Two-View Underwater Structure and Motion for Cameras under Flat Refractive Interfaces

Lai Kang[1,3], Lingda Wu[1,2], and Yee-Hong Yang[3]

[1] College of Information System and Management, National University of Defense Technology, Changsha, China
lkang.vr@gmail.com
[2] The Key Laboratory, Academy of Equipment Command & Technology, Beijing, China
wld@nudt.edu.cn
[3] Department of Computing Science, University of Alberta, Edmonton, Canada
yang@cs.ualberta.ca

Abstract. In an underwater imaging system, a refractive interface is introduced when a camera looks into the water-based environment, resulting in distorted images due to refraction. Simply ignoring the refraction effect or using the lens radial distortion model causes erroneous 3D reconstruction. This paper deals with a general underwater imaging setup using two cameras, of which each camera is placed in a separate waterproof housing with a flat window. The impact of refraction is explicitly modeled in the refractive camera model. Based on two new concepts, namely the Ellipse of Refrax (EoR) and Refractive Depth (RD) of a scene point, we show that provably optimal underwater structure and motion under L_∞-norm can be estimated given known rotation. The constraint of known rotation is further relaxed by incorporating two-view geometry estimation into a new hybrid optimization framework. The experimental results using both synthetic and real images demonstrate that the proposed method can significantly improve the accuracy of camera motion and 3D structure estimation for underwater applications.

1 Introduction

Structure and motion from images is an active research topic in computer vision [1]. While remarkable success has been achieved in the last decade for land-based systems, accurate 3D reconstruction from images captured by underwater cameras, however, has not attracted much attentions in the computer vision community only until recently [2][3]. The key challenge is that the refraction of light occurs when a light ray passing through different media, rendering the perspective camera model invalid.

In early works, the effects of refraction in underwater 3D reconstruction are simply ignored [4] or modeled by approximate methods, such as focal length adjustment [5], lens radial distortion [6] and a combination of the two [7]. Unfortunately, these methods are insufficient since the effect of refraction is known to be highly non-linear and depends on the 3D location of a scene. As shown by

A. Fitzgibbon et al. (Eds.): ECCV 2012, Part IV, LNCS 7575, pp. 303–316, 2012.

Treibitz et al. [8], assuming a single viewpoint (SVP) model can be erroneous for camera calibration in underwater applications.

A more desirable method to compensate for refraction is to use a physically correct refractive camera model. Chari and Sturm [2] analyze using theoretical analysis the underlying multi-view relationships between two cameras when the scene has a single refractive planar surface separating two different media. The authors demonstrate the existence of geometric entities such as the refractive fundamental matrix and the refractive homography matrix. Nevertheless, no practical application of these theoretical results is given in [2]. Chang and Chen [3] study a similar configuration involving multiple views of a scene through a single interface. Refractive distortion is explicitly modeled as a function of depth. In [3], an additional piece of hardware called inertial measurement unit (IMU) is required to provide the roll and pitch angles of the camera. Also, the normal of the refractive interface is assumed to be known. Based on this additional information, the authors derive a linear solution to the relative pose problem and a closed-form solution to the absolute pose problem. More recently, Agrawal et al. [9] show that the underlying refractive geometry corresponds to an axial camera and develop a general theory of calibrating such systems using a planar checkerboard.

Sedlazeck and Koch [10] study the calibration of housing parameters for underwater stereo camera rigs. Rather than minimizing the reprojection error in the image space, the error on the outer interface plane is minimized by deriving the virtual perspective projection [11] for each 3D point. One issue of this method is, as reported in [10], that the optimization process is time consuming (in the order of 3 hours). Compared with [10], our proposed algorithm allows more general configuration of cameras and can minimize the reprojection error in image space efficiently. Another limitation for most existing underwater photography works is that a calibration target with known dimensions is required to perform system calibration [12][11][8].

In this paper, we focus on structure and motion estimation for a general underwater imaging setup consisting of two cameras, of which each camera is placed in a separate waterproof housing with a flat window, without using any calibration object. The main contributions of this paper are: 1) Two new concepts, namely the Ellipse of Refrax (EoR) and the Refractive Depth (RD) of a scene point for the refractive camera model are presented. These two concepts facilitate the derivation of an algorithm which yields globally optimal estimation of relative camera translation, interface distances and 3D structure under L_∞-norm, given known camera rotation and the interface normal; 2) A new hybrid optimization framework is proposed to perform two-view underwater structure and motion. Within this framework, the constraint of known rotation is further relaxed and the reprojection errors in image space are minimized.

2 Refractive Camera Model

This section gives a brief review of the refractive camera model and presents two new concepts to facilitate the recovery of underwater structure and motion.

2.1 Notations and Background

The refractive camera model which consists of a conventional perspective camera model and a refractive interface is shown in Fig. 1(a). This subsection introduces notations used in the refractive camera model, backward projection and forward projection.

