Absolute Scale Structure from Motion Using a Refractive Plate

Akira Shibata, Hiromitsu Fujii, Atsushi Yamashita and Hajime Asama

***Abstract*— Three-dimensional (3D) measurement methods are becoming more and more important to obtain information about the surrounding environment in many fields. Structure from Motion is one 3D measurement method using a single camera. This technique calculates the 3D reconstruction of objects from images captured by a single camera with motion. The camera motion can be estimated simultaneously as well as 3D information of objects. However, it is impossible to calculate the absolute scale of objects using Structure from Motion. Only the relative position can be obtained. In this paper, we propose a Structure from Motion method which can calculate the absolute scale of objects using refraction. Refraction occurs when light ray passes through different media. The absolute scale can be calculated using this change in light ray path. In order to generate the refraction, a refractive plate is placed in front of a camera. In a previous study we proposed a Structure from Motion method using refraction. However, the refractive plate was required to be placed perpendicular to the optical axis of the camera. In this new method, the refractive plate can be placed with any orientation in front of camera. In this aspect, the proposed method is more general than the previous one.**

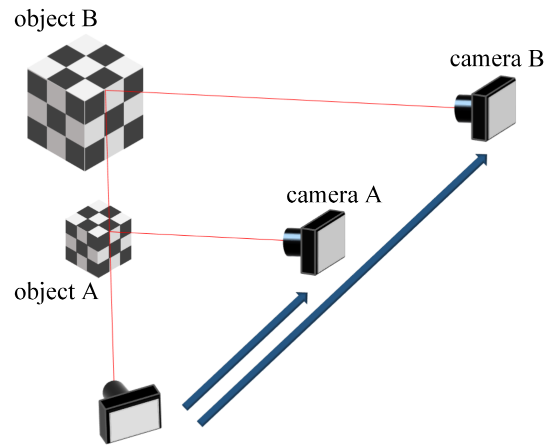


Fig. 1. In this case, the image of object A captured by camera A and the image of object B captured by camera B should be similar. This is the reason why it is impossible to estimate the absolute scale of object using Structure from Motion.

I. INTRODUCTION

Three-dimensional (3D) measurement is important to obtain information about the environment. Structure from Motion is a 3D measurement methods. This technique uses images of objects captured by a single camera with motion, and calculates both the 3D information of objects and the camera motion simultaneously. Structure from Motion is useful in various situations because only a single camera is required, and is actively studied these days [1]. However, the scale of camera translation cannot be estimated in conventional Structure from Motion. Therefore, it is impossible to calculate the absolute scale of objects. Some position constraints on the camera and on objects or geometric information is required to reconstruct objects with absolute scale using conventional Structure from Motion. The Structure from Motion with absolute scale when the camera is placed on a vehicle was proposed in [2]. This method uses the geometric relation of vehicle motion. However, this method is effective only when vehicles can be used. In order to solve this problem, we developed the scale reconstructable Structure from Motion method using refraction [3]. Refraction induced by introducing a different medium results in a change in the light ray path. Especially in water, refraction is studied in the field of computer vision. The images obtained in water environment are distorted because of refraction. The method to remove the distortion by calculating the effect of refraction was reported in [4], [5]. Studies of Structure from Motion considering the refraction in water were also proposed [6]–[8]. However, the purpose of these studies was to remove the effect of refraction as a distortion.

On the other hand, some studies of 3D measurement methods use the refraction actively [9]–[12]. Still, a study of Structure from Motion using refraction actively has not been conducted.

Therefore, we propose a Structure from Motion method which can calculate the absolute scale of object using refraction. In our previous study [3], the refractive plate was required to be placed perpendicular to the optical axis of the camera. In this paper, we propose a method with no constraint of position and orientation of the refractive plate. Therefore, this method could be considered more general than the previous one.

II. STRUCTURE FROM MOTION METHOD WITH ABSOLUTE SCALE

A. Approach

Structure from Motion can estimate the 3D information of objects from only images captured by a single camera. The camera motion can be calculated simultaneously. However the problem is that the absolute scale of object cannot be estimated.

