Scale-Reconstructable Structure from Motion Using Refraction with a Single Camera

Akira Shibata¹, Hiromitsu Fujii¹, Atsushi Yamashita¹, and Hajime Asama¹

Abstract—Three-dimensional (3D) measurement is an important means for robots to acquire information about their environment. Structure from Motion is one of these 3D measurement methods. The 3D reconstruction of objects in the environment can be obtained from pictures captured with single camera in Structure from Motion. Furthermore, the camera motion can be obtained simultaneously. Because of its flexibility, Structure from Motion has been implemented in various ways. However there is an essential problem in that the scale of the measured objects cannot be computed by Structure from Motion. In order to compute the absolute scale, other information is required. However this is difficult for robots in an unknown situation. In this paper, we propose a method that can reconstruct the absolute scale of objects using refraction. Refraction changes the light ray path between the objects and the camera. This method is implemented using only a refractive plate and single camera. The results of simulations show the effectiveness of the proposed method in both air and other media (e.g., water).

I. INTRODUCTION

A remote control robot is an important technology that is needed to explore areas too dangerous for people, such as nuclear plants or disaster sites. These robots need detailed information about their surrounding situation to conduct their activities. In order to acquire the position of surrounding objects, 3D measurement methods are required.

Structure from Motion is a 3D measurement method that is extensively studied in computer vision [1]. With this method, we can simultaneously estimate both the 3D structure of objects and the motion of the camera (rotational and translational parameters) with a single camera. Because of the flexibility of camera motion, Structure from Motion is suited to robot sensing. However, it is impossible to determine absolute translational parameters from the images alone. This is because when the scale of both scene and camera transfer change according to the same ratio, the camera will capture the same images that cannot be distinguished from the earlier ones. As a result, the absolute 3D structure of the objects in the scene cannot be estimated. In order to obtain absolute scale with Structure from Motion using a single camera, other information such as absolute scale in the scene or constraints on the camera motion is required [1]–[4]. However, it is difficult to add scale information when exploring an unknown environment by remote control robot. Furthermore constraints on the camera motion decrease the flexibility of Structure from Motion. Therefore, another approach to reconstruct the scale is required.

In this paper, we propose an absolute Structure from Motion method using refraction. Refraction changes the light ray path, which causes images to differ, even if the scale of both scene and camera transfer change at the same ratio. This difference enables us to compute the absolute translational parameter and reconstruct 3D objects. The proposed method is implemented with a simple system composed of a refractive plate and a single camera.

II. RELATED WORK

Refraction occurs when a light ray passes through the boundary surface between different media (e.g., air and water). It is well known that the direction of the light ray changes in accordance with Snell’s law [5]–[13]. Refractive geometry in computer vision is a well studied topic, especially for underwater environments [5]–[8]. However these studies focused on removing the effect of refraction as distortion. On the other hand, the proposed method uses refraction actively to reconstruct the scale of objects.

Refraction is used for 3D measurement in several studies. The depth map is obtained from the difference of two images with and without the refractive plate between the camera and the scene [9]. In another study, the depth map was obtained from images captured through a refractive plate rotating around the optical axis of the camera [10]. Furthermore, there are studies on 3D measurement from single images with refractive objects such as a biprism [11] or special waterproof case [12]. These studies measure 3D scenes using active refraction. However, the measurement range is limited. Therefore, these methods are not suited to the large-scale measurement that is required by an exploring robot. In contrast, the proposed method can measure a wider range than the other methods using refraction.

¹A. 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 113-8656, Japan shibata@robot.t.u-tokyo.ac.jp
III. MEASUREMENT METHOD

A. Refractive Plate System

The system of this method is shown in Fig. 1. We use a general single camera that is calibrated and place a refractive plate between the camera and the scene objects. The refractive plate should be placed such that its surface is orthogonal to the optical axis of the camera. The distance between the camera and refractive plate l as well as the thickness of the refractive plate w are known. Let the medium between the camera and refractive plate be the internal medium, and the medium between refractive plate and the measurement object be the external medium. The refractive indices of these media are known.

