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

Calibration, Positioning and Tracking in a Refractive and Reflective Scene

Tobias Palmér
Center for Mathematical Sciences
Lund University
Lund, Sweden
Email: tobiasp@maths.lth.se

Giuseppe Bianco, Mikael T. Ekvall, Lars-Anders Hansson
Aquatic ecology, Department of Biology
Lund University
Email: giuseppe.bianco@biol.lu.se
mikael.ekvall@biol.lu.se
lars-anders.hansson@biol.lu.se

Kalle Åström
Center for Mathematical Sciences
Lund University
Lund, Sweden
Email: kalle@maths.lth.se

***Abstract*—We propose a framework for calibration, positioning and tracking in a scene viewed by multiple cameras, through a flat refractive surface and one or several flat reflective walls. Refractions are explicitly modeled by Snell's law and reflections are handled using virtual points. A novel bundle adjustment framework is introduced for solving the nonlinear equations of refractions and the linear equations of reflections, which in addition enables optimization for calibration and positioning. The numerical accuracy of the solutions is investigated on synthetic data, and the influence of noise in image points for several settings of refractive and reflective planes is presented. The performance of the framework is evaluated on real data and confirms the validity of the physical model. Examples of how to use the framework to back-project image coordinates, forward-project scene points and estimate the refractive and reflective planes are presented. Lastly, an application of the system on real data from a biological experiment on small aquatic organisms is presented.**

I. Introduction

There has been a lot of interest in the computer vision society on the subject of using and modeling cameras whose line of sight is somehow disturbed by refractions [1]–[5]. There are many different problems that can be posed by assuming different configurations of known relationships between refractive planes, cameras, scene points etc. In the case where the presence of refractions is due to the camera being under water, there are additional radiometric effects that alters the imagery [6]–[9]. Modeling of refractions has also been combined with reflections. For example, the recovery of refractive and reflective objects has been studied by light-path triangulation [10], [11], assuming known scene points. Furthermore, the problem of reconstruction or pose estimation has been studied in the case of flat reflections [12] and catadioptric cameras with parabolic reflections [13], [14].

However, positioning and tracking of objects within a scene where there are both refractions and reflections present has to our knowledge not been considered in previous research. This is an important application in for example biological sciences, where it is common to study the behavior of objects (e.g. animals) inside of an aquarium while viewing the scene from the outside [15]–[17]. An equivalent problem is to estimate structure inside a refractive (and possibly reflective) scene. We therefore present theory and methods for the calibration of refractive and reflective scene parameters, assuming pre-calibrated cameras. We describe the subsequent use of methods for finding the positions of objects within the scene. A framework is created for this purpose and is applied for a study on the behavior of small aquatic organisms.

In Section II, we explain how refractions and reflections are modeled, using Snell's law and virtual points. In Section III, we show how the models are applied to create a residual vector and how to minimize it efficiently by bundle adjustment. Examples of how to use the framework to forward-project a scene point (Section III-D), back-project an image point (Section III-D) and estimate the refractive and reflective planes (Section IV) are presented. The numerical accuracy of the method is thoroughly tested on synthetic data in Section V-A, and the validity of the framework on real data is tested using images of a calibration pattern submerged in an aquarium in Section V-B. Lastly, an actual application of the system is presented in Section VI, providing a qualitative evaluation of the system on real data.

A. Related work

In [3], Agrawal et al. presents theory of flat refractive geometry, deriving e.g. that a pinhole camera viewing a scene through (multiple) parallel flat refractive surfaces corresponds to an axial camera. It is also shown that the case of double refractions in the case of air-glass-water, where the glass is thin, is well approximated by disregarding the glass. This result is also reported in [4].

A method for structure-from-motion in underwater settings where the camera-refractive plane pose is known is presented by Jordt-Sedlazeck and Koch in [1], deriving an error function for virtual cameras for use in bundle adjustment. Here, the standard methods for bundle adjustment in refractive structure-from-motion are deemed infeasible.

The problem of estimating the absolute pose of a calibrated camera viewing a refractive scene is solved minimally and near-minimally for a few different settings of assumptions, by Haner and Åström [4].

