# Global Optimization with Viewpoint Selection for Scale-reconstructible Structure from Motion Using Refraction

Yukari Okumura, Hiromitsu Fujii, Atsushi Yamashita, and Hajime Asama
Department of Precision Engineering
The University of Tokyo
Tokyo, Japan
Email: {okumura, fujii, yamashita, asama}@robot.t.u-tokyo.ac.jp

***Abstract*—Structure from motion (SfM) is a three-dimensional (3D) measurement method using a single moving camera. This method can simultaneously estimate the 3D positions of objects and camera poses. However, conventional SfM methods cannot determine the real-world scales of objects. To solve this problem, a scale-reconstructible SfM method using refraction was proposed. In the method, the shapes of objects can be reconstructed with their real-world scales by using just two images captured through a refractive plate. However, the method is considerably influenced by quantization errors which occur in spatial sampling when images are digitized, and reconstruction accuracy is insufficient. In this paper, we propose a novel method to apply a global optimization technique using bundle adjustment for scale-reconstructible SfM using refraction. The proposed method improves the accuracy of 3D reconstruction by selecting viewpoints of the single moving camera with consideration for a geometrical consistency in the measurement.**

***Keywords*—Structure from motion; Refraction; Bundle adjustment**

## I. Introduction

Structure from motion (SfM) is a three-dimensional (3D) measurement method using a single moving camera. The method can simultaneously estimate the 3D positions of objects and camera poses, namely, the rotations and the directions of the translations. However, conventional SfM methods cannot calculate the real-world scales of the translations of the camera. Thus, the true sizes of objects cannot be determined.

To reconstruct 3D objects with their real-world scales, camera positions or geometric information are often required in the conventional methods [1]–[4]. However, it is impossible to apply these methods when camera positions or geometric information are completely unknown. To solve this problem, Shibata *et al.* [5], [6] proposed a scale-reconstructible SfM using refraction. In the methodology, refraction is produced by setting a refractive medium such as an acrylic plate between the camera and the object. It is verified that the shapes of objects can be reconstructed with their real-world scales by using just two images taken through the refractive plate. However, the method is considerably influenced by quantization errors which occur in spatial sampling when images are digitized, and hence requires sub-pixel accuracy of correspondence detection. As a result, reconstruction accuracy is insufficient using a camera with pixel resolution.

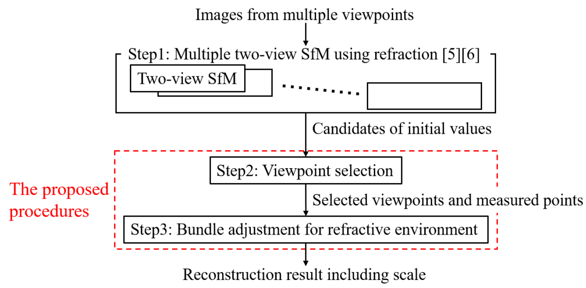



Fig. 1. The schematic of our reconstruction method. Two-view SfM is conducted multiple times to obtain the candidates of the initial values. Then the viewpoints are selected.

Global optimization using bundle adjustment is widely applied to SfM to minimize the influence of measurement errors for accurate reconstruction [7]. Bundle adjustment is an optimization technique using multiple images from the multiple viewpoints by considering a global consistency of the measurement results. This kind of technique is widely used in SfM [8]–[11]. However, there are few studies using bundle adjustment with consideration for the influence of refraction. As an example, Jordt-Sedlazeck *et al.* [12] proposed this kind of reconstruction method. In the study, 3D reconstruction was succeeded in the water by using refraction produced by a refractive plate, and bundle adjustment was used to improve the accuracy of reconstruction. However, the method needs the initial values for the optimization using bundle adjustment in advance.

In this paper, to improve the accuracy of 3D reconstruction, a method of global optimization using bundle adjustment for SfM with refraction is proposed. We propose two approaches. First, a viewpoint selection method is proposed to obtain effective initial values for bundle adjustment. Second, an evaluation function of reprojection is introduced. This evaluation function can treat the effect of the refraction. Both of our approaches use a geometrical consistency of the measurement. In simulation experiments, the effectiveness and stability of our proposed method are verified.

