2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

3D Reconstruction under Light Ray Distortion from Parametric Focal Cameras

Satoshi Morinaka Fumihiko Sakaue Jun Sato
Nagoya Institute of Technology
Email: {morinaka@cv., sakaue@, junsato@}nitech.ac.jp

Kazuhisa Ishimaru Naoki Kawasaki
NIPPON SOKEN, INC.
Email: {kazuhisa_ishimaru,naoki_kawasaki}@soken1.denso.co.jp

***Abstract*—In this paper, we propose a new camera model for reconstructing 3D objects under light ray distortion caused by refractive medias. The proposed method can reconstruct 3D scene, even if light rays projected into the cameras are refracted by the refractive media, such as glasses and raindrops. For this objective, we represent light ray projection of multiple cameras by using a pair of planes shared by the multiple cameras in the scene. By using this model, intrinsic and extrinsic camera parameters as well as the refractive properties of the refractive media can be represented efficiently. By using the newly defined camera model, we propose a method for recovering 3D points and camera parameters with refractive properties simultaneously. The experimental results show the efficiency of the proposed camera model and reconstruction method.**

I. Introduction

Recently, 3D measurement from stereo cameras is widely used in various kinds of fields. Especially, in the field of ITS (Intelligent Transportation Systems), vehicle cameras are used for measuring the 3D distance toward objects and the danger of collision on the road [1], [2], and almost all the vehicles equip multiple cameras for various kinds of purpose in recent years.

The vehicle cameras are in general equipped inside of the vehicle for avoiding rains and dusts, and they observe road scenes through windshield glasses. However, such setup causes serious problems which do not occur under ordinary camera setups. If cameras observe 3D scenes through glasses, light rays projected into the cameras are refracted by the glasses, and then, observed images are often distorted as shown in Fig.1 (right). In addition, these refractions occur not only by the glasses but also by raindrops on the glasses as shown in Fig.1 (left). In such cases, the light rays expected from images by using the ordinary camera model is extremely different from the actual light rays in the 3D space as shown in Fig.2. Thus, if we reconstruct 3D points from these distorted images, the reconstructed points are far from the real 3D points as shown in Fig.2.

Thus, in this paper, we propose a new method for reconstructing 3D scenes from images observed under refractive media such as windshields. In particular, we propose a new camera model for representing the arbitrary refraction of light fields, and propose a method for estimating the parameters of this camera model and 3D points simultaneously by using a bundle adjustment. Thus, the proposed method can recover not just 3D position of objects, but also the refractive properties of the refractive media. The proposed method can reconstruct 3D scene accurately, even if the projected light rays to the cameras are refracted by the media such as glasses and raindrops.

Fig. 1. Image distortion by refractive media. The left image is distorted by windshield of a vehicle and right image is distorted by raindrops on the windshield.

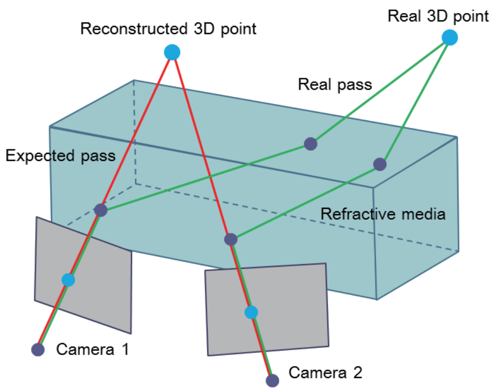


Fig. 2. Incorrect 3D reconstruction caused by a refractive media. In this case, light rays are refracted by the media, and thus, a reconstructed 3D point is very far from a real 3D point.

II. Related works

For reconstructing 3D scene under light refractions, several methods were proposed. In these methods, not the ordinary projective cameras, but non-single focal cameras, such as parametric focal camera model [3], [4], [5], [6], were used. In these camera models, light rays projected into the cameras do not need to be converged to a single optical center. Thus, the paths of light rays can be arbitrary, and the model can represent various kinds of image distortions caused by refractive medias.