This is because the captured images should be similar when the camera translation and size of the object change in the same ratio. For example, in Fig. 1, the images captured by camera A and camera B should be similar. Therefore we cannot distinguish these images. We solve this problem using refraction. Refraction occurs when light ray enters a

A. Shibata,H. Fujii, A. Yamashita, and H. Asama are with Department of Precision Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan {shibata, fujii, yamashita, asama}@robot.t.u-tokyo.ac.jp

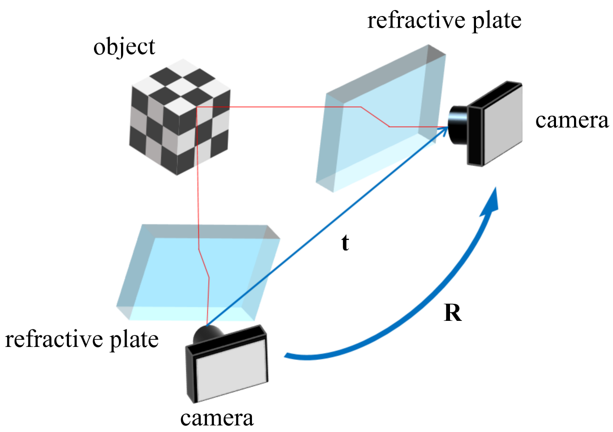


Fig. 2. Our approach to solve the problem of Structure from Motion is using refraction. In order to generate the refraction, the refractive plate is placed in front of a camera. In the previous method, the refractive plate is required to be placed perpendicular to the optical axis of the camera. On the other hand, the proposed method is more general method because the refractive plate can take any orientation.

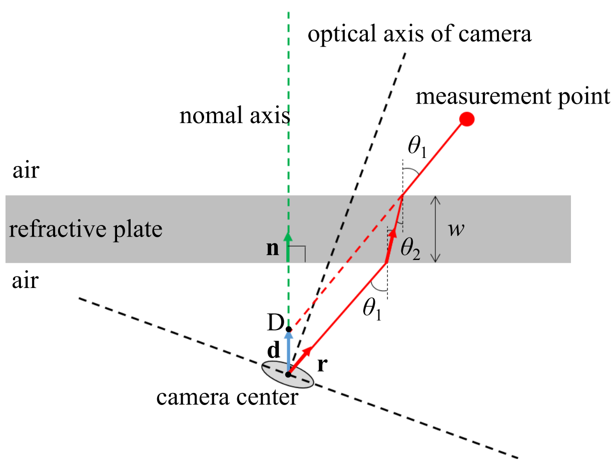


Fig. 3. The light ray path from the measurement point to the camera center. The refraction occurs twice between air and the refractive plate. The difference caused by refraction can be expressed as vector d which has the same direction as the normal of the refractive plate.

different medium, and changes its path. The refraction angle depends on the positions of camera and object. Therefore, in the situation mentioned before, the images become different. It is possible to calculate the absolute scale of object by using this change in proposed method.

The refractive plate is placed in front of the camera to generate the refraction (Fig. 2). In the proposed method, the refractive plate can take any orientation.

B. Calculation method

In this section, the procedure of calculation considering the refraction is explained.

The light ray path from the measurement point to the center of the camera is calculated. Our system is shown in Fig. 3. The refractive plate is placed in front of the camera. The camera is calibrated and the thickness and the normal vector of the refractive plate are known. The medium between the measurement point and the refractive plate, and

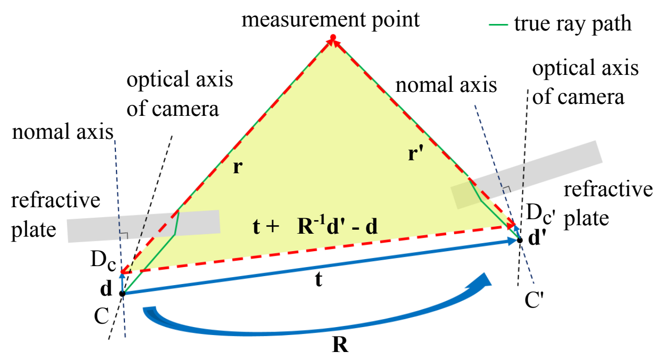


Fig. 4. The geometric relation between two cameras and the measurement points. Ray vectors $\mathbf{r}$ and $\mathbf{r}'$, and the vector between D and D$'$ should be on the same plane.

between the refractive plate and the camera is air. Refractive indices of the air and the refractive plate are known.

Refraction occurs twice between measurement point and the camera in this system. The ray path between the measurement point and the refractive plate is called outer ray, and the ray path between the refractive plate and the camera is called inner ray. As a result, refraction causes a difference between inner ray and outer ray. Let the intersection point of outer ray and the normal vector of the refractive plate from the camera center be point D (Fig. 3). The difference caused by refraction can be expressed as vector $\mathbf{d}$ which is the position vector of point D,

$$\mathbf{d} = d\mathbf{n}, \tag{1}$$

where $\mathbf{n}$ is normal vector of the refractive plate. d is defined as a parameter of the effect of refraction.