## II. Proposed method

### A. Approach for improving reconstruction accuracy

The schematic of our reconstruction method is shown in Fig. 1. At the step 1, the two-view SfM using refraction is

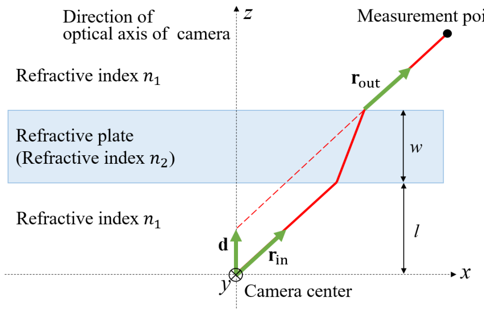



Fig. 2. Relationship between positions of the camera and the refractive plate. The refractive plate is placed vertical to the optical axis of the camera. The red solid line shows the optical ray which is influenced by refraction.

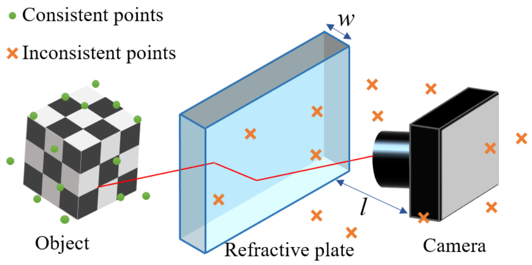



Fig. 3. Geometrical consistency of estimated points. Green points are in the object side of the refractive plate, and orange points are in the camera side or inside of the refractive plate.

conducted multiple times by using the images from multiple viewpoints of the single moving camera. The two-view SfM is based on the previous works [5], [6]. The output of the step 1 is the parameters of the viewpoints and the measured points from each viewpoint. At the step 2, the viewpoints are selected. The parameters of the viewpoints are the rotations and the translations of cameras and the positions of the measured points, and these are optimized at the step 3. The initial values of the parameters greatly affect the optimization result. This is because many unknown parameters have to be adjusted, and an evaluation function is highly non-linear. Thus, initial values for bundle adjustment have to be configured appropriately. We propose a method to select viewpoints of the cameras by verifying images based on geometric constraints (Section II – $C$). At the step 3, bundle adjustment with the evaluation function of the reprojection considering the refraction is conducted. We propose a new evaluation function considering the geometric constraints (Section II – $D$). The viewpoint selection and the design of the evaluation function for the optimization are our main contributions.

## *B. Two-view SfM using refraction for obtaining candidates of initial values*

The schematic of the relationship between positions of a camera and a refractive plate is shown in Fig. 2. The details of following settings are according to the previous works [5], [6]. The camera coordinate is defined as a right-hand coordinate system whose $z$-axis is the direction of the optical axis of the camera and whose origin corresponds to the camera center. The refractive plate is vertical to the optical axis of the camera. In this measurement system, the optical ray is influenced by refraction and traces the red solid line as shown in Fig. 2. Let $\mathbf{r}_{\text{in}}$ denote the inner ray vector which is between the camera and the refractive plate, and let $\mathbf{r}_{\text{out}}$ denote the outer vector which is between the refractive plate and the object. The variable $\mathbf{d}$ represents the variation vector of the ray which is caused by refraction and whose direction is the normal direction of the refractive plate.

The scale-reconstructible SfM can be conducted by this measurement system. The object is measured from multiple viewpoints of a moving camera, and the measurement results are used as candidates of the initial values for the following optimization steps. The candidates are obtained as follows. At the beginning, each of $n$ input images is taken from different viewpoint, and each camera coordinate is set to each viewpoint. One of these coordinates can be defined as the world coordinate without loss of generality. Next, two-view SfM using refraction is conducted multiple times between the image of the standard coordinate and the other $(n-1)$ images, respectively. In this step, SfM is conducted $(n-1)$ times. Finally, the results of the $(n-1)$ two-view SfM are used as the candidates of the initial values.