In particular, the ray-pixel camera model [3] has extremely high freedom for representing input light rays. In this model, paths of the light rays are recorded pixel by pixel. As a result, it can represent arbitrary complicated light rays under refraction. However, this camera model needs very accurate camera calibration for using it practically. In addition, the camera parameters of this camera model change drastically under camera motions, even if the camera motions are small. Furthermore, it consists of a large number of camera parameters for representing light rays pixel by pixel, and thus, it often becomes unstable if the calibration images are not sufficient.

To cope with this problem, parametric models for representing light rays were also studied[6], [5]. In these methods, light rays were represented by two planes like 4D light field representation, and non-linear parametric mapping from the image plane to these two planes were estimated. By using the parametric models, the number of camera parameters can be reduced, and the stable calibration of cameras can be achieved. However, the number of parameters of these models is still much larger than the ordinary projective camera model, and thus, we need large number of images for calibrating these cameras accurately. Although a bundle adjustment based on the parametric focal camera model has also been proposed [4], it often becomes unstable since a large number of parameters must be estimated.

One of the reason of increasing of the number of parameters in the existing methods is the separation of intrinsic and extrinsic camera parameters. In the existing camera models, there are two kinds of parameters. One is intrinsic camera parameters, which represent mapping from an image plane to calibration planes. The other one is extrinsic camera parameters, which represent the relationship among multiple cameras. However, the parametric focal camera model has large freedom in intrinsic parameters, and the intrinsic parameters can describe not only internal parameters of the cameras but also the relationship among multiple cameras. Thus, we can decrease the number of parameters, if we integrate the extrinsic parameters with the intrinsic parameters in the camera models. Based on this observation, we in this paper propose a new camera model, which can represent multiple non-single focal cameras efficiently by using small number of parameters.

III. Ray-Pixel Camera Model

A. *Light Ray Representation using Calibration Planes*

We first describe the existing ray-pixel camera model [3]. In this model, light rays projected into images are not converged to an optical center of a camera. Therefore, the model cannot be represented by the ordinary linear camera model using a 3×4 projection matrix. Instead of the projection matrix based on the optical center, these light rays are represented by two planes set in the 3D scene as shown in Fig.3. The orientation and the position of a light ray are described by two points on these planes. For example, if a light ray from a point $\mathbf{X}$ passes through a point $\Pi_1(\mathbf{m})$ on the plane Π_1 and a point $\Pi_2(\mathbf{m})$ on the plane Π_2, then the light ray is represented by these two points. Thus, the degree of freedom of a light ray is 4, when the light ray goes straight in the 3D space. This light ray description is also used for representing light fields

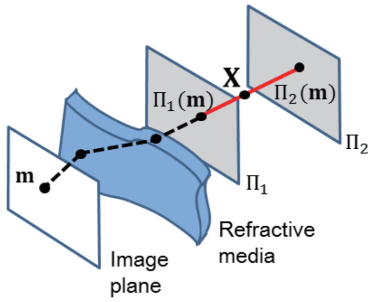


Fig. 3. Light ray representation using two calibration planes

in resent years. In this paper, we call these two planes a set of calibration planes.

Let us consider a case where 3D point $\mathbf{X}$ is observed through some refractive medias such as glasses, and is projected to $\mathbf{m} = [m_x, m_y]^\top$ in the image as shown in Fig.3. Although a light ray from a 3D point $\mathbf{X}$ is refracted by the media, the ray goes straight before and after the media. Therefore, the light ray which goes through the 3D point $\mathbf{X}$ can be described by using two points on the calibration planes, Π_1 and Π_2, in the 3D space. Suppose Z_1 and Z_2 are the depth of these two planes in the camera coordinates, and let $\mathbf{x}_1 = [x_1, y_1]^\top$ and $\mathbf{x}_2 = [x_2, y_2]^\top$ be 2D coordinates of points on these planes. Then, the 3D point $\mathbf{X}$ can be represented by using $\Pi_1(\mathbf{m}) = [x_1, y_1, Z_1]^T$ and $\Pi_2(\mathbf{m}) = [x_2, y_2, Z_2]^T$ as follows:

$$\lambda\widetilde{\mathbf{X}} = a\widetilde{\Pi}_1(\mathbf{m}) + b\widetilde{\Pi}_2(\mathbf{m}) \tag{1}$$

where $\tilde{}$ denotes a homogeneous coordinates, and a and b denote coefficients which indicate the 3D point. λ is a scalar.

If we have complete map from $\mathbf{m}$ to $\Pi_i(\mathbf{m})$ pixel by pixel, arbitrary light refraction by refractive media can be described precisely. By using light ray tracing technique, we can obtain the pixel-wise map from images to $\Pi_i(\mathbf{m})$. However, the estimation of the pixel-wise map is not realistic in the real scene, since the map will change drastically under camera motions, even if the camera motions are small. In addition, the calibration of the light ray needs many calibration images and time. Thus, the pixel-wise light ray calibration is not practical.

To cope with this problem, non-linear parametric mapping is used in general. In this case, a 2D image point $\mathbf{m}$ is mapped to calibration planes Π_i by using parametric functions such as polynomial functions. Furthermore, the relationship among multiple cameras is described by using the relative positions and orientations among the calibration planes of multiple cameras. The map from the image plane to the calibration planes corresponds to the intrinsic parameters in the ordinary projective camera and the relationship among calibration planes of cameras corresponds to the extrinsic parameters.

This description based on the intrinsic and extrinsic parameters is useful for generic scenes where the relationship among cameras is not constraint. However, the description is redundant, and is not efficient. For example, when cameras are set on a grid, i.e. camera array, not only intrinsic

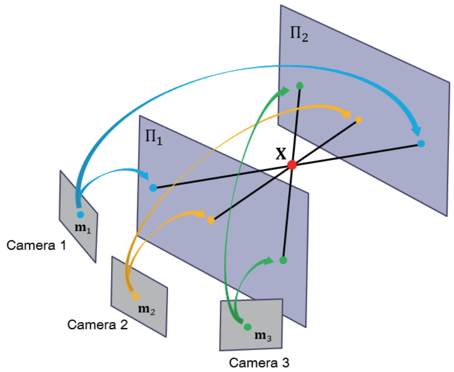


Fig. 4. Parametric focal camera model for multiple cameras based on a single pair of calibration planes

parameters but also extrinsic parameters, i.e relative position and orientation of cameras, can be described by using the parametric mapping functions. In this paper, we propose a new parametric focal camera array model, in which the extrinsic and intrinsic camera parameters are represented by a single parametric mapping function. In the proposed method, only a single pair of calibration planes are used for describing light rays projected into multiple cameras. By using this description, we achieve efficient and stable camera calibration and 3D reconstruction.

IV. Parametric Focal Camera Array Model

A. *Camera Array Model Using Shared Planes*

We next propose a new parametric focal camera array model, in which the light rays go into multiple cameras are described by using a single pair of parallel calibration planes as shown in Fig.4. In the existing parametric focal camera models, a set of light rays projected into N cameras is described by N pairs of calibration planes defined at each camera. In contrast, all light rays projected into all the cameras are described by a single pair of common planes in the proposed method. Therefore, not only intrinsic camera parameters but also extrinsic camera parameters are described by mapping from image planes to a pair of calibration planes. The proposed camera model can be calibrated efficiently and stably, since the number of camera parameters is much less than the existing parametric focal camera models.

In the following sections, we describe a method for reconstructing 3D scene based on the proposed camera model.

B. *Nonlinear image mapping*

We first consider nonlinear mapping from an image plane to a pair of calibration planes in the 3D space. This nonlinear mapping enables us to represent a correspondence between an image point and a light ray in the 3D space under the existence of light refractions. In our method, we use polynomial functions for mapping points.

