical Engineering, Kookmin University, Seoul, Korea
(Tel : +82-2-910-5067; E-mail: {indina | ahn | gm1004 | dhkim }@koomin.ac.kr)

Abstract: Depth recovery in robot vision is an essential problem to infer the three dimensional geometry of scenes from a sequence of the two dimensional images. In the past, many studies have been proposed for the depth estimation such as stereopsis, motion parallax and blurring phenomena. Among cues for depth estimation, depth from lens translation is based on shape from motion by using feature points. This approach is derived from the correspondence of feature points detected in images and performs the depth estimation that uses information on the motion of feature points. The approaches using motion vectors suffer from the occlusion or missing part problem, and the image blur is ignored in the feature point detection. This paper presents a novel approach to the defocus technique based depth from lens translation using sequential SVD factorization. Solving such the problems requires modeling of mutual relationship between the light and optics until reaching the image plane. For this mutuality, we first discuss the optical properties of a camera system, because the image blur varies according to camera parameter settings. The camera system accounts for the camera model integrating a thin lens based camera model to explain the light and optical properties and a perspective projection camera model to explain the depth from lens translation. Then, depth from lens translation is proposed to use the feature points detected in edges of the image blur. The feature points contain the depth information derived from an amount of blur of width. The shape and motion can be estimated from the motion of feature points. This method uses the sequential SVD factorization to represent the orthogonal matrices that are singular value decomposition. Some experiments have been performed with a sequence of real and synthetic images comparing the presented method with the depth from lens translation. Experimental results have demonstrated the validity and shown the applicability of the proposed method to the depth estimation.

Keywords: Blurring phenomena, Depth from lens translation, Sequential SVD factorization and defocus.

1. INTRODUCTION

The imaging system has been processed by many researches because of supplying the variety of information in the environments. Recovering three-dimensional geometry of scene and motion of the camera has been studied for the position measurement and recognition of objects due to estimating the three-dimensional geometry of scenes from the stream of two-dimensional images. The researches for depth estimation such as stereopsis, motion parallax and blurring phenomena are based on the camera system and optical phenomena.

Among the researches, depth from lens translation in a specific case of shape from motion is used for a zooming effect that causes by translating the lens center along the optical axis of the lens [1-5]. This effect can be used for both a fixed camera and an unfixed camera because the camera zoom function is equal to the camera translation itself along the optical axis of the lens.

Depth from lens translation using the sequential SVD factorization recovers both the shape of an object and its motion from a sequence of images using many images and tracking many feature points to obtain feature position information. The shape and motion can be obtained by using an orthogonal matrix derived from singular value decomposition (SVD). The method robustly processes the feature trajectory information using SVD, taking advantage of the linear algebraic properties of orthographic projection. The sequential SVD factorization not only solves a problem of the initialization in a disadvantage of Kaman filter, but also processes more rapid than a method using the Kaman filter [6, 7].

In the task of extracting 3D geometry, the approach suffers from the occlusion or missing part problem [6, 7]. The problem of occlusion occurs when the inserted objects are placed behind real objects in the scene. The approach has a disadvantage to limit the image information with the image blur because the image blur is ignored in the feature point detection [4].

This paper presents a novel depth from lens translation to apply to the defocus technique to recovery the 3D geometry of coordinate from the sequence of 2D images. This approach solving the above problem requires how light interacts with the optics until reaching the imaging surface. This interaction is described by the so-called brightness function. The brightness function defines the widths of blur of feature points detected in the edge of objects. The widths of image blur are used in order to recover the depth information [4].

In the following section, we first discuss the optical properties of a camera system because the image blur varies according to three parameter settings, i.e. zoom as well as focus and iris settings. The camera system accounts for the camera model integrating a thin lens based camera model to explain the light and optical properties and a perspective projection camera model to explain the depth from lens translation. Using this camera model, depth from lens translation is proposed to use the feature points detected in edges of the image blur. Finally, some experiments have been performed with real and synthetic images and evaluated the accuracy of the depth information recovered from the proposed method. Experimental results have demonstrated the validity and shown the applicability of the proposed method to the depth estimation.