## *C. Viewpoint selection*

The parameters to be optimized in bundle adjustment are the rotations and translations of cameras and the positions of the measured points. The candidates of the initial values of the parameters are calculated by two-view SfM using refraction (Section II – $B$).

In two-view SfM, however, the geometrically inconsistent points are often obtained. This consistency is that reconstruction points are obviously in the object side of the refractive plate, and that the estimated position of the object should not be in the camera side or inside of the refractive plate (Fig. 3). This result is influenced by the quantization errors. It is verified in the experiments that when inconsistent points are involved in the initial values, correct solutions cannot be obtained even if bundle adjustment is conducted. This is why the proposed method verifies the geometrical consistency. If they involve the estimated points which are in the camera side or inside of the refractive plate, the viewpoints and the reconstruction results from these viewpoints are removed from initial values.

Concretely, viewpoint selection is conducted as follows. Let the two camera coordinates be $C$ and $C'$, the values of the depth $z$ of each point be ${}^{C}z$ and ${}^{C'}z$, respectively. As shown in Fig. 2, $w$ denotes the thickness of the refractive plate and $l$ denotes the distance between the camera center and the refractive plate. That is, all the reconstructed points should satisfy the following conditions of Eq. (1).

$$ {}^{C}z > w + l \text{ and } {}^{C'}z > w + l. \tag{1} $$

If all points satisfy Eq. (1), the viewpoints are adopted, and the estimated positions of the points are used to calculate the initial values of measurement positions.

The following explains the process for decision of points using as initial values. First, all candidates of the initial values obtained by two-view SfM are verified whether the estimated points from these viewpoints satisfy Eq. (1), or not. The viewpoints whose all of the estimated points satisfy Eq. (1)

are selected. The initial value of each measurement point is calculated as the average of all the estimated positions from the selected viewpoints. The parameters of the selected viewpoints and these average positions of measurement points are used as the initial values of bundle adjustment.

### D. Evaluation function for bundle adjustment using refraction

In this method, to obtain the appropriate parameters by bundle adjustment considering the refraction, the evaluation function that includes two kinds of the reconstruction errors of the optimization is introduced. Let the set of the optimized camera coordinates denote $C^*$, and let the set of the optimized measurement points denote $P^*$. The sets are optimized as follows:

$$\{C^*, P^*\} = \underset{\{C,P\}}{\arg\min}(e_1 + \alpha e_2), \tag{2}$$

where $e_1$ denotes the ray-gap function, and $e_2$ denotes the penalty function. The parameter $\alpha$ is a coefficient.

The reason for introducing the ray-gap function $e_1$ is the following. In general bundle adjustment without considering the influence of the refraction, the evaluation function is the distance between the point in the input image and the reprojected point of the estimated point. However, in the refractive environment, it is impossible to evaluate correctly the positions of the camera coordinates in the geometrically inconsistent states as mentioned in the Section II$-C$. For example, when the $z$ coordinate of the estimated $\mathbf{r}_{\mathrm{out}}$ is negative, the estimated points in this area are reprojected on the image from the back of the camera, and the points are incorrectly regarded as being estimated at the actual positions.

Concretely, $e_1$ is as follows:

$$e_1 = \sum_{c\in\mathcal{C}}\sum_{p\in\mathcal{P}} \left\| {}^{c}\bar{\mathbf{r}}^{p}_{\mathrm{in}} - {}^{c}\bar{\hat{\mathbf{r}}}^{p}_{\mathrm{out}} \right\|^2, \tag{3}$$

where $\mathcal{C}$ denotes the set of the $(n-1)$ camera coordinates, $\mathcal{P}$ denotes the set of the estimated measurement points. Let $\hat{\mathbf{r}}_{\mathrm{out}}$ be defined the estimated outer vector, and the normal vectors of $\mathbf{r}_{\mathrm{in}}$ and $\hat{\mathbf{r}}_{\mathrm{out}}$ be $\bar{\mathbf{r}}_{\mathrm{in}}$ and $\bar{\hat{\mathbf{r}}}_{\mathrm{out}}$, respectively. The upper indices of ${}^{c}\bar{\mathbf{r}}^{p}_{\mathrm{in}}$ and ${}^{c}\bar{\hat{\mathbf{r}}}^{p}_{\mathrm{out}}$ indicate that they denote the values about the point $p$ seen from the camera coordinate $c$.