Concerning reflections and ray-tracing, Sturm and Bonfort considered the task of computing the pose of an object without a direct view [12], viewing a calibration pattern through a flat reflective surface. In [11], Chari and Sturm use both reflections and refractions for estimating the structure of transparent refractive objects. In the case of catadioptric cameras, i.e. cameras

978-1-5090-4846-5/16/$31.00 ©2016 IEEE

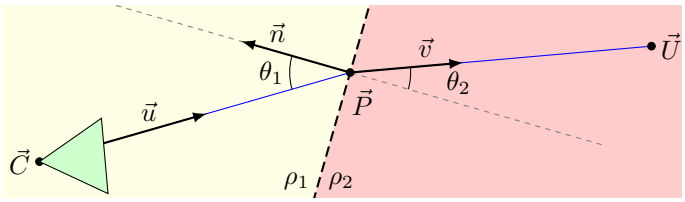

Fig. 1. Snell's law. A ray originating from the camera center C with direction $\vec{u}$ undergoes a change of direction according to $\rho_1 \sin\theta_1 = \rho_2 \sin\theta_2$. This causes the usually linear equations for projections, for example, to become nonlinear and much harder to solve.

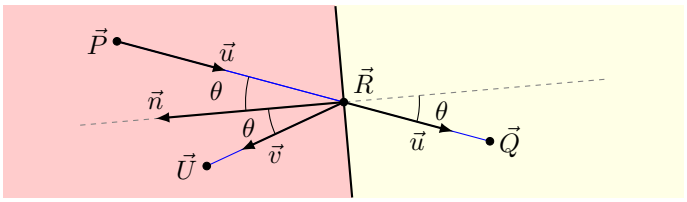

Fig. 2. Reflections. A ray with direction $\vec{u}$ originating from a point $\vec{P}$ intersects a reflective plane $(\vec{n}, d)$ at a point $\vec{R}$, where the ray changes direction to $\vec{v}$ and subsequently intersects a point $\vec{U}$. Note that the ray segment from $\vec{R}$ to $\vec{U}$ is the mirror image of the virtual ray segment from $\vec{R}$ to $\vec{Q}$.

with parabolic mirrors, Geyer and Daniilidis derive epipolar constraints and use them for structure and motion in [13], [14].

In the field of biology, Ekvall et al. studied the movement of small aquatic organisms [15] inside a rectangular aquarium using four cameras, positioned pairwise and orthogonal to two of the surfaces and approximating the cameras as orthogonal.

II. Theory of Refractions and Reflections

In this section, we introduce the equations for modeling refractions and reflections that are later used for defining the error functions in Section III-B. Note that some indices of variables, later introduced in Section III-A, are omitted here.

A. Refractions (Snell's law)

Snell's law describes how the angle θ_1 of an impinging ray $\vec{u}$, with respect to the refractive plane normal, is related to the angle θ_2 of the outgoing ray $\vec{v}$ by

$$\rho_1 \sin(\theta_1) = \rho_2 \sin(\theta_2), \tag{1}$$

where ρ_1 and ρ_2 are the refractive indices of the surfaces. A visualization of this is given in Fig 1. This nonlinear relation is the main source of the challenges encountered in underwater structure-from-motion. Another issue is ambiguity in Eq (1) that gives rise to false solutions.

In order to express Snell's law as a function of the ray directions and plane normal, a property of the cross product is applied:

$$\vec{u} \times \vec{n} = \|\vec{u}\|\|\vec{n}\| \sin(\theta_1)\vec{w}, \tag{2}$$

where $\vec{w}$ is a unit vector orthogonal to both $\vec{u}$ and $\vec{n}$, and θ_1 is the angle between $\vec{u}$ and $\vec{n}$. Since any vector orthogonal to $\vec{u}$ and $\vec{n}$ is also orthogonal to $\vec{n}$, Eq (2) can be used to reformulate Eq (1) to

$$\rho_1 \frac{\vec{u} \times \vec{n}}{\|\vec{u}\|\|\vec{n}\|} = \rho_2 \frac{\vec{v} \times \vec{n}}{\|\vec{v}\|\|\vec{n}\|}, \tag{3}$$

where $\vec{u}$ and $\vec{v}$ are the previously defined ray directions. Assuming that $\|\vec{n}\| = 1$ and multiplying by $\|\vec{u}\|$ and $\|\vec{v}\|$ gives the expression

$$\rho_1 \|\vec{v}\|(\vec{u} \times \vec{n}) = \rho_2 \|\vec{u}\|(\vec{v} \times \vec{n}). \tag{4}$$

Squaring both sides of the equation element-wise gives expressions that are polynomial in all variables, of which only two are independent (ref e.g. [4]). Solutions to this equation will later be sought for the purpose of computing forward projections.