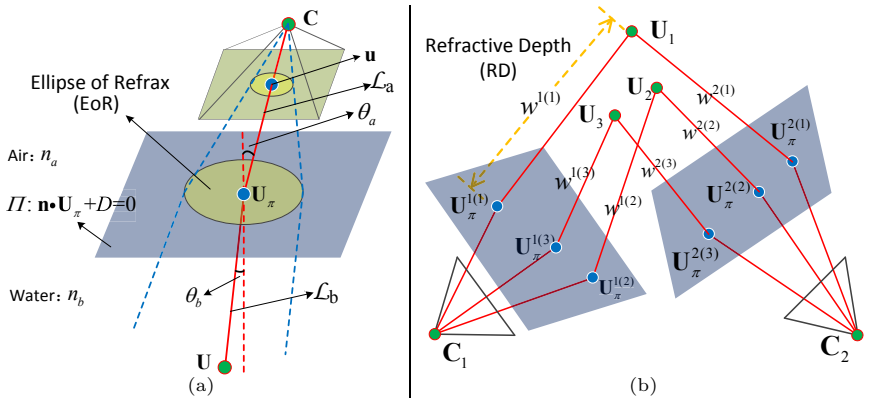

Fig. 1. Illustration of the refractive camera model and two new related concepts. (a) A perspective camera centered at $\mathbf{C}$ in the air observes a 3D point $\mathbf{U}$ in water. The light ray is refracted at the refractive interface Π. The Ellipse of Refrax (EoR) of $\mathbf{U}$ also lies on Π. (b) The Refractive Depth (RD) of 3D point. See text for details.

Back Projection calculates the refracted light ray which originates from the camera center and goes through the corresponding 3D scene point for a given image point. Let the camera projection matrix be of the form $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$, where $\mathbf{K}$ is the calibration matrix of the camera, $\mathbf{R}$ the rotation matrix and $\mathbf{t}$ the translation vector. As shown in Fig. 1(a), the corresponding light ray $\mathcal{L}_a$ in the air is determined by the camera center $\mathbf{C}$ and its direction is given by $\mathbf{r}_a = \mathbf{R}^{-1}\mathbf{K}^{-1}\mathbf{u}$, where $\mathbf{u}$ stands for the homogeneous coordinates of the image point. Given $\mathcal{L}_a$, the point $\mathbf{U}_\pi$ (which is called refrax according to [13]) where the refraction occurs can be determined by computing the intersection of $\mathcal{L}_a$ and the refractive interface Π. In order to calculate the direction of the refracted ray $\mathcal{L}_b$, Snell's law is applied, namely $n_a \sin\theta_a = n_b \sin\theta_b$, where n_a and n_b are the refractive indices of air and water, respectively. The direction of $\mathcal{L}_b$ can thus be written as [14]:

$$\mathbf{r}_b = \frac{n_a}{n_b}\frac{\mathbf{r}_a}{\|\mathbf{r}_a\|_2} - \left(\frac{n_a}{n_b}\cos\theta_a - \sqrt{1-\sin^2\theta_b}\right)\mathbf{n}, \tag{1}$$

where $\sin\theta_b$ can be rewritten as a function of $\cos\theta_a = \mathbf{n}\cdot\frac{\mathbf{r}_a}{\|\mathbf{r}_a\|_2}$ and $\mathbf{n}$ is the normal of Π.

Forward Projection calculates the projection of a 3D scene point onto the image plane. The forward projection of a 3D point under the refractive

camera model corresponds to solving a 4th degree polynomial. Details on forward projection can be found in [3] and [9].

2.2 Ellipse of Refrax (EoR) and Refractive Depth (RD) of a Scene Point

Let $\underline{\mathbf{u}}$ and $\mathbf{u}$ be the ground truth and the measured homogenous image coordinates (with the 3rd element equals to 1) of a scene point $\mathbf{U}$, respectively. For a perspective camera with camera projection matrix $\mathbf{P}$, the reprojection error is:

$$d(\mathbf{P},\mathbf{U},\mathbf{u}) = \|\mathbf{u}-\underline{\mathbf{u}}\|_2 = \frac{\left\|[\mathbf{P}]_1\tilde{\mathbf{U}} - [\mathbf{u}]_1[\mathbf{P}]_3\tilde{\mathbf{U}}, [\mathbf{P}]_2\tilde{\mathbf{U}} - [\mathbf{u}]_2[\mathbf{P}]_3\tilde{\mathbf{U}}\right\|_2}{[\mathbf{P}]_3\tilde{\mathbf{U}}} \tag{2}$$

where $\tilde{\mathbf{U}} = [\mathbf{U}^\top 1]^\top$ and $[\mathbf{P}]_k$ represents the k-th row vector of matrix $\mathbf{P}$. Also, without loss of generality, we assume that $[\mathbf{P}]_3\tilde{\mathbf{U}} > 0$. For the refractive camera model, it is incorrect to use Eq. (2) to calculate the reprojection error. However, since the light ray between the camera center $\mathbf{C}$ and refrax $\mathbf{U}_\pi$ is a straight line (see Fig. 1(a)), we can get a similar equation on $\mathbf{U}_\pi$ by replacing $\mathbf{U}$ with $\mathbf{U}_\pi$ in Eq. (2). In addition, the refrax $\mathbf{U}_\pi$ should lie on the refractive interface, which gives another linear constraint. Given a 3D scene point $\mathbf{U}$ with image point $\mathbf{u}$, we define its Ellipse of Refrax (EoR) as:

$$\mathcal{R}(\mathbf{P},\mathbf{n},\mathbf{u}) = \{\mathbf{U}_\pi | d(\mathbf{P},\mathbf{U}_\pi,\mathbf{u}) \le \gamma, \mathbf{n}\cdot\mathbf{U}_\pi + D = 0,\} \tag{3}$$

where γ is a threshold ($\gamma = 3$ pixels in this paper) specifying the largest reprojection error on image plane. For a camera with projection matrix $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$, assume that $\mathbf{K}$ and $\mathbf{R}$ are known, the first constraint of $\mathcal{R}$ can be rewritten as:

$$\|f_1(\mathbf{x}'), f_2(\mathbf{x}')\|_2 \le \gamma f_3(\mathbf{x}'), \tag{4}$$

where f_1, f_2 and f_3 are affine functions with unknown vector $\mathbf{x}' = (\mathbf{t}^\top, \mathbf{U}_\pi^\top)^\top$ and coefficients determined by $\mathbf{K}$, $\mathbf{R}$ and $\mathbf{u}$. For a fixed γ, Eq. (4) is known to be a Second Order Cone (SOC), which is convex [15][16]. Note that $\mathcal{R}$ corresponds to the intersection of a SOC and a plane, which is an ellipse (see Fig. 1(a)). Also, it is easy to see that, by assuming that the normal of the refractive interface $\mathbf{n}$ is known, the second constraint of $\mathcal{R}$ is linear in $\mathbf{x}'' = (\mathbf{U}_\pi^\top, D)^\top$. Based on the above discussion, we conclude that EoR defines a convex set for known $\mathbf{K}$, $\mathbf{R}$, $\mathbf{n}$ and image measurement $\mathbf{u}$.

Since EoR directly imposes constraint on the refrax $\mathbf{U}_\pi$ rather than on scene point, solely using EoR does not make sense for reconstructing a 3D point. Suppose a scene consisting of N 3D points $\mathbf{U}_j (j = 1, \cdots, N)$ is observed by two cameras with camera center $\mathbf{C}_i (i = 1, 2)$, the refrax of the j-th 3D point on the i-th interface is denoted by $\mathbf{U}_\pi^{i(j)}$ (see Fig. 1(b)). According to backward projection, each image measurement (or each refrax) imposes a linear constraint on its corresponding 3D scene point. For instance, the constraint for $\mathbf{U}_\pi^{i(j)}$ is given by:

$$\mathbf{U}_j = \mathbf{U}_\pi^{i(j)} + w^{i(j)}\mathbf{r}_b^{i(j)}, \tag{5}$$

where $\mathbf{r}_b^{i(j)}$ denotes the direction of the refracted ray that corresponds to refrax $\mathbf{U}_\pi^{i(j)}$. As the direction of the refracted ray is uniquely determined by $\mathbf{K}$, $\mathbf{R}$, $\mathbf{n}$ and image measurement $\mathbf{u}$ (see subsection 2.1), Eq. (5) generates three new independent linear constraints on $\mathbf{U}_j, \mathbf{U}_\pi^{i(j)}$ and the coefficient $w^{i(j)}$, which we call the Refractive Depth (RD) of 3D point $\mathbf{U}_j$ with respect to the i-th camera.

3 Underwater Structure and Motion with Known Rotation

In this subsection, we show that the constraints from EoR and RD presented in the aforementioned section can be imposed in a new formulation of the underwater with known rotation problem. In the context of this problem, the term rotation refers to the rotation of the perspective camera and the normal of the refractive interface.

3.1 Underwater Known Rotation Problem with Provably Optimal Solution

The underwater with known rotation problem (UKRP1) is formulated as the following min-max problem:

$$\begin{aligned} \mathsf{UKRP1} \quad & \min\max_{ij}\ d(\mathbf{P}_i, \mathbf{U}_\pi^{i(j)}, \mathbf{u}^{i(j)}) \\ & \text{subject to } \mathbf{n}_i \cdot \mathbf{U}_\pi^{i(j)} + D_i = 0, \\ & \qquad \mathbf{U}_j = \mathbf{U}_\pi^{i(j)} + w^{i(j)} \mathbf{r}_b^{i(j)}, \\ & \qquad \forall i = 1, 2, \\ & \qquad \forall j = 1, \cdots, N. \end{aligned} \tag{6}$$

with unknown vector

$$\mathbf{X} = \left(\mathbf{U}_1^\top, \cdots, \mathbf{U}_N^\top, {\mathbf{U}_\pi^{1(1)}}^\top, \cdots, {\mathbf{U}_\pi^{2(N)}}^\top, w^{1(1)}, \cdots, w^{2(N)}, \mathbf{t}_1^\top, \mathbf{t}_2^\top, D_1, D_2 \right)^\top. \tag{7}$$

The UKRP1 minimizes the L_∞-norm of the vector of reprojection errors on image plane. More conveniently, UKRP1 can be rewritten in its equivalent form:

$$\begin{aligned} \mathsf{UKRP2} \quad & \min\ \gamma \\ & \text{subject to } d(\mathbf{P}_i, \mathbf{U}_\pi^{i(j)}, \mathbf{u}^{i(j)}) \leq \gamma, \\ & \qquad \mathbf{n}_i \cdot \mathbf{U}_\pi^{i(j)} + D_i = 0, \\ & \qquad \mathbf{U}_j = \mathbf{U}_\pi^{i(j)} + w^{i(j)} \mathbf{r}_b^{i(j)}, \\ & \qquad \forall i = 1, 2, \\ & \qquad \forall j = 1, \cdots, N. \end{aligned} \tag{8}$$