In a previous study [3], the refractive plate should be placed perpendicular to the optical axis of the camera. Therefore, the vector $\mathbf{d}$ was always $\mathbf{d} = (0, 0, d)$. However in the proposed method, $\mathbf{d}$ is expressed as Eq. (1), because the direction of normal of refractive plate does not have to be equal to the optical axis of the camera.

In general, refraction occurs according to Snell's law as follows.

$$n_1 \sin\theta_1 \quad = \quad n_2 \sin\theta_2, \tag{2}$$

where n_1 and n_2 are the refractive indices of the air and the refractive plate respectively, and known parameters. θ_1 and θ_2 are the angles between the normal of refractive plate and ray path in air and in the refractive plate respectively.

The light ray paths from camera center to the measurement point are on the same plane. Therefore the length of $\mathbf{d}$ can be calculated using a geometric relation as follows.

$$d = w\left(1 - \frac{\mathbf{r}^{\mathrm{T}} \cdot \mathbf{n}}{\sqrt{\left(\frac{n_2}{n_1}\right)^2 - \|\mathbf{r} \times \mathbf{n}\|^2}}\right), \tag{3}$$

where w is the thickness of the refractive plate. Equation (3) shows that $\mathbf{d}$ depends on the thickness of the refractive plate and is independent of the distance between the refractive plate and the camera center. This means that the refractive plate can be placed at arbitrary position in front of camera. $\mathbf{r}$ is the ray vector in air and $\mathbf{n}$ is the normal vector of the refractive plate. $\mathbf{r}$ can be calculated from image coordinates of corresponding point.

Now, the measurement point is captured from two viewpoints C and C′ (Fig. 4). The point D of each camera coordinate system is $\mathrm{D_c}$ and $\mathrm{D_{c'}}$, respectively. Let $\mathbf{r}$ and $\mathbf{r}'$ be ray vectors to the measurement point from $\mathrm{D_c}$ and $\mathrm{D_{c'}}$, respectively. Ray vectors $\mathbf{r}$ and $\mathbf{r}'$, and the vector between $\mathrm{D_c}$ and $\mathrm{D_{c'}}$ should be on the same plane. This relation can be expressed as follows.

$$\left\{(\mathbf{t}+\mathbf{R}^{-1}\mathbf{d}'-\mathbf{d})\times\mathbf{R}^{-1}\mathbf{r}'\right\}^{\mathrm{T}}\mathbf{r}=0, \tag{4}$$

where $\mathbf{R}$ is rotation matrix from coordinate C to C′. $\mathbf{t}$ is the transfer vector from the center of coordinate C to the center of coordinate C′. Vectors $\mathbf{d}$ and $\mathbf{d}'$ are position vectors of $\mathrm{D_c}$ and $\mathrm{D_{c'}}$ in the camera coordinate system.

Let the elements of vectors be $\mathbf{r}=(x,y,z)^{\mathrm{T}}$, $\mathbf{d}=(d_1,d_2,d_3)^{\mathrm{T}}$, $\mathbf{r}'=(x',y',z')^{\mathrm{T}}$, $\mathbf{d}'=(d'_1,d'_2,d'_3)^{\mathrm{T}}$, and Eq. (4) can be expressed using the orthogonality of the rotation matrix, as follows.

$$\begin{pmatrix} xx' \\ yx' \\ zx' \\ xy' \\ yy' \\ zy' \\ xz' \\ yz' \\ zz' \\ d_3yx'-d_2zx'+d'_3xy'-d'_2xz' \\ d_1zx'-d_3xx'+d'_3yy'-d'_2yz' \\ d_2xx'-d_1yx'+d'_3zy'-d'_2zz' \\ d_3yy'-d_2zy'+d'_1xz'-d'_3xx' \\ d_1zy'-d_3xy'+d'_1yz'-d'_3yx' \\ d_2xy'-d_1yy'+d'_1zz'-d'_3zx' \\ d_3yz'-d_2zz'+d'_2xx'-d'_1xy' \\ d_1zz'-d_3xz'+d'_2yx'-d'_1yy' \\ d_2xz'-d_1yz'+d'_2zx'-d'_1zy' \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} r_{12}t_3-r_{13}t_2 \\ r_{13}t_1-r_{11}t_3 \\ r_{11}t_2-r_{12}t_1 \\ r_{22}t_3-r_{23}t_2 \\ r_{23}t_1-r_{21}t_3 \\ r_{21}t_2-r_{22}t_1 \\ r_{32}t_3-r_{33}t_2 \\ r_{33}t_1-r_{31}t_3 \\ r_{31}t_2-r_{32}t_1 \\ r_{11} \\ r_{12} \\ r_{13} \\ r_{21} \\ r_{22} \\ r_{23} \\ r_{31} \\ r_{32} \\ r_{33} \end{pmatrix} = 0. \tag{5}$$