Let us consider a point $\mathbf{m} = [m_1, m_2]^T$ in the image and 2D points $\mathbf{x}^i = [x_1^i, x_2^i]^T$ on a pair of calibration planes $\Pi_i (i = 1, 2)$. Then, the point mapping from the image to the 2D points on a pair of calibration planes can be described by using polynomial functions as follows:

$$x_j^i = \sum_{k=0}^{K} \sum_{l=0}^{K-k} a_{kl}^{ij} m_1^k m_2^{k-l} (i = 1, 2; j = 1, 2) \tag{2}$$

where K is the order of the polynomial function. The coefficients a_{kl}^{ij} are camera parameters in this model. Therefore, the camera calibration of this model is equivalent to the estimation of these coefficients. Note, the polynomial function can represent more complicated mapping when K is large. However, the estimation (calibration) of the coefficients becomes more unstable if K is large, since the number of parameters becomes large. Thus, K may be three or less than three in practice.

By using the nonlinear mapping, flexibility of the camera model become higher than the ordinary single focal camera models. However, this mapping is not enough for representing both intrinsic and extrinsic camera parameters.

To clarify the problem, let us consider the case where $K = 1$. In this case, the mapping function based on the polynomial function is equivalent to the 2D affine transformation. In the ordinary projective camera model, the affine transformation is not sufficient to represent rotation of the cameras, and we need the projective transformation to represent image distortions caused by the camera rotation.

Thus, we in this paper consider a new image mapping function by combining polynomial mapping functions and projective image transformation as follows:

$$\lambda x_j^i = \sum_{k=0}^{K} \sum_{l=0}^{K-k} a_{kl}^{ij} m_1^k m_2^{k-l} (i = 1, 2; j = 1, 2) \tag{3}$$

$$\lambda = \sum_{k=0}^{1} \sum_{l=0}^{K-k} a_{kl}^{i3} m_1^k m_2^{k-l} (i = 1, 2) \tag{4}$$

In this mapping, K-th order polynomial functions and a 2D projective transformation are combined together, so that the new mapping function can represent 2D projective transformation as well as K-th order polynomial transformation. For example, if $K = 1$, the combined mapping function becomes a standard 2D projective transformation. Since the camera rotation and planar translation can be represented by a 2D projective transformation, the mapping function shown in Eq.(3) can represent not only the intrinsic parameters but also the extrinsic parameters of the camera efficiently.

C. *3D Reconstruction by Proposed Camera Model*

We next consider 3D reconstruction from the newly defined parametric focal camera array model. Let us consider the case where there are N cameras in the scene and they are described by our new camera model. Suppose a light ray which goes through a 3D point $\mathbf{X}$ is projected to $\mathbf{m}_n$ in the image of the n-th camera as shown in Fig. 5. Then, the 3D point $\mathbf{X}$ can be described as follows:

$$\mathbf{X} = \alpha \mathbf{B}_n + \mathbf{X}_n^1 \tag{5}$$

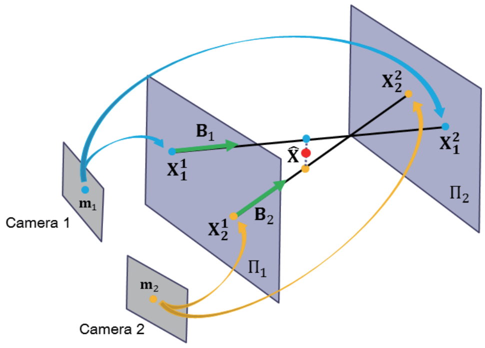


Fig. 5. 3D reconstruction from a set of estimated light rays.