The first two constraints in UKRP2 correspond to the EoR and the third constraint corresponds to the RD defined in subsection 2.2. As the constraints in UKRP2 are convex for a fixed γ, the solution to UKRP2 can be found by solving a

sequence of feasibility problems within a bisection procedure [16]. In particular, the underwater feasibility problem (UFSBP) is given by:

$$\begin{aligned} \mathsf{UFSBP} \qquad & \text{Given } \gamma \\ & \text{does there exist } \mathbf{X} \\ & \quad \text{subject to } d(\mathbf{P}_i, \mathbf{U}_\pi^{i(j)}, \mathbf{u}^{i(j)}) \leq \gamma, \\ & \qquad \mathbf{n}_i \cdot \mathbf{U}_\pi^{i(j)} + D_i = 0, \\ & \qquad \mathbf{U}_j = \mathbf{U}_\pi^{i(j)} + w^{i(j)} \mathbf{r}_b^{i(j)}, \\ & \qquad \forall i = 1, 2, \\ & \qquad \forall j = 1, \cdots, N. \end{aligned} \tag{9}$$

Since a feasibility problem does not have an objective function, we only need to examine whether all the constraints are satisfied for a given γ. Because all the constraints of UFSBP are convex, the feasibility problem UFSBP is also convex and can be solved efficiently using convex optimization [15].

3.2 Robust Formulation of the Underwater Known Rotation Problem

While the algorithm proposed in subsection 3.1 can estimate camera translation, interface distances and scene structure optimally, minimization under the L_∞-norm is known to be particularly sensitive to outliers [16]. In this paper, outliers are handled by introducing auxiliary variables as in [17]. Instead of solving a sequence of convex problems, satisfactory estimation of structure and motion can also be obtained by solving the following single convex optimization problem:

$$\begin{aligned} \mathsf{UKRP3} \qquad & \min \textstyle\sum_{j=1}^{N} s_j \\ & \text{subject to } d(\mathbf{P}_i, \mathbf{U}_\pi^{i(j)}, \mathbf{u}^{i(j)}) [\mathbf{P}_i]_3 \mathbf{U}_\pi^{i(j)} \leq \gamma [\mathbf{P}_i]_3 \mathbf{U}_\pi^{i(j)} + s_j, \\ & \qquad \mathbf{n}_i \cdot \mathbf{U}_\pi^{i(j)} + D_i = 0, \\ & \qquad \mathbf{U}_j = \mathbf{U}_\pi^{i(j)} + w^{i(j)} \mathbf{r}_b^{i(j)}, \\ & \qquad \forall i = 1, 2, \\ & \qquad \forall j = 1, \cdots, N. \end{aligned} \tag{10}$$

with unknown vector

$$\widetilde{\mathbf{X}} = \left(\mathbf{X}^\top, s_1, \cdots, s_N \right)^\top. \tag{11}$$

Again, for a fixed γ, the UKRP3 is convex and can be solved efficiently. The case $s_j > 0$ in the solution to UKRP3 indicates that the reprojection error of the j-th 3D point is larger than γ in at least one image, and thus can be identified as outlier.

4 Underwater Structure and Motion with Rotation Estimation

For two general underwater cameras, a minimal set of 11 parameters is required to model the two view geometry (intrinsic parameters are assumed to be known).

Since we assume that the image plane of each camera is nearly parallel to its refractive interface, the required number of parameters is reduced to 7, of which 5 are for the relative pose of the two cameras and 2 for the distances between the cameras and their refractive interfaces. From subsection 3.1, we know that the relative translation and the distance between each camera and its refractive interface can be optimally estimated. In this section, the algorithm proposed in subsection 3.1 is incorporated into Differential Evolution (DE), which is one of the most powerful population-based stochastic function minimizer [18], resulting in a new hybrid framework. Consequently, the underwater structure and motion problem is reduced to a small scale optimization problem over the rotation space, which can be concisely parameterized by only 4 parameters using quaternion.

4.1 Two-View Geometry Estimation Using Hybrid Optimization

Similar to many other evolutionary algorithms, DE maintains a population of N_p individuals. N_p new trial vectors are generated from the perturbation (scaled difference between two randomly selected population vectors) of points in the current generation. The trial vector competes against the population vector of the same index, and the vector with a better fitness value will be marked as a member of the next generation. In our problem, each individual $\mathbf{\Theta}$ is a 4-dimensional real-valued trial vector, which corresponds to a possible solution. Each individual in the initial population is randomly selected under uniform distribution in the rotation space. Without loss of generality, the coordinate system of the first camera is chosen to coincide with the world coordinate system. Given a trial vector $\mathbf{\Theta}$, the rotation matrices of the two cameras are given by $\mathbf{R}_1 = \mathbf{I}_{3\times3}$ (3×3 identity matrix) and $\mathbf{R}_2 = R_m(\mathbf{\Theta})$, where $R_m(.)$ transforms a quaternion into its equivalent rotation matrix. The normals of the two interfaces are given by $\mathbf{n}_1 = (0, 0, 1)^\top$ and $\mathbf{n}_2 = R_m^{-1}(\mathbf{\Theta})(0, 0, 1)^\top$.

Our proposed hybrid optimization consists of three stages. In the first stage, we search for the best camera rotation using DE [18]. In this stage, a subset of outlier free image correspondences are used and the individual evaluation for a given trial vector $\mathbf{\Theta}$ is performed as follows: first, retrieve the system parameters (camera rotation and interface normal) specified by the given trial vector; then, estimate the provably optimal structure and motion by solving UKRP2 (see subsection 3.1) and finally calculate the RMS reprojection error of reconstructed 3D scene as the fitness of $\mathbf{\Theta}$. In the second stage, we use all image correspondences (may contain outliers) and the best rotation estimated in the first stage to remove outliers and obtain robust estimates by solving the UKRP3 (see subsection 3.2). In the final stage, both system parameters and 3D structure are further refined by bundle adjustment as shown in the next subsection.