B. Light Ray Path Tracing

In this section, the path the light traces from image points to 3D points considering refraction is explained. The light ray refracts at both faces of the refractive plate. Therefore, three different ray vectors exist between the camera center and the object. Let the unit vectors of the ray from the camera to the measurement object be \( \mathbf{r}_{\text{in}} = (\alpha_1, \beta_1, \gamma_1)^T \), \( \mathbf{r}_{\mid} = (\alpha_2, \beta_2, \gamma_2)^T \) and \( \mathbf{r}_{\text{out}} = (\alpha_3, \beta_3, \gamma_3)^T \), respectively. These vectors and the normal vector of the refractive plate are on the same plane (Fig. 2). On this plane, \( \theta_1 \) and \( \theta_2 \) are the angle of the incident light and refractive light from the internal medium to the refractive plate, respectively. In addition, \( \theta_3 \) is the angle of refractive light from the refractive plate to the external medium. The relation between these angles are explained by Snell’s law as follows.

\[
\begin{align*}
    n_1 \sin \theta_1 &= n_2 \sin \theta_2, \quad (1) \\
    n_2 \sin \theta_2 &= n_3 \sin \theta_3, \quad (2)
\end{align*}
\]

where \( n_1, n_2 \) and \( n_3 \) are the refractive indices of the internal medium, refractive plate and external medium, respectively.

Using these relations, \( \mathbf{r}_{\text{in}}, \mathbf{r}_{\mid} \) and \( \mathbf{r}_{\text{out}} \) can be computed. First, \( \mathbf{r}_{\text{in}} \) can be computed from the coordinates of the image point. When the coordinates of the image point are \((u, v)\) and the focal length of camera is \( f \), \( \mathbf{r}_{\text{in}} \) is expressed as normalized \((u, v, f)\) (Fig. 3). Furthermore, we define the \( \psi \)-axis whose direction is a projection of \( \mathbf{r}_{\text{in}} \) on the \( x-y \) plane. The refraction occurs on the \( \psi-z \) plane.

Next, \( \mathbf{r}_{\mid} \) can be computed. Because the three vectors \( \mathbf{r}_{\text{in}}, \mathbf{r}_{\mid} \) and the normal vector of the refractive plate \( \mathbf{N} = (0, 0, 1) \) are on the same plane, \( \mathbf{r}_{\mid} \) is expressed as follows.

\[
\mathbf{r}_{\mid} = \begin{pmatrix}
    \alpha_2 \\
    \beta_2 \\
    \gamma_2
\end{pmatrix} = p_1 \begin{pmatrix}
    \alpha_1 \\
    \beta_1 \\
    \gamma_1
\end{pmatrix} + q_1 \begin{pmatrix}
    0 \\
    0 \\
    1
\end{pmatrix},
\]

where \( p_1 \) and \( q_1 \) are constant.

Next, \( \cos \theta_1 \) is calculated from the inner product of \( \mathbf{r}_{\text{in}} \) and \( \mathbf{N} \).

\[
\mathbf{r}_{\text{in}} \cdot \mathbf{N} = |\mathbf{r}_{\text{in}}||\mathbf{N}| \cos \theta_1,
\]

\[
\therefore \cos \theta_1 = \gamma_1,
\]

and \( \sin \theta_1 \) is calculated from the cross product of \( \mathbf{r}_{\text{in}} \) and \( \mathbf{N} \).

\[
|\mathbf{r}_{\text{in}} \times \mathbf{N}| = |\mathbf{r}_{\text{in}}||\mathbf{N}| \sin \theta_1,
\]

\[
\therefore \sin \theta_1 = \sqrt{(\alpha_1)^2 + (\beta_1)^2}.
\]

Furthermore, \( \cos \theta_2 \) is calculated from the inner product of \( \mathbf{r}_{\mid} \) and \( \mathbf{N} \).

\[
\mathbf{r}_{\mid} \cdot \mathbf{N} = |\mathbf{r}_{\mid}||\mathbf{N}| \cos \theta_2,
\]

\[
\therefore \cos \theta_2 = \gamma_2 = p_1 \gamma_1 + q_1,
\]

and \( \sin \theta_2 \) is calculated from the cross product of \( \mathbf{r}_{\mid} \) and \( \mathbf{N} \).