2. CAMERA MODEL

The coordinate system in Fig. 1 is applied to depth from lens translation. The coordinate system is composed a real world coordinate system ($\vec{x}_f, \vec{y}_f, \vec{z}_f$, the rotation of a camera)

and a shape of objects ($\vec{s}_p$, the coordinate of feature points of an object surface), translation ($\vec{t}_f$).

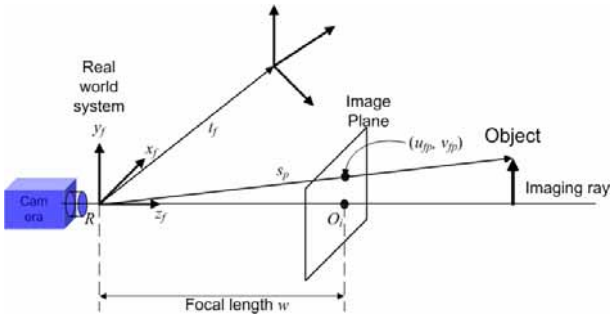


Fig. 1 Coordinate system.

In the imaging systems, zoom lens camera systems have three fundamental parameters, zoom, focus and iris. Those imaging effects and optical properties are observed as shown in Table 1 [3, 4].

Table 1 Imaging effects and optical properties of three fundamental parameters.

	Optical properties	Imaging effects
Zoom	Allow to take images with arbitrary magnification of a scene without changing the focused distance and image intensity	M : ○ F : △ B : △
Focus	Enable to take focused images of an object at any distance with varying the image magnification without changing image intensity	M : △ F : ○ B : △
Iris	Be capable of controlling the image intensity with changing the depth of field	M : X F : △ B : ○

M : magnification, F : focusing, B : brightness
○ : main effect, △ : side effect, X : no effect

Thin lens camera model is built based on the above consideration as shown in Fig. 2.

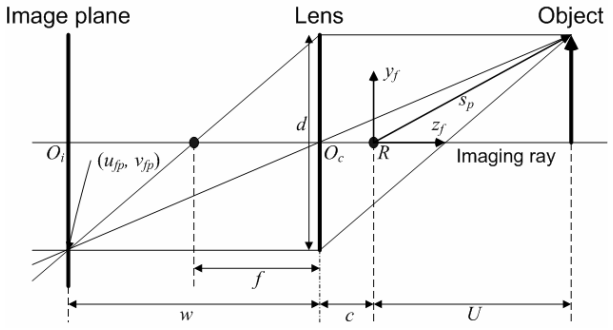


Fig. 2 Thin lens image system of camera model.

The camera model parameters are composed a real, lens and model parameters to describe the properties of the camera model. The real parameters denote the actual value of zoom, focus and iris setting. We define zoom Z from 0(wide) to 16384(tele), focus F from 4096(infinite) to 49152(near), iris I from 0(open) to 15(close). The lens parameters represent the optical characteristics of the lens. The model parameters are the effective values that describe the relationship between zoom, focus and iris. An additional effective parameter $c(Z, F)$ represented the lens center position on the optical axis of the lens is defined the distance between the reference point R that is the origin of the real world coordinate and the lens center O_c. Table 2 summarizes these camera model parameters and some notations [4]. Other parameters stated in section 4.

Table 2 Camera model parameters.

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

3. A DEFOCUS BASED DEPTH FROM LENS TRANSLATION USING SEQUENTIAL SVD FACTORIZATION

3.1 Depth estimation using the width of blur

It illustrates how a point object P is projected on the image plane in the image blur. The object distance D is a function of model parameters d, w, c, lens parameter U and width b of blur. From the geometric relation in the Fig. 3, the object distance D is obtained by a following equation.

$$D = \begin{cases} \dfrac{wd(U+c)}{wd+b(U+c)} - c & (D < U) \\ \dfrac{wd(U+c)}{wd-b(U+c)} - c & (D > U) \end{cases} \quad (1)$$

Where w, c, d and U are given by zoom, focus and iris settings. The depth D is computed from any setting of the three real parameters [4].