Equation (5) can be simply expressed as an inner product as follows.

$$\mathbf{u}^{\mathrm{T}}\mathbf{g} \;=\; 0, \tag{6}$$

where, $\mathbf{u}$ is a known vector which has known parameters and $\mathbf{g}$ is a vector which has unknown parameters. r_{ij} is the i, j-th element of rotational matrix $\mathbf{R}$, and t_i is the i-th element of transfer vector $\mathbf{t}$.

For each corresponding point, Eq. (6) is obtained. Therefore, let the known vector $\mathbf{u}$ for the k-th point of all n points be $\mathbf{u}_k$, and define $\mathbf{U}$ as follows,

$$\mathbf{U}=(\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3,\cdots,\mathbf{u}_k,\cdots,\mathbf{u}_n)^{\mathrm{T}}, \tag{7}$$

then Eq. (6) can be simply expressed as,

$$\mathbf{U}\mathbf{g}=\mathbf{0}. \tag{8}$$

Therefore $\mathbf{g}$ can be calculated using least-squares method. At least 17 points are required to solve Eq. (8) in this method. The initial solution for the least-squares method is obtained from the eigenvector for smallest eigenvalue of $\mathbf{U}^{\mathrm{T}}\mathbf{U}$. This is the same way as conventional Structure from Motion.

Let g_i be the i-th element of $\mathbf{g}$. Then $g_{10}\sim g_{18}$ are the same as each element of $\mathbf{R}$, respectively. Therefore, from the orthonormality of the rotation matrix, the constraints are obtained as follows.

$$\begin{aligned} {g_{10}}^2+{g_{11}}^2+{g_{12}}^2 &= 1, \\ {g_{13}}^2+{g_{14}}^2+{g_{15}}^2 &= 1, \\ (g_{10},g_{11},g_{12})^{\mathrm{T}}\times(g_{13},g_{14},g_{15})^{\mathrm{T}} &= (g_{16},g_{17},g_{18})^{\mathrm{T}}, \\ (g_{10},g_{11},g_{12})\cdot(g_{13},g_{14},g_{15})^{\mathrm{T}} &= 0. \end{aligned} \tag{9}$$

These constraints enable us to calculate the norm of $\mathbf{g}$. This is why the absolute scale of object can be estimated in proposed method. Equations (9) are applied to the least-squares method using the Lagrange multiplier method.

When the refractive plate is placed perpendicular to the optical axis of the camera, the normal vector of the refractive plate should be $\mathbf{n}=(0,0,1)^{\mathrm{T}}$. This leads the 18-th element of known vector $\mathbf{u}$ to be 0. In this case,

$$\mathbf{g}=(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1)^{\mathrm{T}}, \tag{10}$$

can be obtained as the initial solution to Eq. (8) by the method mentioned before. However it is clear that Eq. (10) does not satisfy the constraint in Eqs. (9). Therefore we should avoid Eq. (10) while choosing the initial solution. Therefore, placing the refractive plate perpendicular to the optical axis of the camera is a particular case of this method.

After $\mathbf{g}$ is obtained, $\mathbf{R}$ and $\mathbf{t}$ can be calculated. $\mathbf{R}$ can be obtained from $g_{10}\sim g_{18}$ directly. On the other hand, $\mathbf{t}$ can be calculated using matrix $\mathbf{E}$ whose elements are $g_1\sim g_9$. $\mathbf{E}$ matrix is the product of $\mathbf{R}$ and $\mathbf{T}$ as follows.

$$\begin{aligned} \mathbf{E} &= \begin{pmatrix} g_1 & g_2 & g_3 \\ g_4 & g_5 & g_6 \\ g_7 & g_8 & g_9 \end{pmatrix} \\ &= \begin{pmatrix} r_{12}t_3-r_{13}t_2 & r_{13}t_1-r_{11}t_3 & r_{11}t_2-r_{12}t_1 \\ r_{22}t_3-r_{23}t_2 & r_{23}t_1-r_{21}t_3 & r_{21}t_2-r_{22}t_1 \\ r_{32}t_3-r_{33}t_2 & r_{33}t_1-r_{31}t_3 & r_{31}t_2-r_{32}t_1 \end{pmatrix} \\ &= \mathbf{R}\mathbf{T}, \end{aligned} \tag{11}$$

$$\mathbf{T} = \mathbf{R}^{-1}\mathbf{E}, \tag{12}$$

TABLE I
SIMULATION CONDITIONS

n_1	1.0 (air)
n_2	1.49 (acryl)
w	50 mm
$\mathbf{R}$	(-0.15π,-0.15π, 0.10π) rad (Euler angles)
$\mathbf{t}$	(600, -300, 50) mm
$\mathbf{n}$	(0.454, -0.405, 0.794)