Note that in the case where $\vec{u}$ and $\vec{n}$ are known, solving for $\vec{v}$ gives the expression

$$\vec{v} = r\vec{u} + (r\cos\theta_1 - \text{sign}(\cos\theta_1)\cos(\theta_2))\vec{n}, \tag{5}$$

where

$$\begin{aligned} \cos\theta_1 &= -\vec{n} \cdot \vec{u}, \\ \cos\theta_2 &= \sqrt{1 - r^2(1 - \cos^2\theta_1)}, \end{aligned} \tag{6}$$

and $r = \rho_1/\rho_2$. This provides a convenient method for computing the refraction of directions and will later be used for back-projections.

B. Reflections

The law of reflection states that the angle θ_1 of an impinging ray is related to the angle θ_2 of the outgoing ray as

$$\theta_1 = \theta_2, \tag{7}$$

with respect to the normal of the reflective surface at the point of reflection, as depicted in Fig 2. This relation can be reformulated on vector form as a linear transformation from the incoming ray direction $\vec{v}$ to the outgoing ray direction $\vec{w}$ as

$$\vec{w} = (\vec{I} - 2\vec{n}\vec{n}^\top)\vec{v}, \tag{8}$$

where $\vec{n}$ is the normal direction of the surface at the point of intersection.

An equivalent way to model reflections is to reflect points instead of directions. The observation of a point $\vec{U}$ that is reflected in a surface appears the same as the observation of a virtual point $\vec{Q}$ that is the mirror image of $\vec{U}$ and travels straight through the surface. Adapting Eq 8 for points gives the following linear transformation for the reflection of a point $\vec{U}$ in $\vec{\pi}$:

$$\begin{bmatrix} \vec{Q} \\ 1 \end{bmatrix} = \begin{bmatrix} (\vec{I} - 2\vec{n}\vec{n}^\top) & -2\vec{n}d \\ \vec{0}^\top & 1 \end{bmatrix} \begin{bmatrix} \vec{U} \\ 1 \end{bmatrix}. \tag{9}$$

This property of modeling the reflection of rays provides a convenient method for computing forward projections through one or multiple reflections.

III. Method

In this section, we describe how the theory of reflections and refractions introduced in Section II can be applied. First the notation is introduced, then the equations of refractions and reflections are adapted for use in a residual vector, and lastly initialization and practical application is presented.