4.2 Sparse and Dense Underwater 3D Reconstruction

Given the rotation parameters and a set of outlier affected image correspondences, the sparse 3D structure and updated motion can be obtained by solving

the robust underwater known rotation problem UKRP3. We minimize the following objective function:

$$\mathcal{J} = \sum_{i=1}^{2} \sum_{j=1}^{N} [d'(\mathbf{P}_i, \mathbf{n}_i, D_i, \mathbf{U}_j, \mathbf{u}^{i(j)})]^2, \tag{12}$$

where $d'(\mathbf{P}_i, \mathbf{n}_i, D_i, \mathbf{U}_j, \mathbf{u}^{i(j)})$ is the reprojection error of the j-th 3D point $\mathbf{U}_j$ in the i-th image. The projection of a 3D point can be analytically computed using forward projection [3][9]. The objective function defined in Eq. (12) is a typical non-linear function and its scale can be large for 3D reconstruction problem. In this paper, we adopt a general purpose sparse Levenberg-Marquart (splm) algorithm [19] to improve the efficiency of optimization. For the dense 3D reconstruction, a modified version of the patch-based multi-view stereo (PMVS) algorithm [20][3] is used and it generates a (quasi) dense set of oriented patches covering the surface of scene, which can be converted into a mesh in a post processing stage.

5 Experiments

In order to evaluate the performance of our proposed method, we implemented the algorithms in C++ and carried out extensive experiments using both synthetic data and real images. The academic version of MOSEK [21] was used to solve the convex optimization problems. The refractive index of water is set to 1.33. In order to establish feature correspondences between two images, SIFT image features were detected and matched using the methods proposed in [22]. For the first stage of our proposed hybrid optimization, a subset of outlier free image correspondences are selected manually. The error metrics for quantitative evaluation are defined as follows: 1) the error in the relative camera rotation $\Delta\mathbf{R}$ is measured as the angle (in degrees) in the axis-angle representation of the rotation $\mathbf{R}_{est}\mathbf{R}_{gt}^{\top}$, where $\mathbf{R}_{gt}$ and $\mathbf{R}_{est}$ are the ground truth and the estimated relative camera rotation, respectively; 2) the error in the relative camera translation $\Delta\mathbf{t}$ is measured as the angle (in degrees) between the estimated relative camera translation $\mathbf{T}_{est}$ and the ground truth relative camera translation $\mathbf{T}_{gt}$; and 3) the error in the relative interface distance ΔD is measured as

$$\Delta D = \frac{1}{2}\left(\left|\frac{d_{est1}}{d_{gt1}} \cdot \frac{\|\mathbf{T}_{gt}\|}{\|\mathbf{T}_{est}\|} - 1\right| + \left|\frac{d_{est2}}{d_{gt2}} \cdot \frac{\|\mathbf{T}_{gt}\|}{\|\mathbf{T}_{est}\|} - 1\right|\right), \tag{13}$$

where d_{est1}, d_{est2} are the estimated distances between each camera and its refractive interface, and d_{gt1}, d_{gt2} are the corresponding ground truth distances. Since the magnitude of $\mathbf{T}_{est}$ cannot be recovered in metric 3D reconstruction, a scale factor $\frac{\|\mathbf{T}_{gt}\|}{\|\mathbf{T}_{est}\|}$ is used to normalize ΔD.

5.1 Synthetic Data

Our first set of experiments uses synthetic data, where a 3D scene consists of randomly generated 3D points within a unit cube. The two cameras were placed

two units away from the center of the cube, looking toward the center of the cube. Both the relative camera rotation and translation were randomly perturbed to generate various setups. The distance between each camera and its interface was randomly chosen from 0.2 units to 1 unit.

First, we evaluate the performance of the globally optimal structure and motion estimation algorithm described in subsection 3.1 using noise free data sets. Examined quantities are $\Delta\mathbf{T}$ and ΔD as defined earlier. Three data sets with a different number of 3D scene points are generated, each of which consists of 500 randomly generated instances. The statistical results using noise free data are shown in Fig. 2(a), which demonstrate that our proposed algorithm can estimate the camera and interface parameters accurately in the absence of noise, and that using more image correspondences improves the accuracy. Note that the accuracy of estimation can be further improved by specifying a smaller error tolerance in the bisection procedure. Next, we study the performance of the globally optimal algorithm under different amounts of noise. In addition to uniformly distributed noise under which it yields provably optimal estimation, the influence of Gaussian noise is also investigated. For each level of noise, 30 instances are analyzed statistically. Shown in Fig. 2(b) and Fig. 2(c) are the results of camera pose and interface parameter estimation under Uniform and Gaussian noise, respectively. Attributed to efficient convex programming solver provided in MOSEK, the running time is stable and increases approximately linearly with respect to the scale of problem. Specifically, it takes approximately 0.1 seconds for the synthetic scene consisting of 16 scene points and 0.5 seconds for 64 3D scene points.