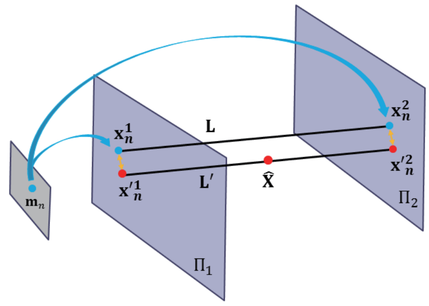


Fig. 6. Reprojection error on calibration planes. In the proposed parametric focal camera model, the reprojection error is defined not on the image plane but on the calibration planes.

where, $\mathbf{X}_n^i = [x_{1n}^i, x_{2n}^i, Z_i]^T$ is a 3D point on a plane Π_i mapped from $\mathbf{m}_n$, and $\mathbf{B}_n$ is a 3-vector which represents the orientation of the light ray as follows:

$$\mathbf{B}_n = \frac{\mathbf{X}_n^2 - \mathbf{X}_n^1}{\|\mathbf{X}_n^2 - \mathbf{X}_n^1\|} \tag{6}$$

Then, a 3D point $\hat{\mathbf{X}}$ can be reconstructed from a set of N light rays by the least means square method as follows:

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmin}} \sum_{n=1}^{N} (\|\mathbf{X} - \mathbf{X}_n^1\|^2 - (\mathbf{B}_n^T(\mathbf{X} - \mathbf{X}_n^1))^2) \tag{7}$$

In this equation, distances from the reconstructed point to each light rays are minimized.

V. Simultaneous Camera Calibration and 3D Reconstruction Using Bundle Adjustment

We next consider simultaneous camera calibration and 3D reconstruction using a framework of bundle adjustment. The recovered camera parameters include not just the intrinsic parameters of usual cameras, but also the refractive properties of the refractive media. Hence, the proposed method can reconstruct 3D scene, even if the refractive properties and cameras are unknown.

In the ordinary camera model, i.e. projective camera model, the bundle adjustment can be achieved by minimizing reprojection errors in input images. However, the definition and estimation of the reprojection error in the parametric focal camera model is not so simple because of the non-linear mapping. In this model, the estimation of reprojection error on the image planes needs non-linear computation, and thus, the estimation of reprojection error on the image plane is complicated. Thus, we in this research define reprojection errors on a pair of calibration planes Π_1 and Π_2. The new reprojection errors do not require non-linear estimation, and thus, we can minimize the reprojection errors efficiently.

Let us consider the case where a 3D point $\hat{\mathbf{X}}$ is reconstructed from N cameras in our proposed framework. Let $\mathbf{x}_n^j (j = 1, 2 : n = 1, \cdots, N)$ be a 2D point on Π_j mapped from the n-th camera. $\mathbf{L}$ represents a light ray defined by a pair of points on the calibration planes, and $\mathbf{L}'$ represents a ray which is parallel to $\mathbf{L}$ and goes through the 3D point $\hat{\mathbf{X}}$ as shown in Fig.6. The intersection of $\mathbf{L}'$ with a plane Π_j is denoted by $\mathbf{x'}_n^j (j = 1, 2 : n = 1, \cdots, N)$. Then, we define the reprojection error E on the calibration planes Π_1 and Π_2 as follows:

$$E = \sum_{n=1}^{N} \sum_{j=1}^{2} \|\mathbf{x}_n^j - \mathbf{x'}_n^j\|^2 \tag{8}$$

If there are Q 3D points in the scene and $E_q (q = 1, \cdots, Q)$ is the reprojection error of the q-th 3D point $\mathbf{X}_q$, then the optimal camera parameters $\{\hat{a}_{kl}^{ijn}\}$ and a set of reconstructed points $\{\hat{\mathbf{X}}_q\}$ can be estimated as follows:

$$\{\{\hat{a}_{kl}^{ijn}\}, \{\hat{\mathbf{X}}_q\}\} = \underset{a_{kl}^{ijn}, \mathbf{X}_q}{\operatorname{argmin}} \sum_{q=1}^{Q} E_q \tag{9}$$

The initial values of $\{\mathbf{X}_q\}$ are given by 3D reconstruction from the standard single focal camera model, and the initial values of $\{a_{kl}^{ijn}\}$ are computed from the initial values of $\{\mathbf{X}_q\}$.