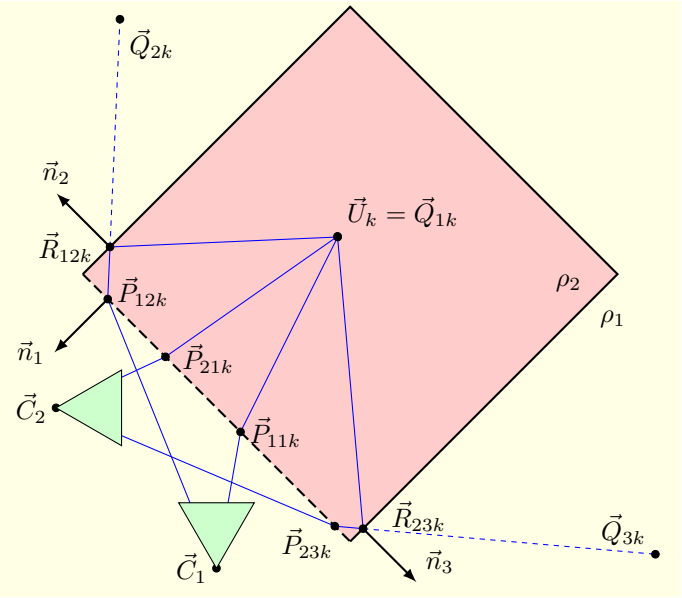

Fig. 3. The refraction and reflections of two pairs of rays from two camera focal points $\vec{C}_i$, all corresponding to the same scene point $\vec{U}_k$. All of the rays intersect the plane of refraction $(\vec{n}_1, d_1)$ where they change direction, resulting in second linear pieces of the ray originating at the points of refraction $\vec{P}_{ijk}$. Two of the rays also intersect a plane of reflection $(\vec{n}_j, d_j)$, $j > 1$, resulting in additional linear pieces of the ray originating at the points of reflection $\vec{R}_{ijk}$. In the case of forward projection, computing the points of reflection is conveniently avoided by instead forward projecting the virtual point $\vec{Q}_{jk}$ that is the reflection of the scene point in zero, one or several planes of reflection.

A. Data structure and notation

The modeling of a ray subject to refraction and reflection is based on the use of a *help point* at the refractive plane and a *virtual point* that is the reflected scene point, as shown in Fig 3. Scene point k is denoted $\vec{U}_k$ and the reflection of $\vec{U}_k$ in *reflective plane configuration* $j \in \mathbb{N}$ is denoted $\vec{Q}_{jk}$. A reflective plane configuration can be for example reflection in $\vec{\pi}_2$, followed by reflection in $\vec{\pi}_4$. By defining $j = 1$ as the empty reflection, $\vec{Q}_{jk}$ denotes an *end point* that is either the scene point for $j = 1$ or a virtual point for $j > 1$. The observation of an end point by camera i corresponds to a ray in two linear pieces - from $\vec{Q}_{jk}$ to $\vec{P}_{ijk}$ and from $\vec{P}_{ijk}$ to $\vec{C}_i$, where $\vec{C}_i$ is the focal point and $\vec{P}_{ijk}$ is a *help point*. The *help point* $\vec{P}_{ijk}$ lies on the refractive plane and is defined such that the angles of the vectors $\vec{u} = \vec{P}_{ijk} - \vec{C}_i$ and $\vec{v} = \vec{Q}_{jk} - \vec{P}_{ijk}$ satisfies Eq (4).

A ray corresponding to the observation of an *end point* $\vec{Q}_{jk}$ by camera i is represented by the points it interpolates as

$$\left(\vec{C}_i, \vec{P}_{ijk}, \vec{Q}_{jk}\right), \tag{10}$$

where $\vec{P}_{ijk}$ is the *help point*. The normalized image coordinates in camera i relating to an end point $\vec{Q}_{jk}$ are denoted $\vec{u}_{ijk}$ and are given by the projection of the help point $\vec{P}_{ijk}$.

Note that in the notation used throughout the paper, it is assumed that all scene points are subject to a single refraction and that the plane of refraction separates the end points from the camera focal points.

B. Bundle adjustment

Equations describing refractions and reflections, as introduced in Section II-A and Section II-B, together with the data structures introduced in Section III-A are used to create a residual vector $\vec{r}$.

The error function for the refraction of a ray $(\vec{C}_i, \vec{P}_{ijk}, \vec{Q}_{jk})$ is adapted from Eq (4) and defined as

$$\begin{aligned}\vec{r}_{refr}(\vec{C}_i,\vec{P}_{ijk},\vec{Q}_{jk},\vec{\mu},\vec{n}_1) = \\ \mu_1^2 \left(\left(\vec{C}_i - \vec{P}_{ijk}\right) \times \vec{n}_1\right)^2 \left\|\vec{Q}_{jk} - \vec{P}_{ijk}\right\|_2^2 \\ - \mu_2^2 \left(\left(\vec{Q}_{jk} - \vec{P}_{ijk}\right) \times \vec{n}_1\right)^2 \left\|\vec{C}_i - \vec{P}_{ijk}\right\|_2^2,\end{aligned} \tag{11}$$

where the square of the cross product is computed element-wise. The refractive error function is complemented by an error function constraining the help points to lie on the plane of refraction.