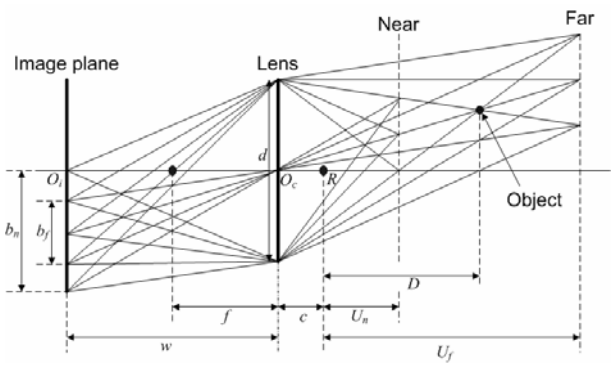


Fig. 3 Width b of blur of a point object.

The method to measure the width b of blur is obtained from the brightness function. The brightness function is the first derivatives of a step edge as shown in Fig. 4. Here, the edge position is the position of the maximum (rising edge) and minimum (falling edge) value and the brightness range is normalized between -255 and 255. The width of the blurred edge is determined the size between rising and falling point (on the opposite) corresponding to the profile across the blurred edge extracted from an image.

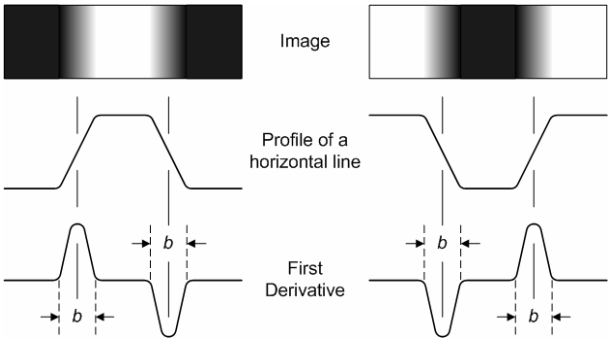


Fig. 4 Brightness function.

3.2 Depth from lens translation

In general, the pth object point $s_p = (x_p, y_p, z_p)^T$ represented in the world coordinates is the perspective projection on the fth image frame coordinates as $i_{fp} = (u_{fp}, v_{fp})^T$. The standard perspective projection equation is

$$i_{fp} = \boldsymbol{P}\left(M_f s_p + t_f\right) \tag{2}$$

where M_f and t_f denote the rotation matrix and translation vector respectively, and P is defined as a perspective projection operator. The M_f and t_f can be obtained by the rotation and translation of the camera respectively. The Eq. (2) is rewritten as the following equation [4, 5].

$$\frac{w}{z}\begin{bmatrix} u \\ v \end{bmatrix} = \boldsymbol{P}\begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{3}$$

where the z_p of pth object point is the depth D obtained by depth estimation. We can know the pth object point s_p, and recover the three-dimensional geometry of scene. Also, the tracking of feature points is used the method that the current feature point positions are derived from the previous features on the fixed objects and the translated camera. Thus, we can solve the problem of the occlusion and missing part by detecting the estimated object point in the object point extracted from the image. Based on the least squares criterion, the minimization of the following error function yields the optimal tracking of feature points.

$$Error = \sum_{f,p} \left\| \tilde{s}_{fp} - s_{fp} \right\| \tag{4}$$

where $\tilde{s}_{fp}$ represents the pth object points extracted from the images and s_{fp} is the estimated object points.

3.3 Sequential SVD factorization

For the shape and motion space are separated by using the SVD, the Eq. (3) by an arbitrary point is

$$u_{fp} = m_f \cdot s_p + t_{xf}, \quad v_{fp} = n_f \cdot s_p + t_{yf} \tag{5}$$

where

$$m_f = \frac{w}{z_f} x_f, \quad n_f = \frac{w}{z_f} y_f. \tag{6}$$

The Eq. (5) is rewritten as the following matrix equation by accounting for the pth object point and the fth image frame.