\[
|\mathbf{r}_{\mid} \times \mathbf{N}| = |\mathbf{r}_{\mid}||\mathbf{N}| \sin \theta_2,
\]

\[
\therefore \sin \theta_2 = \sqrt{(\alpha_2)^2 + (\beta_2)^2} = \sqrt{(p_1 \alpha_1)^2 + (p_1 \beta_1)^2}.
\]

From Eqs. (1), (7) and (11), \( p_1 \) is obtained.

\[
p_1 = \frac{\sin \theta_2}{\sin \theta_1} = \frac{n_1}{n_2}.
\]

From Eqs. (5), (9) and (12), \( \cos \theta_2 \) is expressed.

\[
\cos \theta_2 = p_1 \gamma_1 + q_1 = \frac{n_1}{n_2} \cos \theta_1 + q_1.
\]

We note that \( \cos \theta_2 \) can also be expressed as follows according to its trigonometric identity.

\[
\cos \theta_2 = \sqrt{1 - \sin^2 \theta_2} = \sqrt{1 - \left(\frac{n_1}{n_2}\right)^2 \sin^2 \theta_1}.
\]
From Eqs. (13) and (14), \( q_1 \) is obtained.

\[
q_1 = - \frac{n_1}{n_2} \cos \theta_1 + \sqrt{1 - \left( \frac{n_1}{n_2} \right)^2 \sin^2 \theta_1}. \tag{15}
\]

Therefore, from Eqs. (3), (12) and (15), we can calculate \( \mathbf{r}_{\text{mid}} \).

\[
\mathbf{r}_{\text{mid}} = \frac{n_1}{n_2} \mathbf{r}_m - \left\{ \frac{n_1}{n_2} \cos \theta_1 - \sqrt{1 - \left( \frac{n_1}{n_2} \right)^2 \sin^2 \theta_1} \right\} \mathbf{N}. \tag{16}
\]

Vector \( \mathbf{r}_{\text{out}} \) is computed by the same procedure.

\[
\mathbf{r}_{\text{out}} = \frac{n_2}{n_3} \mathbf{r}_{\text{mid}} - \left\{ \frac{n_2}{n_3} \cos \theta_2 - \sqrt{1 - \left( \frac{n_2}{n_3} \right)^2 \sin^2 \theta_2} \right\} \mathbf{N}. \tag{17}
\]

Here, let the intersection point of \( \mathbf{r}_{\text{out}} \) and the optical axis of camera be start point \( D \), whose position vector is expressed as follows.

\[
\mathbf{d} = (0, 0, d), \tag{18}
\]

where \( d \) is distance between point \( D \) and the camera center, which is calculated geometrically (Fig. 2).

\[
d = l + w - \frac{l \tan \theta_1 + w \tan \theta_2}{\tan \theta_3}. \tag{19}
\]

**C. Geometrical Relation**

An object to be measured is observed from two different viewpoints in Structure from Motion. These two camera coordinates are \( C \) and \( C' \) respectively. The z-axis of each coordinate corresponds to each optical axis of the camera and the world coordinates correspond to the camera coordinates \( C \). The rotation matrix \( \mathbf{R} \) is a transformation matrix from camera coordinate \( C \) to \( C' \), and transfer vector \( \mathbf{t} \) is defined as the position vector of the center of the camera coordinate \( C' \). Vector \( \mathbf{r}_{\text{out}} \) and \( \mathbf{d} \) of each coordinate point can be expressed as follows.

\[
\mathbf{r}_{\text{out}} = (x, y, z)^T, \tag{20}
\]

\[
\mathbf{d} = (0, 0, d)^T, \tag{21}
\]

\[
\mathbf{r}'_{\text{out}} = (x', y', z')^T, \tag{22}
\]

\[
\mathbf{d}' = (0, 0, d')^T. \tag{23}
\]