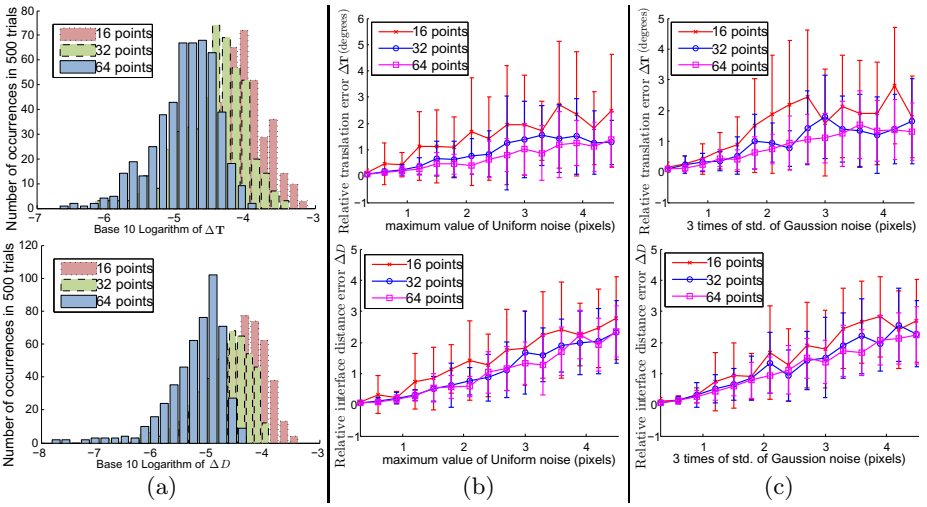

Fig. 2. Accuracy of parameter estimation (solution to UKRP2) (a) using noise free data sets, (b) under Uniform noise and (c) under Gaussian noise with a different number of 3D scene points

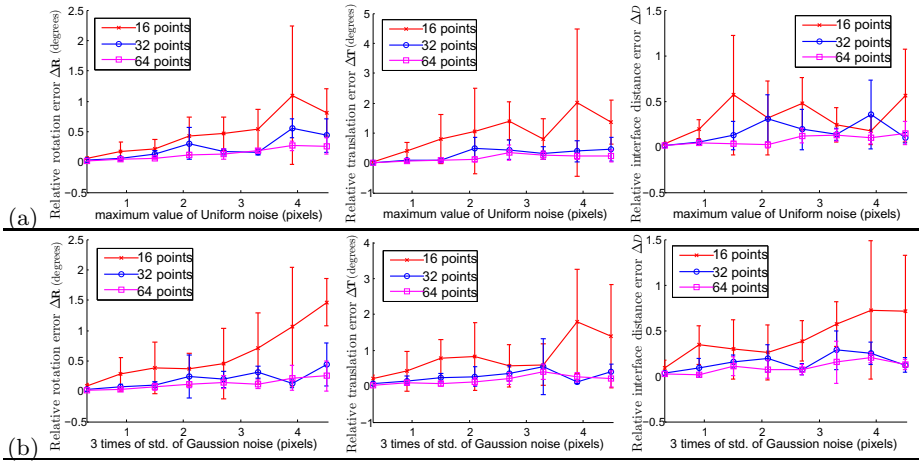

Fig. 3. Accuracy of parameter estimation using hybrid optimization for data sets (a) under Uniform noise and (b) under Gaussian noise

Then, we evaluate the performance of two-view geometry estimation using our proposed hybrid optimization described in subsection 4.1. Statistics over 20 randomly generated instances with a specified number of 3D points under each level of noise are shown in Fig. 3(a) (under Uniform noise) and Fig. 3(b) (under Gaussian noise). The results show that our proposed two-view geometry estimation method can obtain accurate estimation of camera pose and interface parameters. It is noteworthy that even though the solution to UKRP2 (see Fig. 2(b) and Fig. 2(c)) is provably optimal under Uniform noise, the proposed hybrid method can significantly improve the accuracy of camera and interface parameter estimation. The improvements indicate that hybrid optimization is more suitable for the refractive camera model which possesses highly intrinsic non-linearity. In addition, this set of experiments once again confirms that using more image correspondences can improve the accuracy of geometry estimation.

5.2 Synthetically Rendered Images

While subsection 5.1 presents the performance of proposed algorithms statistically, this subsection presents the comparison of results using synthetically rendered images. Since rendered images can be very realistic, they provide an

Table 1. Comparison of the accuracy of camera localization

Method	$\Delta\mathbf{R}$	$\Delta\mathbf{T}$
NDist	3.157	6.123
FAdj	3.044	3.438
RDist	2.236	6.179
FAdj+RDist	2.869	4.985
NEW	0.017	0.051

alternative way to evaluate practical performance of our proposed algorithms. In this experiment, we use POVRay, an publicly available ray tracer, to generate synthetically rendered images. In particular, we create a square water-filled pool of size 1.15×1×0.55 (L×W×H unit). The Stanford bunny model standing on an isosceles right angled prism is placed at the bottom of the pool. Each camera is placed in a glass housing deployed underwater to one side of the pool. The thickness of the glass is set to 0.01 units, the distance between each camera and its refractive interface is set to 0.1 units. The focal length of the camera is 800 pixels and the resolution of the image is 1024×768. Such settings result in a 65° horizontal field of view (FOV) of camera. The setup and two rendered images are shown in Fig. 4(a).