$$\begin{bmatrix} u_{11} & \cdots & u_{1p} \\ \vdots & \vdots & \vdots \\ u_{f1} & \cdots & u_{fp} \\ v_{11} & \cdots & v_{1p} \\ \vdots & \vdots & \vdots \\ v_{f1} & \cdots & v_{fp} \end{bmatrix} = \begin{bmatrix} m_1 \\ \vdots \\ m_f \\ n_1 \\ \vdots \\ n_f \end{bmatrix} \begin{bmatrix} s_1 & \cdots & s_p \end{bmatrix} + \begin{bmatrix} t_{x1} \\ \vdots \\ t_{xf} \\ t_{y1} \\ \vdots \\ t_{yf} \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}$$

or simply,

$$W = MS + T\begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \tag{7}$$

where W is the 2FxP measurement matrix whose each column contains the observations for a single point, while each row contains the observed u or v –coordinates for a single frame, M is the 2Fx3 motion matrix whose rows are the m_f and n_f vectors, S is the 3xP shape matrix whose columns are the s_p vectors and T is the 2Fx1 translation vector whose elements are the t_x and t_y.

The translation vector T of Eq. (7) is the subtracted from W, leaving a registered measurement matrix as the following equation.

$$W^* = W - T\begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} = MS \tag{8}$$

where W^* is the product of a 2Fx3 motion matrix M and a 3xP shape matrix S, its rank is at most three. When noise is present in the input, the W^* will not be exactly of rank three. We use SVD to remove the noise presenting the image and separate the M and S [8].

If the SVD is the measurement matrix W^*, the rank of W^* get the values from the very high three singular values relatively and the singular values approximated the zero by the

noise. The values excepting to the three singular values make zero. Thus, the SVD of $W^* \in R^{2FxP}$ is represented by two orthogonal matrix $Q_1 \in R^{2Fx3}$, $Q_2 \in R^{Px3}$ and diagonal matrix $\sum \in R^{3x3}$.

$$W^*_{SVD} = Q_1 \sum Q_2^T \tag{9}$$

where $\sum = diag(\sigma_1, \sigma_2, \sigma_3)$ and $\sigma_1 \geq \sigma_2 \geq \sigma_3 > 0$.

We can separate the W^*_{SVD} such as the following equation.

$$W^*_{SVD} = \hat{M}\hat{S} \tag{10}$$

$$\hat{M} = Q_1 \Sigma, \quad \hat{S} = Q_2^T \tag{11}$$

where $\hat{M} = [\hat{m}_1 \cdots \hat{m}_f \quad \hat{n}_1 \cdots \hat{n}_f]^T$ and $\hat{S} = [\hat{s}_1 \cdots \hat{s}_p]$.

The column space spanning by $\hat{M}$ is the motion space and the row space spanning by $\hat{S}$ is the shape space. We can know the dimension of the space by the rank theory. In view of the results, the high dimension input space is to detect the two subspaces $\hat{M}$ and $\hat{S}$ [9].

4. EXPERIMENTS

4.1 Camera Calibration

We used a zoom lens camera system SONY FCB-EX780 to take both real and synthetic images. The spatial and brightness resolutions of images were 640x480 pixels and 256 gray levels for each RGB signal. The optical characteristics of the zoom lens are given as focal length f = 2.4㎜ ~ 60㎜, focused distance U = (wide)35㎜ ~ ∞ and (tele)800㎜ ~ ∞ and F-number A = 1.6 ~ 2.7. The camera calibration experiments have been done based on the methods in [3, 4].

Using the above optical characteristics and the calibration target that consists of a structured pattern, we obtained the following relationship between real and lens parameters.

$$f(Z) = 2.2352 \times 10^{-2} Z^2 + 2.4 \times 10^{-3} \tag{12}$$

$$U(Z,F) = \frac{3.5251 \times 10^4}{F - 4096} + 4.6692 \times 10^{-5} Z - 7.437410^{-1} \tag{13}$$

$$A(I) = 1.6\sqrt{2}^{(1.00653 \times 10^{-1} I)} \tag{14}$$

Also, in the model parameters, the effective focal length $w(Z, F)$ and position of lens center $c(Z, F)$ are determined by focus and zoom settings. Thus, the image magnification has been evaluated due to F and Z. From the geometrical relation between object distance and image size in Fig. 6, we have the following equation.

$$\begin{bmatrix} h/H_0 & -1 \\ h/H_1 & -1 \end{bmatrix} \begin{bmatrix} w \\ l \end{bmatrix} = \begin{bmatrix} D_0 \\ D_1 \end{bmatrix} \tag{15}$$

where H_0 and H_1 denote the size of the target on the image plane when the target is set at the distance D_0 and D_1 from the reference point respectively, and h is the width of the target [4].