In Eq. (3), it is evaluated whether the two vectors coincide or not. The two vectors are the inner ray vector $\mathbf{r}_{\mathrm{in}}$ and the outer ray vector $\mathbf{r}_{\mathrm{out}}$. Equation (3) compares each component of the normal inner ray vector ${}^{c}\bar{\mathbf{r}}^{p}_{\mathrm{in}}$ and the estimated normal outer ray vector ${}^{c}\bar{\hat{\mathbf{r}}}^{p}_{\mathrm{out}}$ in order that the direction of ${}^{c}\bar{\hat{\mathbf{r}}}^{p}_{\mathrm{out}}$ coincides with that of ${}^{c}\bar{\mathbf{r}}^{p}_{\mathrm{in}}$.

The inner ray vector $\mathbf{r}_{\mathrm{in}}$ can be calculated from the image coordinate of the corresponding points. The estimated outer ray vector can be calculated as ${}^{c}\hat{\mathbf{r}}^{p}_{\mathrm{out}} = {}^{c}\hat{\mathbf{p}}^{p} - {}^{c}\mathbf{d}^{p}$, where ${}^{c}\hat{\mathbf{r}}^{p}_{\mathrm{out}}$, ${}^{c}\hat{\mathbf{p}}^{p}$, and ${}^{c}\mathbf{d}^{p}$ denote the estimated outer ray vector, the estimated position vector, and the variation vector about the point $p$ seen from the camera coordinate $c$, respectively. The position ${}^{c}\hat{\mathbf{p}}^{p}$ denotes the objective variable of bundle adjustment. The variation ${}^{c}\mathbf{d}^{p}$ can be calculated geometrically.

Even if the inconsistent candidates of initial values are removed by viewpoint selection, the positions of the point become inconsistent during optimization. Thus, it is needed to consider the inconsistent estimated values in the evaluation function by introducing the penalty function $e_2$. The penalty function $e_2$ is as follows:

$$e_2 = \sum_{c\in\mathcal{C}}\sum_{p\in\mathcal{P}} \left\{ \max\left(0, w + l - {}^{c}\hat{z}^{p}\right) \right\}^2, \tag{4}$$

where ${}^{c}\hat{z}^{p}$ denotes the $z$ coordinate of ${}^{c}\hat{\mathbf{p}}^{p}$ which is the estimated position vector of point $p$.

The penalty is given to the geometrically inconsistent points according to Eq. (4) to ensure that the optimization is conducted on the object side of the refractive plate. This is because it is impossible to consider the inconsistent points in the camera side and inside of the refractive plate by the evaluation of the two ray vectors. If point $p$ does not satisfy Eq. (1), the penalty is given according to the distance from the plate.

Finally, the optimization is conducted by Levenberg-Marquardt method as described in Eq. (2). As a result, the optimized camera coordinates from the selected viewpoints and the optimized measurement points can be obtained.

## III. Simulation Experiment

### A. Experiment setup using Stanford Bunny

To verify the effectiveness of the proposed method, simulation experiments were performed. In the experiments, a 3D model of the Stanford Bunny consisting of 1,428 points was measured. Ten images taken from different ten viewpoints were prepared, and correspondences between images were given. The refractive index of air $n_1$ and the refractive plate $n_2$ were set to 1.00 and 1.49, respectively. The thickness of the refractive plate $w$ was 200 mm, and the distance between the camera and the refractive plate $l$ was 50 mm. The coefficient $\alpha$ was set to 1. The quantization error of the spatial sampling of the measurement points was set to pixel accuracy to confirm that the proposed method can reconstruct with high accuracy even if the quantization error is large.