We next consider reconstruction ambiguity of the bundle adjustment in our proposed model. In general, reconstructed results by uncalibrated camera include ambiguity such as projective ambiguity from projective cameras. This ambiguity can be eliminate using some kinds of prior knowledge such as camera parameters or basis points in the input scene. In our method, we use known 3D points in the scene as basis points to eliminate the ambiguity.

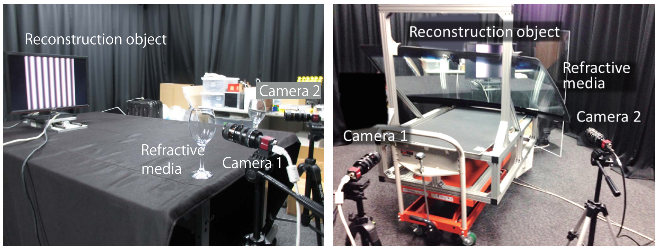
Let us consider the case when there are R 3D points $\mathbf{X}_r^B (r = 1, \cdots, R)$ and their 3D position is known. When reprojection error of these points is E_r^B, 3D points except these basis points and camera parameters can be estimated as follows:

$$\{\{\hat{a}_{kl}^{ijn}\}, \{\hat{\mathbf{X}}_q\}\} = \underset{a_{kl}^{ijn}, \mathbf{X}_q}{\operatorname{argmin}} (\sum_{q=1}^{Q} E_q + w \sum_{r=1}^{R} E_r^B) \tag{10}$$

where w is a weight of the basis points. By using the basis points, ambiguity of the camera parameters and reconstructed 3D points can be eliminated.

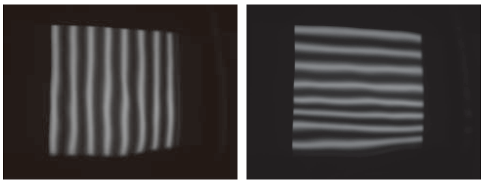
VI. Experimental Results

We next show experimental results from the proposed method. In our first experiments, wine glasses were put in

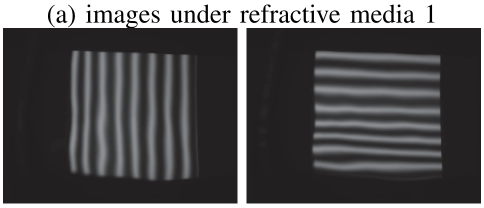


(a) wine glass (b) windshield

Fig. 7. Experimental environment

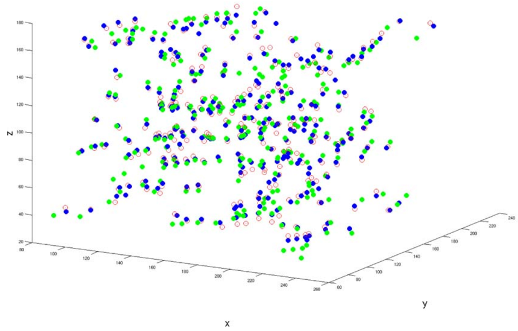


(a) images under refractive media 1

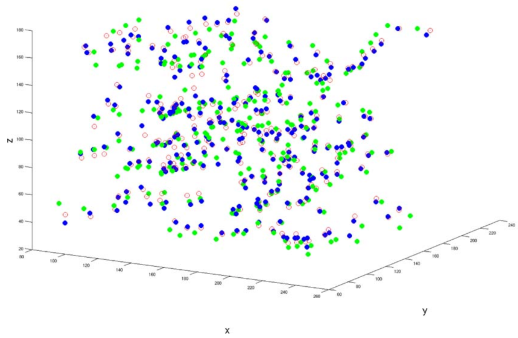


(b) images under refractive media 2

Fig. 8. Examples of input images with distortion



(a) refractive media 1



(b) refractive media 2

Fig. 9. Reconstructed 3D points from the proposed method under two different refractive media. In this reconstruction Eq.(3) with 3rd order and projective transformation was used.