Each of the \( \mathbf{r}_{\text{out}} \) vectors of the two coordinates and the vector between each of the start points \( D \) of both coordinates are on the same plane (Fig. 4). These three vectors are expressed in world coordinates as \( \mathbf{r}_{\text{out}}, \mathbf{R}^{-1} \mathbf{r}_{\text{out}} \) and \( \mathbf{t} + \mathbf{R}^{-1} \mathbf{r}'_{\text{out}} - \mathbf{d} \). Therefore, we obtain the equation for this geometrical relation as follows.

\[
\{ (\mathbf{t} + \mathbf{R}^{-1} \mathbf{d}' - \mathbf{d}) \times \mathbf{R}^{-1} \mathbf{r}'_{\text{out}} \}^T \mathbf{r}_{\text{out}} = 0. \tag{24}
\]

If there is no refractive plate, Eq. (24) holds when vector \( \mathbf{t} \) is multiplied. This is the reason why the scale of vector \( \mathbf{t} \) cannot be acquired by conventional method.

Equation (24) can be converted into a formula consisting of an inner product of the vector \( \mathbf{u} \), composed of known quantities and vector \( \mathbf{g} \), composed of unknown quantities by the orthonormality of rotation matrix.

\[
\begin{pmatrix}
x' \\
y' \\
z' \\
-xx' + d'xy' \\
-yy' + d'yy' \\
zz' + d'zz'
\end{pmatrix}^T
\begin{pmatrix}
x_{12}t_3 - x_{13}t_2 \\
x_{12}t_3 - x_{13}t_2 \\
x_{11}t_2 - x_{12}t_1 \\
x_{11}t_2 - x_{12}t_1 \\
x_{11}t_3 - x_{12}t_1 \\
x_{11}t_3 - x_{12}t_1 \\
x_{12}t_3 - x_{13}t_2 \\
x_{12}t_3 - x_{13}t_2 \\
x_{11}t_2 - x_{13}t_1 \\
x_{11}t_2 - x_{13}t_1 \\
x_{12}t_3 - x_{13}t_2 \\
x_{12}t_3 - x_{13}t_2 \\
x_{13}t_2 - x_{12}t_1 \\
x_{13}t_2 - x_{12}t_1 \\
x_{13}t_3 - x_{12}t_1 \\
x_{13}t_3 - x_{12}t_1
\end{pmatrix}
= 0, \tag{25}
\]

where \( r_{ij} \) is the \( i, j \)-th element of rotational matrix \( \mathbf{R} \), and \( t_i \) is the \( i \)-th element of transfer vector \( \mathbf{t} \).

Because Eq. (26) is obtained from every measurement point, a homogeneous equation is obtained from \( n \) points.

\[
\mathbf{Ug} = 0, \tag{27}
\]

where

\[
\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_n)^T. \tag{28}
\]

In order to compute the unknown vector \( \mathbf{g} \), a least-squares method is applied to Eq. (27). Let \( g_i \) be the \( i \)-th element of \( \mathbf{g} \). Then \( g_{10} \sim g_{12} \) and \( g_{13} \sim g_{15} \) are the same as the first and second columns of \( \mathbf{R} \) respectively. This leads to two constraints because of the orthonormality of the rotation matrix.

\[
g_{10}^2 + g_{11}^2 + g_{12}^2 = 1, \tag{29}
\]

\[
g_{13}^2 + g_{14}^2 + g_{15}^2 = 1. \tag{30}
\]
Therefore, Eq. (19) is modified as follows.

\[
\begin{align*}
\mathbf{g}^{\text{external and internal media are air,}} & = \mathbf{A} \mathbf{g}^{\text{external media are not specified. It can be applied as long as}} \\
& \quad \text{the refractive indices of those media are known. For example, this method can be used for underwater measurement.}
\end{align*}
\]

However, in this section, we focus on measurement in air because many 3D measurements occur in air in practice.

\[E. \text{ Application to Measurement in Air}\]

The proposed method is general because internal and external media are not specified. It can be applied as long as the refractive indices of those media are known. For example, this method can be used for underwater measurement. However, in this section, we focus on measurement in air because many 3D measurements occur in air in practice.