The error function for the reflection of a scene point $\vec{U}_k$ to a virtual point $\vec{Q}_{jk}$ is adapted from Eq (9) as

$$\begin{aligned}\vec{r}_{refl}(\vec{U}_k,\vec{Q}_{jk},\vec{\pi}_j) = \\ \left(\vec{I} - 2\vec{n}_j\vec{n}_j^\top\right)\vec{Q}_{jk} - \vec{U}_k - 2\vec{n}_j d_j.\end{aligned} \tag{12}$$

Also, error functions for the re-projections of help points $\vec{P}_{ijk}$ to image points $\vec{u}_{ijk}$, $\vec{r}_{proj}(\vec{P}_{ijk}, \vec{u}_{ijk})$, and constraints $\|\vec{n}_j\| = 1$ for the plane normal vectors $\vec{n}_j$, $\vec{r}_n(\vec{n}_j)$, are added.

C. Using the Schur complement

Given an end point $\vec{Q}_{jk}$ and a camera center $\vec{C}_i$, there is always a help point $\vec{P}_{ijk}$ such that the ray $(\vec{C}_i, \vec{P}_{ijk}, \vec{Q}_{jk})$ satisfies Snell's law. Finding the help points is computationally very cheap - each help point is subject to four constraints and only a few iterations using Gauss-Newton's method are enough to find it accurately. Furthermore, all help points are independent from other help points. The existence of $\vec{P}_{ijk}$ and the computational efficiency in estimation can be leveraged to increase effiency in the estimation of other variables as follows.

First, the residual vector and variables are partitioned as

$$\begin{aligned}\vec{r} &= \begin{bmatrix}\vec{r}_1 & \vec{r}_2\end{bmatrix}^\top, \\ \vec{x} &= \begin{bmatrix}\vec{x}_A & \vec{x}_B\end{bmatrix}^\top,\end{aligned} \tag{13}$$

where $\vec{r}_1$ are the residuals for refractions and reflections, $\vec{r}_2$ the residuals for projections and plane normalization, $\vec{x}_A$ is the vector containing help points and $\vec{x}_B$ contains some other disjoint subset of $\vec{x}$, for example the reflective plane parameters. The linearization of the problem at $\vec{x}$ is

$$\begin{cases}\delta\vec{r}_1 = \vec{J}_{1A}\delta\vec{x}_A + \vec{J}_{1B}\delta\vec{x}_B, \\ \delta\vec{r}_2 = \vec{J}_{2A}\delta\vec{x}_A + \vec{J}_{2B}\delta\vec{x}_B.\end{cases} \tag{14}$$

where $\vec{J}_{1A}$ is the matrix of partial derivatives of $\vec{r}_1$ with respect to $\vec{x}_A$ and $\vec{J}_{1B}$, $\vec{J}_{2A}$, $\vec{J}_{2B}$ are defined analogously. First solving for help points with respect to refractions and reflections gives (approximately) $\|\vec{r}_1\| = 0$. Inserting $\vec{r}_1$ into Eq (14) and solving the first line for $\delta\vec{x}_A$ gives $\delta\vec{x}_A$ as a function of $\delta\vec{x}_B$. Inserting the solution $\delta\vec{x}_A$ into the second equation, and solving the second equation for $\delta\vec{x}_B$ results in

$$\begin{cases} \delta\vec{x}_A = -\vec{J}_{1A}^{-1}\vec{J}_{1B}\delta\vec{x}_B, \\ \delta\vec{x}_B = -(\vec{J}_{2A}\vec{J}_{1A}^{-1}\vec{J}_{1B} + \vec{J}_{2B})^{-1}\delta\vec{r}_2, \end{cases} \quad (15)$$

where $\vec{J}_{2A}\vec{J}_{1A}^{-1}\vec{J}_{1B} + \vec{J}_{2B}$ is known as the Schur-complement. This is applied in the estimation of e.g. the refractive plane, by first optimizing the help points, then optimizing in the direction $\delta\vec{x}$ obtained by solving Eq (15) for $\delta\vec{x}_A$ and $\delta\vec{x}_B$. The procedure is repeated from the newly obtained $\vec{x}$ until some stopping criteria is met.

D. Computing the back-projections and forward projections

The bundle adjustment methods assumes that there is an initial value for each variable, thus there is need for methods of initializing those variables. In this section, initialization for forward projections and backward projections are explained.