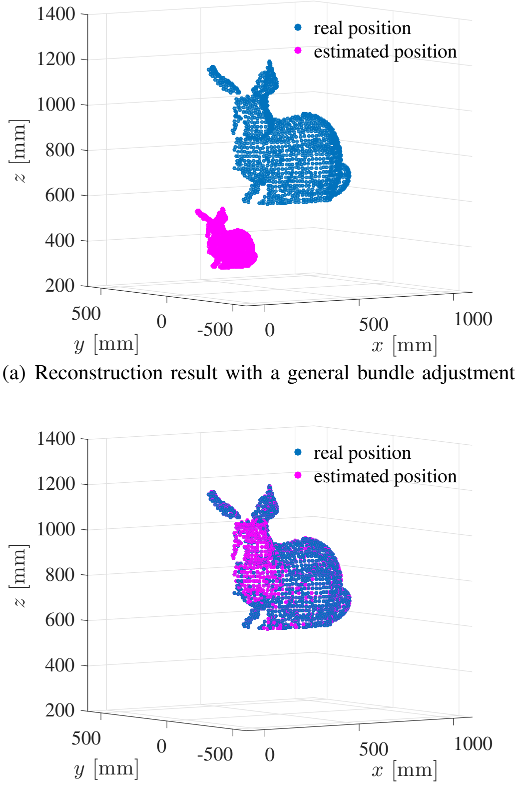
### B. Result of simulation experiment

Reconstruction results are shown in Fig. 4. Figure 4 (a) shows the reconstruction result of the Stanford Bunny with a conventional bundle adjustment without considering influence of the refraction, and Fig. 4 (b) shows the reconstruction result with the proposed method. In the proposed method, five viewpoints were selected from ten candidates of the initial values, and bundle adjustment was conducted. The average of mean absolute error from the actual positions for all points were 705.15 mm in Fig. 4 (a) and 0.64 mm in Fig. 4 (b).

From these results, it is confirmed that the proposed method is effective for SfM using refraction under the large quantization error.

### C. Effect of thickness of refractive plate

To verify the effect of the plate thickness on the proposed method, simulation experiments are performed using plates of different thicknesses. The thickness of the refractive plate was set to 50 mm, 100 mm and 200 mm. Two randomly generated point clouds were reconstructed. The point cloud comprised 200 points, and 20 images from different viewpoints were used in one experiment. Simulation experiments were performed by using different 100 sets of viewpoints of the camera. The




(a) Reconstruction result with a general bundle adjustment



(b) Reconstruction result with the proposed method

Fig. 4. Reconstruction results of the Stanford Bunny with a general bundle adjustment (a) and the proposed method (b).

TABLE I
EFFECT OF THICKNESS OF REFRACTIVE PLATE

| Thickness of plate | 50 mm | 100 mm | 200 mm |
|---|---|---|---|
| Random points 1 | 20 / 100 | 62 / 100 | 100 / 100 |
| Random points 2 | 36 / 100 | 77 / 100 | 100 / 100 |

refractive index $n_1$ and $n_2$, the distance between the camera and the refractive plate $l$, the coefficient $\alpha$ and the quantization error of the measurement points were on the same conditions as Section III – $A$.

In accuracy evaluation, a trial was regarded as successful if the average of mean absolute error from the actual positions for all points was under 1 mm, and the number of the successful trials was counted. Table I shows the results. The thicker the refractive plate was, the more the number of reconstruction with high accuracy was. These results indicate that the proposed method is greatly affected by the thickness of the refractive plate.

### D. Comparison with conventional bundle adjustment

Two randomly generated point clouds were reconstructed to compare the proposed method with the conventional bundle adjustment. The thickness of the refractive plate was set to 200 mm. Other conditions were the same as Section III – $C$. In these experiments, the number of successful trials was also counted in the same way. The results were shown in Table II. These results show that the number of reconstruction with high accuracy increase using the proposed method.