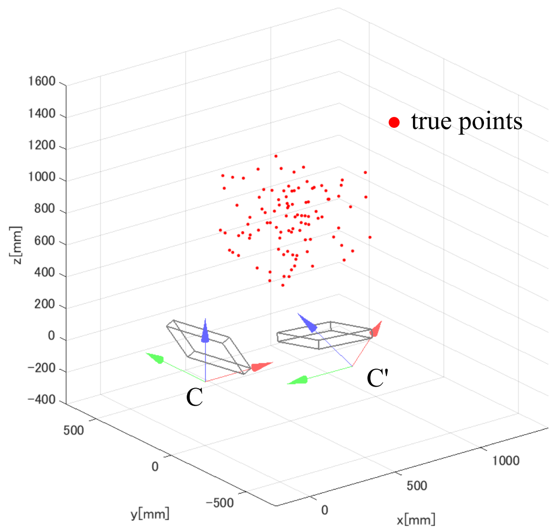


Fig. 5. Two camera positions are set. 100 measurement points are placed randomly in 3D area (red dots).

where

$$\mathbf{T} = \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix}, \tag{13}$$

$\mathbf{t}$ can be obtained from the elements of $\mathbf{T}$.

The 3D positions of measurement points are calculated by triangulation from the position and posture of cameras. However, even if we consider the opposite vector of $\mathbf{g}$, $-\mathbf{g}$, Eq. (8) is held. This means that the case that $\mathbf{t}$ is correct however the opposite matrix of $\mathbf{R}$ can also be solution of Eq. (8). The method of finding the correct $\mathbf{R}$ is explained below.

The measurement points should be in front of each camera. Therefore, the correct solution is the one where the z-coordinates of estimated positions of points are positive. If the z-coordinates of points are negative, the opposite of $\mathbf{g}$ should be the correct solution.

III. EXPERIMENTS

A. Simulation experiment

In order to confirm the effectiveness of the proposed method, simulations were conducted. The conditions of the simulation are shown in Table I. 100 measurement points were put randomly in a 3D area of $200 < x < 800$, $-300 < y < 300$, $600 < z < 1200$. The red dots are the true positions of points in Fig. 5. Then two cameras were placed at positions from where the all points could be captured. The vectors C and C′ express the two camera coordinates in Fig. 5. The images of points captured from each camera considering the refraction were simulated. The proposed method was applied to these images and the points were reconstructed. The result of the proposed method is shown in Fig. 6. The blue dots express the estimated positions of points and the estimated camera positions are expressed by long vectors. It appeared that the positions of points were calculated with absolute scale. The error, which was the average of Euclidean distance between true position and estimated position of points was very small (9.49×10^{-6} mm). Therefore, the proposed method can effectively calculate the absolute scale of object.

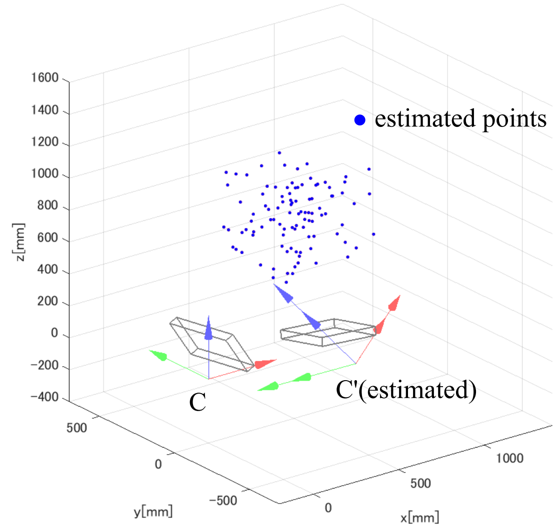


Fig. 6. The result of simulation experiment. Blue dots are the estimated positions of points. Long vectors show the estimated camera coordinate. Both are close to the true value. It means the proposed method is effective.