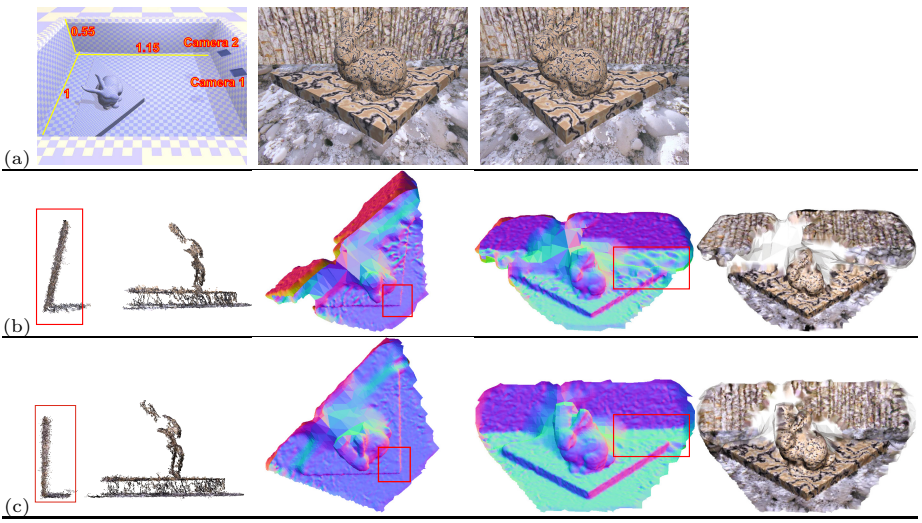

Fig. 4. Experiments with synthetically generated images. Figures in (a) are the simulated setup and two rendered images. Figures in (b) are the results of 3D reconstruction using FAdj+RDist. Figures in (c) are the results of 3D reconstruction using our proposed method. Interested regions are highlighted. Apparent distortion in 3D reconstruction includes: The wall and floor of the pool become non-perpendicular in the first column of (b), the two equal sides of the reconstructed isosceles right angled prism become non-perpendicular in the second column of (b), and the reconstructed scene far from the cameras is noisy in the third column of (b).

The performance of our proposed method (denoted as NEW) are compared with four cases: 1) simply ignore distortion (denoted as NDist); 2) use focal length adjustment (denoted as FAdj); 3) use lens radial distortion (denoted as RDist) and 4) simultaneously adjust focal length and lens radial distortion parameters (denoted as FAdj+RDist). All of the four compared cases are performed with bundler [23]. A comparison of the accuracy of camera localization is shown in Table 1. The results confirm that the conventional image-space distortion

models are incapable in compensating for the refraction effect. On the contrary, our proposed method can handle refractive distortion properly.

The presence of the refractive interface not only affect camera localization, but also results in distorted and incomplete 3D reconstruction. In particular, we found that apart from distortion in the reconstructed dense 3D scene, it fails to reconstruct consistent patches for the part of the scene far from the cameras, since the distortion induced by refraction increases as scene depth increases. For qualitative evaluation, the results of dense 3D reconstruction using FAdj+RDist and our proposed method are shown in Fig. 4(b) and Fig. 4(c), respectively.

5.3 Real Images

The practical performance of our proposed method has also been tested on real images. Two Point Grey Research Flea2 cameras were placed behind two planar faces of a large, water-filled tank. The optical axis of each camera was approximately parallel to the normal of its refractive interface and the intrinsic camera parameters were assumed to be known. In this imaging setup, the thickness of glass is about 6 mm, the distance between each camera and its refractive

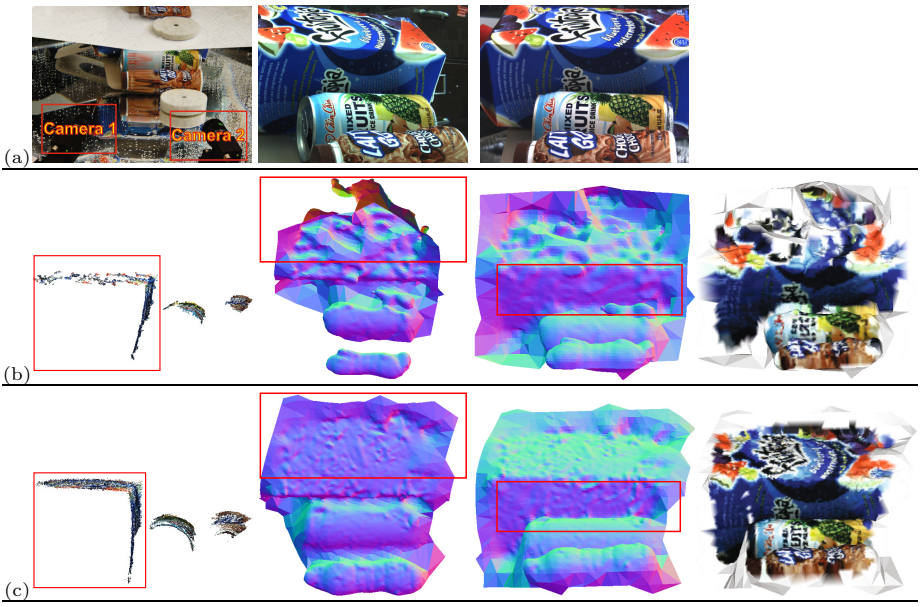

Fig. 5. Experiments with real images. Figures in (a) are the setup constructed in our lab and the two captured images. Figures in (b) are the results of 3D reconstruction using FAdj+RDist. Figures in (c) are the results of 3D reconstruction using our proposed method. Interested regions are highlighted. Apparent distortion in 3D reconstruction includes: The two sides of the box become non-perpendicular in the first column of (b), the reconstructed scene far from the cameras is noisy in the second and the third columns of (b).

interface is about 15 mm. The scene placed about 400 mm away from each camera contains three close together containers leaned on a bracket. The resolution of the captured images is 1032×776 pixels, the focal length of the camera is roughly 1800 pixels, and the horizontal field of view of the camera is about 32°. The experimental setup and captured images are shown in Fig. 5(a).