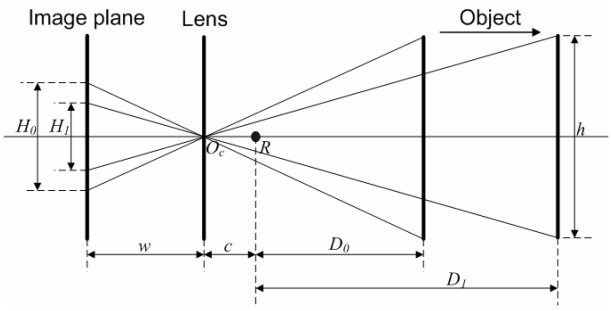


Fig. 5 Calibration of the effective focal length $w(Z, F)$ and position of lens center $c(Z, F)$

The effective lens diameter $d(w, A)$ is determined from the blurred image of a step edge. Fig. 3 illustrates the width of blurred b image of a point object P [4].

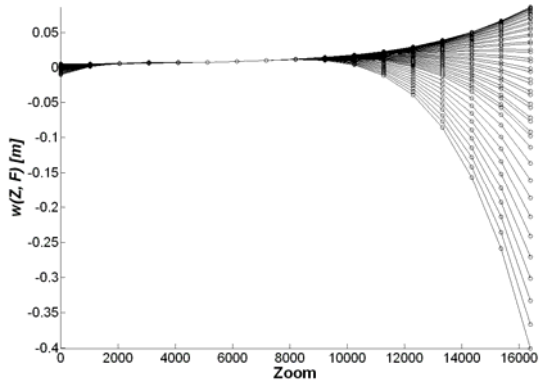
$$d = \frac{b}{w}\left|\frac{1}{D+c} - \frac{1}{U+c}\right|^{-1} \tag{16}$$

The planar target having a step edge was placed at 1004㎜, 1507㎜ and 1998㎜ from the camera position. And from the images taken at F = 4096 to 49152 and Z = 0 to 16384 at intervals of 1048 and 512 respectively with I = 8, we have determined $w(Z, F)$ and $c(Z, F)$ with 4th order polynomials of Z by means of the least squares method and $d(w, A)$ as a linear function of $w(Z, F)$ and A^{-1} by the number of pixels of blurred edge taken the images. The calibration results of effective focal length and lens position are shown in Fig. 6, and the functions $w(Z, F)$, $c(Z, F)$ and $d(w, A)$ are represented as follows.

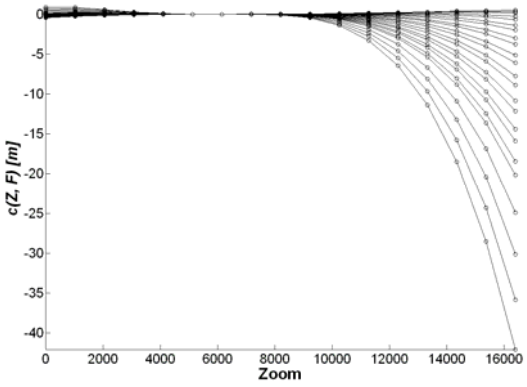
$$\begin{aligned} w(Z,F) = {} & -3.4869 \times 10^{-4} + 6.6566 \times 10^{-7} F - 1.8589 \times 10^{-11} F^2 + \\ & (6.7163 \times 10^{-6} - 7.8093 \times 10^{-10} F + 2.0152 \times 10^{-14} F^2) Z + \\ & (-2.3203 \times 10^{-9} + 2.7422 \times 10^{-13} F - 6.8289 \times 10^{-18} F^2) Z^2 + \\ & (3.3866 \times 10^{-13} - 3.8075 \times 10^{-17} F + 9.2851 \times 10^{-22} F^2) Z^3 + \\ & (-1.6298 \times 10^{-17} + 1.8434 \times 10^{-21} F - 4.4345 \times 10^{-26} F^2) Z^4 \end{aligned}$$