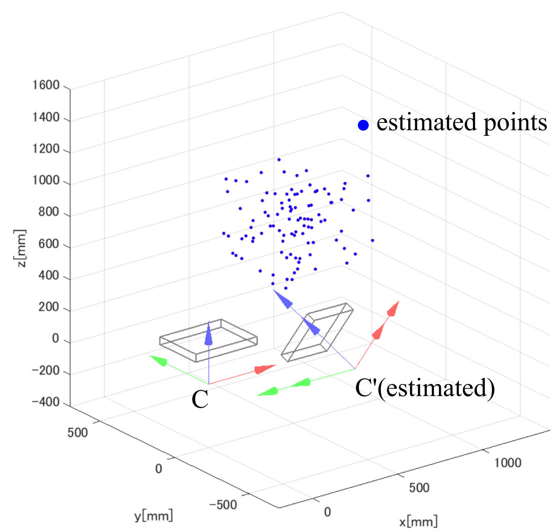
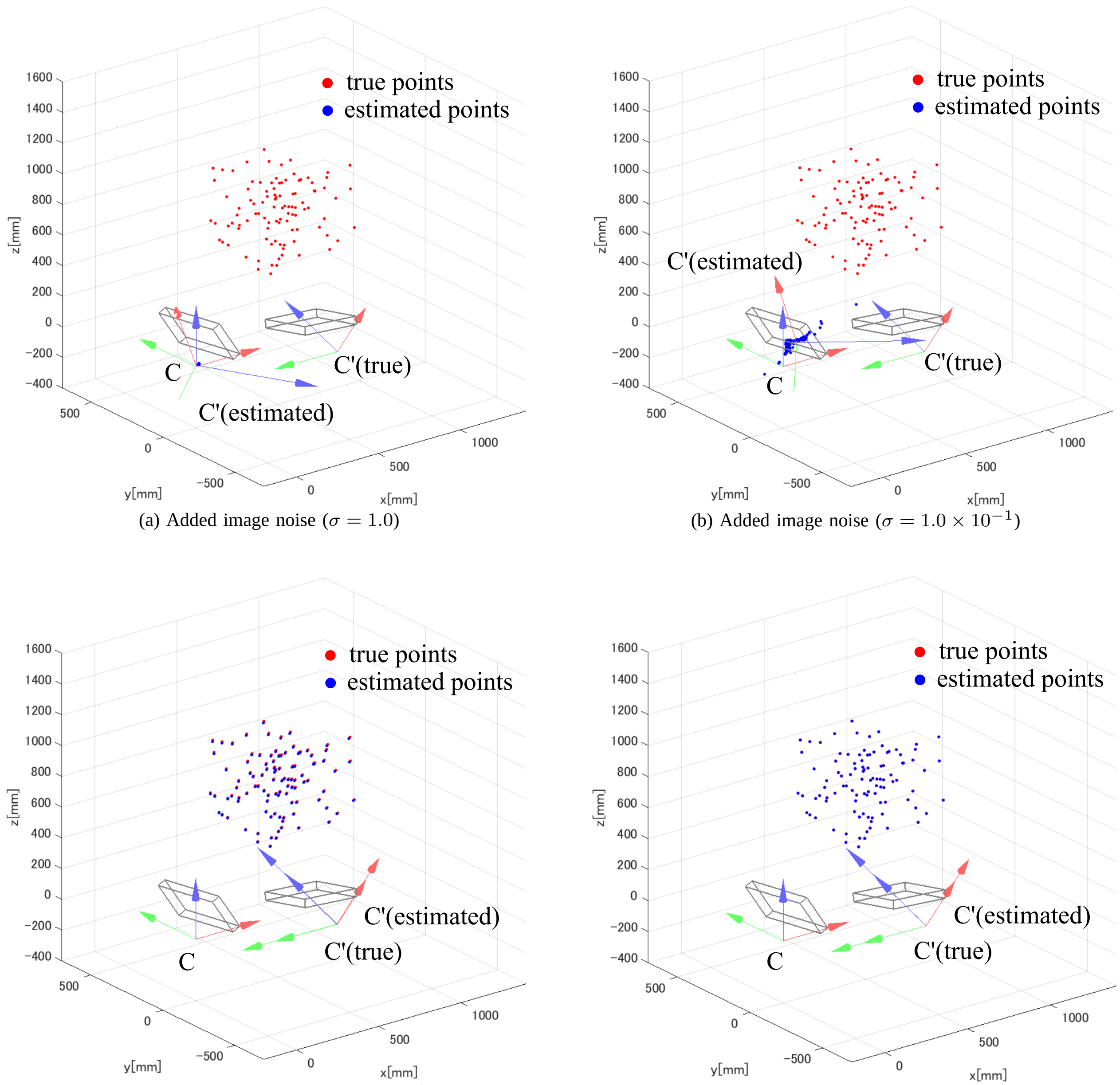


Fig. 7. The result of simulation when the refractive plate is placed perpendicular to the optical axis of the camera. The estimated positions of points and cameras are close to true values. Therefore the proposed method is effective in such particular case.