The refractive index of air is approximately 1.00. When both external and internal media are air, \(n_1 = n_3 = 1.00\) and \(\theta_1 = \theta_3\) modified from Eqs. (1) and (2), respectively, hold. Therefore, Eq. (19) is modified as follows.

\[
d = w - \frac{w \tan \theta_2}{\tan \theta_1},
\]  

\[\text{(35)}\]

\[
\text{TABLE I} \quad \text{SIMULATION CONDITIONS}
\]

<table>
<thead>
<tr>
<th>Situation</th>
<th>Air</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n_1)</td>
<td>1.0 (air)</td>
</tr>
<tr>
<td>(n_2)</td>
<td>1.49 (acryl)</td>
</tr>
<tr>
<td>(n_3)</td>
<td>1.0 (air)</td>
</tr>
<tr>
<td>(l)</td>
<td>200 [mm]</td>
</tr>
<tr>
<td>(w)</td>
<td>50 [mm]</td>
</tr>
<tr>
<td>(R)</td>
<td>(-0.15\pi,-0.15\pi, 0.1\pi) [rad] (Euler angles)</td>
</tr>
<tr>
<td>(t)</td>
<td>(600, -300, 50) [mm]</td>
</tr>
</tbody>
</table>

\[\text{Fig. 5. In the simulation, camera positions are given. Blue vectors indicates the optical axis of each camera. One hundred points (red circles) are placed randomly in 3D space such that each camera can capture all of them.}\]

\[\text{Fig. 6. Images captured by both cameras in the simulation.}\]

In Eq. (35), distance \(d\) has no relation to the length \(l\) between the refractive plate and camera center. Because both internal and external media are air, \(r_{in}\) and \(r_{out}\) are parallel. Therefore, the distance between these vectors does not change with respect to \(l\). Other calculation procedures are the same as the above method.

IV. SIMULATION EXPERIMENTS

A. Simulation in Air

In order to validate the effectiveness of the proposed method, a simulation experiment was conducted. Simulation conditions are shown in Table I. We placed 100 measurement points randomly in a 3D space where both cameras could capture all of them. The blue vectors indicate the optical axis of each camera in Fig. 5. First, images of all the points for both cameras as captured through the refractive plates were simulated (Fig. 6). From those image coordinates, the 3D coordinates of the points were computed using the proposed method and compared with the true position of the points.
The proposed method makes use of refraction to reconstruct the absolute scale of the objects. Therefore, it is assumed that the precision of the reconstruction would be improved by increasing the effect of refraction. In this system, the refraction is generated by the refractive plate, and its effect changes according to the thickness of the plate.

B. Simulation in Water

In order to ensure the effectiveness of proposed method in a general situation, a simulation experiment in water was also conducted. The conditions were the same as those of the air simulation (Table I), excluding \( n_S \). The refractive index of water was set to \( n_S = 1.33 \) in this simulation. The result is shown in Fig. 8. The average of each Euclidean distance between the true and reconstructed positions was extremely small (8.6 \( \times 10^{-6} \) mm). Therefore, the proposed method can reconstruct the scale of a scene object.

C. Influence of Refractive Plate Thickness

The proposed method makes use of refraction to reconstruct the absolute scale of the objects. Therefore, it is assumed that the precision of the reconstruction would be improved by increasing the effect of refraction. In this system, the refraction is generated by the refractive plate, and its effect changes according to the thickness of the plate.

D. Quantization Error

The proposed method uses the effect of refraction. However, that effect should be extremely small. Therefore, this method could be sensitive to error. In this simulation, there is no error in the image coordinates. In practice, however, quantization error will be generated because of the size of the pixel. We usually determine the coordinates to the integer precision of a pixel. In this section, we generate quantization error artificially and inspect the effect of this error to the result of the 3D reconstruction.

During image capture in the simulation, the coordinates of the image points were rounded to zero, one, two, and three decimal places. The proposed method was applied to these coordinates. The medium of this simulation was air. The results are shown in Figs. 10 (a)–(d), respectively. It is clear that the coordinates rounded to zero and one decimal places are not precise enough to reconstruct the real positions. We need precision to at least the two decimal place. Therefore, the proposed method requires sub-pixel accuracy for detection of the corresponding points.