front of two cameras as refractive medias as shown in Fig.7 (a), and the position and the orientation of the glasses were changed randomly, so that we can obtain images under different unknown refractive medias. In our experiments, a display system was set in the scene to generate artificial 3D points with very accurate ground truth data. The display system was put on a moving stage, and its depth was changed by the stage. We showed a set of phase shift patterns on the display system, and their images were taken by a set of stereo cameras. The observed images were used for generating 2D image points in the stereo images. These image points were used for reconstructing 3D points represented by the phase shift patterns at each depth.

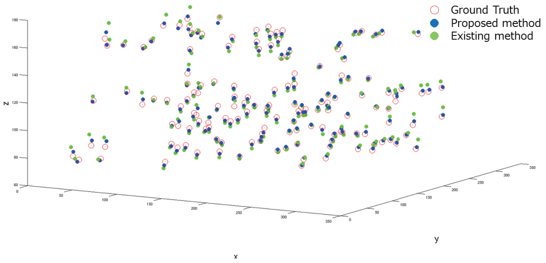
Fig. 8 (a) and (b) show example images taken under two different refractive medias, i.e. two different settings of wine glasses. As shown in these images, the display patterns were strongly distorted according to the refractive medias. We chose 300 corresponding points in stereo images randomly, and reconstructed them. In order to evaluate the effect of high order terms in the polynomial function, the order of the polynomial equation was changed from the 1st to the 3rd, and the accuracy of 3D reconstruction was compared. Also, the camera model with projective transformation in Eq.(3) is compared with that without projective transformation in Eq.(2).

Fig. 9 (a) and (b) show 3D points reconstructed from the proposed method with 3rd order polynomial function under two different refractive medias. The red points show ground truth, the blue points show reconstructed points from the proposed method and the green points show results from the standard single focal camera model. As we can see in these figures, the 3D points recovered from the standard camera model change drastically according to the refractive medias, and the results are very different from the ground truth. On the contrary, the results from the proposed method are almost identical with the ground truth data, even if the properties of refractive media change drastically. Table I shows the comparison of RMS errors of each reconstruction method in different orders of polynomial equation. These results show that the proposed camera model can reconstruct 3D points much more accurately than the ordinary single focal camera model by increasing the order of polynomial function. From these results, we find that the proposed method can recover 3D points accurately under arbitrary refractive medias, even if their refractive properties are unknown.

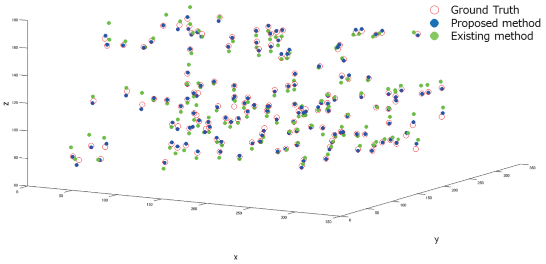
TABLE I
RMS ERRORS OF 3D RECONSTRUCTION [MM]

	proposed model	single focal model
1st order	6.26	6.51
2nd order	4.88	
3rd order	3.39	

We next show the results by using an windshield glass as a refractive media as shown in Fig. 7 (b). The 3D points