TABLE II
COMPARISON OF CONVENTIONAL METHOD AND PROPOSED METHOD

| | Conventional method | Proposed method |
|---|---|---|
| Random points 1 | 13 / 100 | 100 / 100 |
| Random points 2 | 15 / 100 | 100 / 100 |

## IV. CONCLUSION

In this paper, we proposed a novel method of global optimization using bundle adjustment for SfM with refraction to improve the accuracy of 3D reconstruction. The viewpoint selection and the evaluation function of optimization for bundle adjustment using the geometrical consistency of the measurement were proposed. In simulation experiments, the effectiveness of our proposed method was verified.

Improvement of the accuracy using the thinner refractive plates and actual experiments are our future works.


## ACKNOWLEDGMENTS

This work was in part funded by Tough Robotics Challenge, ImPACT Program of the Council for Science, Technology and Innovation (Cabinet Office, Government of Japan). The 3D model of the Stanford Bunny using in the experiments was provided by the Stanford 3D Scanning Repository.


## REFERENCES


[1] Richard Hartley and Andrew Zisserman: Multiple View Geometry in Computer Vision, Cambridge University Press, Second Edition, 2004.

[2] Davide Scaramuzza, Friedrich Fraundorfer, Marc Pollefeys, and Roland Siegwart: "Absolute Scale in Structure from Motion from Single Vehicle Mounted Camera by Exploiting Nonholonomic Constraints", Proceedings of the 2009 IEEE International Conference on Computer Vision, pp. 1413–1419, 2009.

[3] Ashwin P. Dani, Nicholas R. Fischer, and Warren E. Dixon: "Single Camera Structure from Motion", IEEE Transactions on Automatic Control, Vol. 57, No. 1, pp. 238–243, 2012.

[4] Manolis Lourakis and Xenophon Zabulis: "Accurate Scale Factor Estimation in 3D Reconstruction", Proceedings of the International Conference on Computer Analysis of Images and Patterns, Vol. 1, pp. 498–506, 2013.

[5] Akira Shibata, Hiromitsu Fujii, Atsushi Yamashita, and Hajime Asama: "Scale-Reconstructable Structure from Motion Using Refraction with a Single Camera", Proceedings of the 2015 IEEE International Conference on Robotics and Automation, pp. 5239–5244, 2015.

[6] Akira Shibata, Hiromitsu Fujii, Atsushi Yamashita, and Hajime Asama: "Absolute Scale Structure from Motion Using a Refractive Plate", Proceedings of the 2015 IEEE/SICE International Symposium on System Integration, pp. 540–545, 2015.

[7] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon: "Bundle Adjustment - A Modern Synthesis", Vision Algorithms: Theory and Practice, pp. 298–372, 2000.

[8] Ji Zhang, Mireille Boutin, and Daniel G. Aliaga: "Robust Bundle Adjustment for Structure from Motion", Proceedings of the 2006 IEEE International Conference on Image Processing, pp. 2185–2188, 2006.

[9] Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz, and Richard Szeliski: "Building Rome in a day", Proceedings of the 2009 IEEE International Conference on Computer Vision, pp. 72–79, 2009.

[10] Maxime Lhuillier: "Incremental Fusion of Structure-from-Motion and GPS Using Constrained Bundle Adjustments", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 34, No. 12, pp. 2489–2495, 2012.

[11] Soulaiman El Hazzat, Abderrahim Saaidi, and Khalid Satori: "Structure from Motion for 3D Object Reconstruction based on Local and Global Bundle Adjustment", Proceedings of the 2015 Third World Conference on Complex Systems, pp. 1–6, 2015.

[12] Anne Jordt-Sedlazeck and Reinhard Koch: "Refractive Structure-from-motion on Underwater Images", Proceedings of the 2013 IEEE International Conference on Computer Vision, pp. 57–64, 2013.