As explained in Section II, there is the particular case that the refractive plate is placed perpendicular to the optical axis of the camera in the proposed method. Therefore simulation experiment with the normal vector set to $n = (0, 0, 1)$ was



(a) Added image noise ($\sigma = 1.0$)

(b) Added image noise ($\sigma = 1.0 \times 10^{-1}$)

(c) Added image noise ($\sigma = 1.0 \times 10^{-2}$)

(d) Added image noise ($\sigma = 1.0 \times 10^{-3}$)

Fig. 8. Gaussian noise were added on images in simulation. σ was set as 1.0, 1.0×10^{-1}, 1.0×10^{-2} and 1.0×10^{-3}. The result shows that the noise is required to be less than $\sigma = 1.0 \times 10^{-2}$ in this condition. Therefore it is important to remove the image noise in the proposed method.

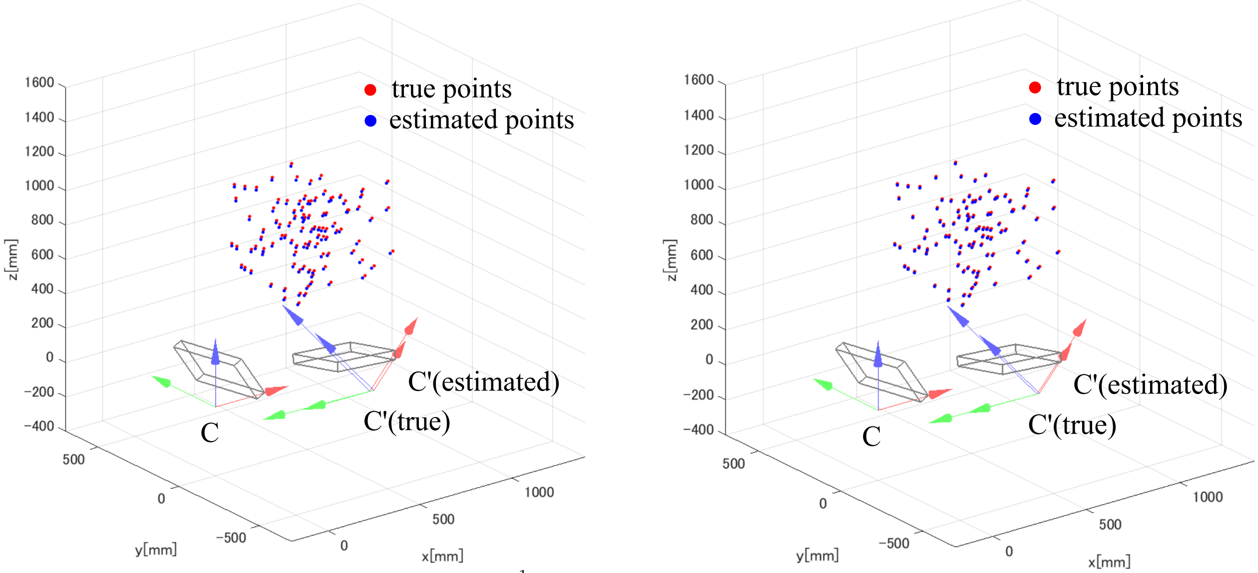
conducted. The result of simulation is shown in Fig. 7. The blue dots express the estimated position of points. The average of error was 4.28×10^{-7} mm. This result shows that the proposed method is effective even when the refractive plate is placed perpendicular to the optical axis of the camera.

B. Influence of image noise

The influence of noise on images to the 3D reconstruction in the proposed method was investigated. Gaussian noise was added on the images obtained by simulation and the precision of results of 3D reconstruction were evaluated. The standard deviation σ was set as 1.0, 1.0×10^{-1}, 1.0×10^{-2} and 1.0×10^{-3}, and the proposed method was applied to each case. Experimental conditions were the same as in Table I. The result is shown in Fig. 8. The red dots express the true position of points and the blue dots express the estimated position of points. When $\sigma = 1.0$ and $\sigma = 1.0 \times 10^{-1}$, the precision of reconstruction was not acceptable. Under these conditions, noise is required to be less than $\sigma = 1.0 \times 10^{-2}$. Therefore, it is important to remove noise on images in the proposed method.