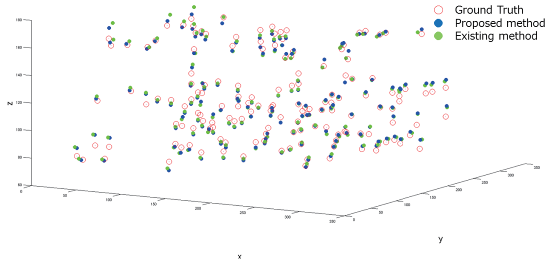


(a) 1st order polynomial

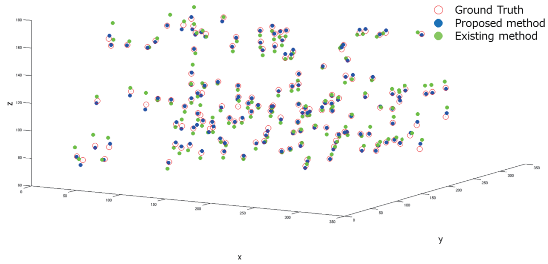


(b) 3rd order polynomial

Fig. 10. Reconstructed results from the proposed method. In this reconstruction, Eq.(3) with projective transformation was used.



(a) 1st order polynomial



(b) 3rd order polynomial

Fig. 11. Reconstructed results from Eq.(2) without projective transformation.

TABLE II
RMS ERRORS OF 3D RECONSTRUCTION [MM]

	with projective transformation	without projective transformation
Single focal	4.29	
1st order	2.22	4.15
2nd order	2.56	2.85
3rd order	2.28	2.21

reconstructed from the proposed camera model are shown in Fig. 10, and the 3D points reconstructed from the camera model without projective transformation in Eq.(2) are shown in Fig. 11. Their RMS errors are also shown in table II. Although the image distortion is not so large in this case, the result from the standard camera model has large errors, while the proposed method provides us accurate 3D reconstruction. Also, we find that the accuracy of the proposed method with projective transformation is much better than that of Eq.(2), even if we use low order polynomial function, i.e. 1st order, as shown in table II. This means that the proposed camera model with projective transformation can represent extrinsic parameters as well as intrinsic parameters and light distortions more efficiently than the polynomial function without projective transformation. Thus, we can decrease the degree-of-freedom of camera model without degrading the accuracy of 3D reconstruction in the proposed method.

VII. Conclusion

In this paper, we proposed a new method for reconstructing 3D objects under the existence of refractive medias in the scene. The proposed method can reconstruct 3D scene accurately, even if light rays projected into cameras are refracted by the refractive media such as glasses. For this objective, we proposed a new camera models by using a pair of planes shared by multiple cameras. By using this model, not only intrinsic camera parameters but also extrinsic camera parameters and refraction of lights can be represented efficiently. In addition, we presented a method for estimating camera parameters and 3D points simultaneously by using the bundle adjustment. The experimental results showed that the proposed method provides us accurate 3D reconstruction, even if we have unknown varying refractive medias.

References

[1] K.Nomoto, F.Sakaue, and J.Sato, "Raindrop complement based on epipolar geometry and spatiotemporal patches," in *International Conference on Computer Vision Theory and Applications*, 2011, pp. 175–180.

[2] Y.Watanabe, F.Sakaue, and J.Sato, "Time-to-contact from image intensity," in *Computer Vision and Pattern Recognition*, 2015, pp. 4176–4183.

[3] M.Grossberg and S.Nayer, "The raxel imaging model and ray-based calibration," *International Journal of Computer Vision*, vol. 61, no. 2, pp. 119–137, 2005.

[4] S.Ramalingam, S.Lodha, and P.Sturm, "A generic structure-from-motion framework," *Computer Vision and Image Understanding*, vol. 103, no. 3, pp. 218–228, 2006.

[5] T.Yano, S.Nobuhara, and T.Matsuyama, "3d shape from silhouettes in water for online novel-view synthesis," *Information Processing Society of Japan, Transactions on Computer Vision and Applications*, vol. 5, no. 0, pp. 65–69, 2013.

[6] M.Nishimura, S.Nobuhara, T.Matsuyama, S.Shimizu, and K.Fujii, "A linear generalized camera calibration from three intersecting reference planes," in *International Conference on Computer Vision*, 2015, pp. 2354–2362.