Fig. 7. Simulation results of the proposed method (in air). The distances between the real position (red dots) and estimated position (blue triangles) are extremely small. This shows that the proposed method can reconstruct scene objects with absolute scale.

Fig. 8. Simulation results of the proposed method (in water). This shows that the proposed method is effective in water.

The result is shown in Fig. 7. The red circles indicate the real positions and the blue triangles indicate their estimated positions. It is clear that the points were reconstructed correctly. The average of each Euclidean distance between the true and reconstructed positions was extremely small (8.6 \( \times 10^{-6} \) mm). Therefore, the proposed method can reconstruct the scale of a scene object.

B. Simulation in Water

In order to ensure the effectiveness of proposed method in a general situation, a simulation experiment in water was also conducted. The conditions were the same as those of the air simulation (Table I), excluding \( n_S \). The refractive index of water was set to \( n_S = 1.33 \) in this simulation. The result is shown in Fig. 8. The average of each Euclidean distance between the true and reconstructed positions was extremely small (1.1 \( \times 10^{-7} \) mm). Therefore, the proposed method can reconstruct the scale of scene objects in water as well as in air. This indicates that if the internal and external media are different, the proposed method is still effective.

C. Influence of Refractive Plate Thickness

The proposed method makes use of refraction to reconstruct the absolute scale of the objects. Therefore, it is assumed that the precision of the reconstruction would be improved by increasing the effect of refraction. In this system, the refraction is generated by the refractive plate, and its effect changes according to the thickness of the plate.

In this section, the thickness of the plate was changed in the simulation to investigate the influence of thickness.

The thickness of refractive plate \( w \) was changed by 10 mm (from 10 to 100 mm), and the error of the reconstruction results were compared. In this experiment, the distance between the camera center and refractive plate \( l \) as well as the positions of the data points were the same as for previous experiments (Table I). The medium of this simulation was air. The result is shown in Fig. 9. The horizontal axis is the thickness of the plate \( w \), and the vertical axis is the amount of error between the true and estimated positions. This indicates that the precision of the measurement is improved when the plate is thicker. Therefore, in the proposed method, the grater the amount of refraction, the more precise the refraction result.

D. Quantization Error

The proposed method uses the effect of refraction. However, that effect should be extremely small. Therefore, this method could be sensitive to error. In this simulation, there is no error in the image coordinates. In practice, however, quantization error will be generated because of the size of the pixel. We usually determine the coordinates to the integer precision of a pixel. In this section, we generate quantization error artificially and inspect the effect of this error to the result of the 3D reconstruction.

During image capture in the simulation, the coordinates of the image points were rounded to zero, one, two and three decimal places. The proposed method was applied to these coordinates. The medium of this simulation was air. The results are shown in Figs. 10 (a)–(d), respectively. It is clear that the coordinates rounded to zero and one decimal places are not precise enough to reconstruct the real positions. We need precision to at least the two decimal place. Therefore, the proposed method requires sub-pixel accuracy for detection of the corresponding points.

V. CONCLUSIONS

In this paper, we proposed a scale-reconstructable Structure from Motion method with a single camera by using refraction and verified its effectiveness in simulation. The
The proposed method is simple because it requires only a refractive plate that causes light rays to change their direction and is attached in front of a single camera. Furthermore, it is clear that this method is applicable to various media such as water as well as air. This indicates that this method could be used for underwater exploration. The experiments also show that the precision of the reconstruction is improved by using a thicker refractive plate because the effect of refraction increases. However, it is also clear that this method is sensitive to error. Even the quantization error has a bad influence on the results of reconstruction. This means that detecting corresponding positions with sub-pixel accuracy is necessary. Therefore, the robustness of this method should be improved.

Experimental results using actual measurements have not been acquired yet. This task will be future work. Furthermore, it would be interesting to examine whether this method would be more accurate when using a special-shaped refractive object instead of a refractive plate.

ACKNOWLEDGMENT

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

REFERENCES