The results of dense 3D reconstruction using FAdj+RDist and our proposed methods are presented in Fig. 5(b) and Fig. 5(c) for qualitative evaluation. Compared with the results on synthetically rendered images (see Fig. 4), the 3D reconstruction on real images is less accurate due to a large amount noise in image measurements. Nevertheless, more complete and accurate surface has been obtained using our proposed method than that using FAdj+RDist, which demonstrates the superiority of our method.

6 Conclusions and Discussions

This paper proposes a method to perform structure and motion from two images captured by a general underwater imaging setup consisting of two cameras, of which each camera is placed in a separate waterproof housing with a flat window. A new formulation of the underwater known rotation problem for the refractive camera model is proposed based on two new concepts, namely Ellipse of Refrax (EoR) and Refractive Depth (RD). The proposed formulation allows one to obtain provably optimal underwater structure and motion under L_∞-norm given known rotation. The constraint of known rotation is further relaxed in a new hybrid optimization framework. Promising results on synthetic data, synthetically rendered images and real images demonstrate that the proposed method can significantly improve the accuracy of camera motion and 3D structure estimation for underwater applications.

Two simplification of the underwater imaging setup were made in this paper. First, the thickness of refractive was ignored. In fact, as pointed out by Treibitz et al. [8], the thickness of the glass interface results in negligible shift in the image because the thickness of interface is normally small and the refractive indices of glass and water are close. Second, the image plane of each camera was assumed to be nearly parallel to its refractive interface, as in most real underwater imaging system. Thus, both assumptions are reasonable for practical underwater applications. As for future work, it would be interesting to investigate scenarios where the refractive indices of media are unknown.

Acknowledgments. This work was partially supported by the Chinese Scholarship Council (grant no. 2010611068), the Hunan Provincial Innovation Foundation for Postgraduate (grant no. CX2010B025), the Natural Sciences and Engineering Research Council of Canada (NSERC) and the University of Alberta. The authors would like to thank the three anonymous reviewers for their constructive comments.

References

1. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, New York (2004)
2. Chari, V., Sturm, P.: Multiple-view geometry of the refractive plane. In: BMVC (2009)
3. Chang, Y., Chen, T.: Multi-view 3d reconstruction for scenes under the refractive plane with known vertical direction. In: ICCV (2011)
4. Queiroz-Neto, J.P., Carceroni, R., Barros, W., Campos, M.: Underwater stereo. In: CGIP, XVII Brazilian Symposium (2004)
5. Ferreira, R., Costeira, J.P., Santos, J.A.: Stereo Reconstruction of a Submerged Scene. In: Marques, J.S., Pérez de la Blanca, N., Pina, P. (eds.) IbPRIA 2005. LNCS, vol. 3522, pp. 102–109. Springer, Heidelberg (2005)
6. Pizarro, O., Eustice, R., Singh, H.: Relative pose estimation for instrumented, calibrated imaging platforms. In: DICTA (2003)
7. Lavest, J.M., Rives, G., Lapresté, J.T.: Underwater Camera Calibration. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1843, pp. 654–668. Springer, Heidelberg (2000)
8. Treibitz, T., Schechner, Y.Y., Singh, H.: Flat refractive geometry. In: CVPR (2008)
9. Agrawal, A., Ramalingam, S., Taguchi, Y., Chari, V.: A theory of multi-layer flat refractive geometry. In: CVPR (2012)
10. Sedlazeck, A., Koch, R.: Calibration of housing parameters for underwater stereo-camera rigs. In: BMVC (2011)
11. Telem, G., Filin, S.: Photogrammetric modeling of underwater environments. ISPRS Journal of Photogrammetry and Remote Sensing 65(5), 433–444 (2010)
12. Kunz, C., Singh, H.: Hemispherical refraction and camera calibration in underwater vision. In: OCEANS (2008)
13. Glaeser, G., Schrocker, H.P.: Reflections on refractions. Journal for Grometry and Graphics 4, 1–18 (2000)
14. Glassner, A.S.: An Introduction to Ray Tracing. Academic Press Ltd., London (1989)
15. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, New York (2004)
16. Kahl, F., Hartley, R., Member, S.: Multiple-view geometry under the L_∞-norm. IEEE TPAMI 30(9), 1603–1617 (2008)
17. Olsson, C., Eriksson, A., Hartley, R.: Outlier removal using duality. In: CVPR (2010)
18. Price, K., Storn, R.M., Lampinen, J.A.: Differential Evolution: A Practical Approach to Global Optimization. Springer-Verlag New York, Inc., Secaucus (2005)
19. Lourakis, M.I.A.: Sparse Non-linear Least Squares Optimization for Geometric Vision. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part II. LNCS, vol. 6312, pp. 43–56. Springer, Heidelberg (2010)
20. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multi-view stereopsis. In: CVPR (2007)
21. MOSEK, `http://www.mosek.com/`
22. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV 60, 91–110 (2004)
23. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: exploring photo collections in 3D. In: ACM SIGGRAPH (2006)