$$\begin{aligned} c(Z,F) = {} & 2.9836 \times 10^{-1} - 7.226 \times 10^{-5} F + 1.7634 \times 10^{-9} F^2 + \\ & (2.3038 \times 10^{-4} + 1.3084 \times 10^{-9} F - 6.773 \times 10^{-14} F^2) Z + \\ & (-1.6693 \times 10^{-7} + 1.1243 \times 10^{-11} F - 2.5425 \times 10^{-16} F^2) Z^2 + \\ & (2.9919 \times 10^{-11} - 2.4334 \times 10^{-15} F + 5.578 \times 10^{-20} F^2) Z^3 + \\ & (-1.6533 \times 10^{-15} + 1.4418 \times 10^{-19} F - 3.3149 \times 10^{-24} F^2) Z^4 \end{aligned}$$

$$d(w, A) = 12.3049 w A^{-1}$$



(a) Effective focal length $w(Z, F)$



(b) Position of lens center $c(Z, F)$

Fig. 6 Effective focal length w and position of lens center c.

We evaluate the depth estimation by these parameters. The accuracy evaluation of average depth, error and standard deviation for each depth are shown in Table 3 and the estimated depths D are shown in Fig. 7. In view of the results, the absolute value of the depth error was very small around the focused distance at each focus setting. The error was relatively small in the nearer distance. These results are an important factor in modeling the lens center translation.

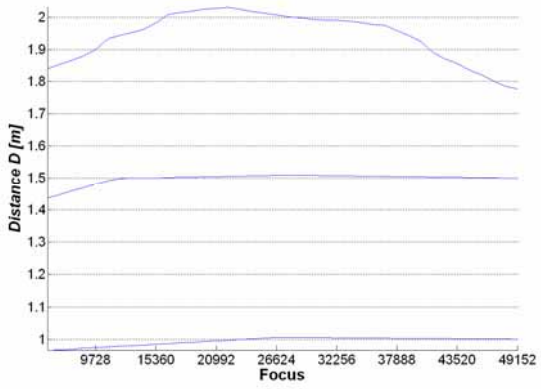


Fig. 7 Accuracy evaluation of depth estimation.

Table 3 Evaluation of depth estimation.

True [m]	Average [m]	Error [m]	S.D.
1.004	0.99476	0.00924	0.01236
1.507	1.49861	0.00839	0.01652
1.998	1.94143	0.05657	0.07573

4.2 Depth from lens translation

The object using in this experiments was a cylinder shape and a bottom with a chess board shapes. The structured pattern of the chess board was 15㎝ x 15㎝ square and distance from the camera was about 1.5m. When the image was captured the object was fixed and the camera was translated the object into the center, the images were taken by Z = 8192, F = 21276 and I = 8.

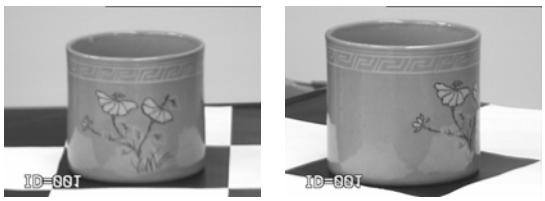


Fig. 8 Cylinder shaped object using in experiments.

In general cases, the strategy for the feature point detections was to use only regions with a rich enough texture. In this spirit, we used the method to detect positions where first-derivatives were sufficiently high. This method assumes that the columns and rows on the images are related to the distance between the camera system and the objects, and the blurring phenomena according to the distance respectively. Based on this assumption, depth information was derived from the width of blur due to the blurring phenomena. The recovered 3D positions of the feature points were shown in Fig. 9. The average error of the depth of the recovered 3D positions was 13.429㎜. This value was small enough to demonstrate the depth stability of this method.