C. Influence of the thickness of the refractive plate

It is assumed that the result of reconstruction is influenced by the amount of refraction. The parameter of refraction d depends on the thicknesses of refractive plate w (Eq. (3)). Therefore, the result of reconstruction was investigated with various thickness. The conditions of simulation were the



(a) $w = 100$ mm (added image noise σ=1.0 $\times$ 10^{-1}) (b) $w = 150$ mm (added image noise σ=1.0 $\times$ 10^{-1})

Fig. 9. The results of the proposed method when the thicker refractive plate is used in the simulation. Thickness was set at $w = 100$ mm and 150 mm. Gaussian noise was added to each image (σ=1.0 $\times$ 10^{-1}). The result of reconstruction was better when the refractive plate was thicker.

same as in Table I. Gaussian noise ($\sigma = 1.0 \times 10^{-1}$) was added on images in simulation. The thickness of plate w was set to 100 mm and 150 mm. The results of simulations are shown in Fig. 9. The red dots express the true position of points and the blue dots express the estimated position of points. It appears that the precision of reconstruction is improved when thicker refractive plate is used. Therefore, the precision of reconstruction of proposed method can be improved with thicker refractive plate even in the presence of noise.

IV. CONCLUSION

In this paper, we proposed a Structure from Motion method which can calculate the absolute scale of objects using refraction. The proposed method is effective even if the refractive plate is not placed perpendicular to the optical axis of the camera. Therefore, this is a generalized method of previous one. Simulation results show the effectiveness of the proposed method. It is also effective when the refractive plate is perpendicular to the optical axis of the camera, though it is important to choose the initial solution for optimization. It is clear that this method is sensitive to noise as much as the previous method. However, when a thicker refractive plate is used, the result of reconstruction becomes more precise.

Future work is to conduct experiments using real images. When real images are used, the quantization error should be occurred. Therefore to improve the robustness of the proposed method is required.

ACKNOWLEDGMENTS

This work was in part supported by Tough Robotics Challenge, ImPACT Program (Impulsing Paradigm Change through Disruptive Technologies Program).

REFERENCES

[1] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, Second Edition, 2004.

[2] D. Scaramuzza, F. Fraundorfer, M. Pollefeys and R. Siegwart, "Absolute Scale in Structure from Motion from a Single Vehicle Mounted Camera by Exploiting Nonholonomic Constraints", *Proceedings of the 2009 IEEE International Conference on Computer Vision*, 2009, pp.1413–1419.

[3] A. Shibata, H. Fujii, A. Yamashita and H. Asama, "Scale-Reconstructable Structure from Motion Using Refraction with a Single Camera ", *Proceedings of the 2015 IEEE International Conference on Robotics and Automation*, 2015, pp. 5239–5244.

[4] T. Treibitz, Y. Y. Schechner and H. Singh, "Flat Refractive Geometry", *Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[5] V. Chari and P. Sturm, "Multi–View Geometry of the Refractive Plane", *Proceedings of British Machine Vision Conference 2009*, 2009, pp. 1–11.

[6] Y. J. Chang and T. Chen, "Multi–View 3D Reconstruction for Scenes under the Refractive Plane with Known Vertical Direction", *Proceedings of the 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011, pp. 351–358.

[7] L. Kang, L. Wu, and Y. H. Yang, "Two-View Underwater Structure and Motion for Cameras under Flat Refractive Interfaces", *Proceedings of European Conference on Computer Vision 2012*, 2012, pp. 303–316.

[8] A. Jordt-Sedlazeck and R. Koch, "Refractive Structure–from–Motion on Underwater Images", *Proceedings of the 2013 IEEE International Conference on Computer Vision*, 2013, pp. 57–64.

[9] D. H. Lee, I. S. Kweon and R. Cipolla, "A Biprism–Stereo Camera System", *Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, 1999, pp. 82–87.

[10] C. Gao and N. Ahuja, "A Refractive Camera for Acquiring Stereo and Super–resolution Images", *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, pp. 2316–2323.

[11] A. Yamashita, Y. Shirane and T. Kaneko, "Monocular Underwater Stereo – 3D Measurement Using Difference of Appearance Depending on Optical Paths –", *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 3652–3657.

[12] Z. Chen, K. K. Wong, Y. Matsushita, X. Zhu and M. Liu, "Self–Calibrating Depth from Refraction", *Proceedings of the 2011 IEEE International Conference on Computer Vision*, 2011, pp. 635–642.