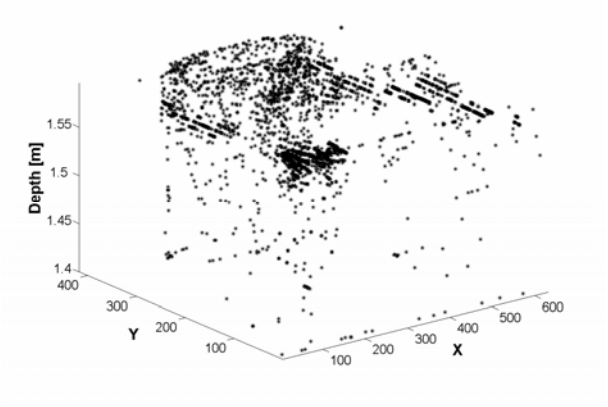


Fig. 9 Recovered 3D positions of the feature points.

The feature point tracking was evaluated to track the feature points by translating the camera. Fig. 10 shows the translated camera. The distances of the image center between the camera

and the object were D_1 = 1508.153㎜ in the first frame and D_2 = 1599.596㎜ in the second frame. The rotation and translation are 45° on the right hand direction and t_x = 1131㎜, t_z = 377㎜ respectively.

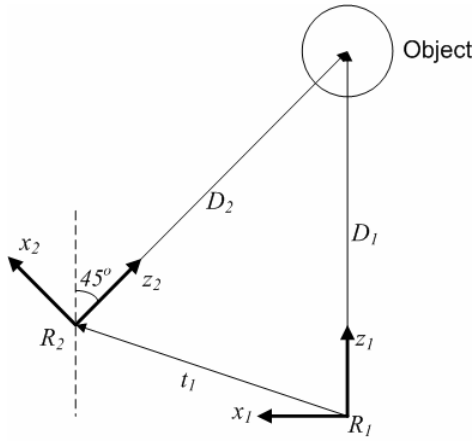


Fig. 10 Experiment for feature point tracking.

We calculated the RMS error between the corresponding real world coordinates of the estimated feature point positions and the corresponding of the feature point positions on the second frame. The RMS error in the experiment results was 168.33㎜. Though this value was not small enough, the missing part problem could be solved.

5. CONCLUSION

This paper has presented a novel depth from lens translation that accounts for the image blur and solves the missing part problem. The optical properties of a camera system are discussed, because the image blur varies according to camera parameter settings. The camera system accounts for the camera model integrating a thin lens based camera model and a perspective projection camera model. This proposed method utilizes the depth estimated the feature points detected in edges of the image blur. Therefore, the proposed method takes account for the missing part problem and the information in the image blur. The experimental results have demonstrated the validity and shown its applicability to the estimation depth, shape and motion.

REFERENCES

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

[2] J.M. Lavest, C. Delherm, B. Peuchot and M. Dhome, "Implicit Reconstructin by zooming," *CVIU*, vol. 66, No.3, pp. 301-305, 1997.

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

[4] M. Baba, N. Asada, A. Oda and T. Migita, "A thin lens based camera model for depth estimation from defocus and translation by zooming," *Proceeding of 15th International Conference on Vision Interface (VI2002)*, pp. 274-281, 2002.

[5] R. Szeliski and S.B. kang, "Recovering 3D Shape and Motion from Image Streans using Non-Linear Least Squares," *Journal of Visual Communication and Image Representation*, 5(1), pp. 10-28, 1994.

[6] J. Marshall, C. Burbeck, and D. Ariely, "Occlusion edge blur: A cue to relative visual depth," *Int'l J. Opeical Soc. Am. A*, Vol. 13, pp. 681-688, 1996.

[7] M. Watanabe and S. Nayar, "Rational Filters for Passive Depth from Defocus," *Int'l J. Computer Vision and Pattern Recognition*, Vol. 27, no. 3, pp.203-225, 1998.

[8] C.J. Poelman and T.Kanade, "A paraperspective factorization method for shape and motion recovery," *Technical Report CMU-CS-93-219*, Carnegie Mellon Univ., 1993.

[9] B.O. Jung, H.G. Kim and H.S. Ko, "3D shape and motion recovery using SVD from image sequence," *KIEE*, Vol. 35, no. 3, pp. 480-488, March 1998.

[10] Y.I. Ohta, K. Maenobu and T. Sakai, "Obtaining surface orientation from texels under perspective projection," *In Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pp. 746-751, August 1981.