In the case of forward-projections, a scene point $\vec{U}_k$ is known and the set of image points $\vec{u}_{ijk}$ that corresponds to $\vec{U}_k$ are sought. This amounts to finding and then projecting the point of refraction $\vec{P}_{ijk}$ for any end point, i.e. for the scene point or any reflection $\vec{Q}_{jk}$ of it. For this purpose, the scene point is reflected in some set of plane configurations, resulting in the set of end points $\{\vec{Q}_{jk}\}$. For each such end point $\vec{Q}_{jk}$ and camera center $\vec{C}_i$, an initial value for the corresponding help point $\vec{P}_{ijk}$ is given by the intersection of the straight line from $\vec{Q}_{jk}$ to $\vec{C}_i$ and the refractive plane $\vec{\pi}_1$. For each camera i and plane configuration j, the error of refraction for the ray $\vec{r}_{refr}(\vec{C}_i, \vec{P}_{ijk}, \vec{Q}_{jk}, \vec{\mu}, \vec{n}_1)$ is subsequently minimized with respect to the help points $\vec{P}_{ijk}$. The help points are then projected by their corresponding camera matrix to provide the sought image coordinates $\vec{u}_{ijk}$.

In the case of back-projecting an image point, assuming that the refractive and reflective planes are known, the help point and refracted ray direction can be obtained by tracing the ray from the camera focal point to the plane of refraction and computing the new direction by Eq (5). The process is analogous for the reflective planes using Eq (8). Note that the bundle adjustment framework is not used for this task.

In the case of back-projecting a set of image points $\vec{u}_{ijk}$ corresponding to the same scene point $\vec{U}_k$, assuming knowledge of the reflective plane configuration j each ray undergoes, the scene point $\vec{U}_k$ is sought. Help points $\vec{P}_{ijk}$ and refracted and reflected ray directions are computed as was presented in the case of single back-projection. This results in a set of points and directions whose intersection is used as an initial value for the scene point $\vec{U}_k$. Note that in the case where there are at least two rays unaffected by reflections, the initial value is more conveniently computed using only those rays. The virtual points are then initialized by reflecting $\vec{U}_k$ in the specified reflective plane configurations using Eq (9), Thus the ray $(\vec{C}_i, \vec{P}_{ijk}, \vec{Q}_{jk})$ is known for each end point. The residual vector for this estimation is composed of $\vec{r}_{refl}(\vec{U}_k, \vec{Q}_{jk}, \vec{\pi}_j)$, $\vec{r}_{refr}(\vec{C}_i, \vec{P}_{ijk}, \vec{Q}_{jk}, \vec{\mu}, \vec{n}_1)$ and $\vec{r}_{proj}(\vec{P}_{ijk}, \vec{u}_{ijk})$, and is subsequently minimized with respect to the virtual points $\vec{Q}_{jk}$ and scene points $\vec{U}_k$.

An issue in tracking applications is that the configuration of reflections that each ray undergoes on its way to the end point can not always be assumed to be known. However, the above methods can be combined to create a method that can be used to find the configurations. Given a putative pair of image point correspondences and assume that their corresponding rays do not undergo reflections before meeting their common end point. Then the end point can be estimated by back-projection as previously explained. If the estimated end point is located outside of the scene (e.g. in the case of an aquarium), the estimate is discarded. A set of reprojected points are given by reflecting the end point in several reflection configurations and forward projecting each of the resulting virtual points. Each such forward projection is feasible only if the help point is located inside of the bounded refractive plane. Lastly, the distances to detected image points are evaluated and points sufficiently close to the re-projections are assumed to correspond to the same scene point. If additional image point correspondences are found, the position of the scene point can subsequently be further optimized.

IV. Calibrating the scene

The calibration process is divided in two steps - pre-calibration and refractive scene calibration. The pre-calibration is, in this case, composed of estimating the intrinsic parameters of each camera and the relative poses of the cameras, i.e. radial distortion, skew, focal length, etc. and position and orientation. This is done using the standard methods with a moving checkerboard [18], [19].

The refractive scene calibration consists of estimating the parameters of the planes of refraction and reflections in the scene, i.e. $(\pi_1, \pi_2, ...)$, and is done using a set of observed image points. It is assumed that the plane configurations that each of the corresponding rays are reflected in before they meet their scene points are known, and that there are sufficiently small errors in the initial values of the scene parameters. The method for back-projection explained in Section III-D is used to find a ray $(\vec{C}_i, \vec{P}_{ijk}, \vec{Q}_{jk})$ for each image point. Each ray is entered into the framework and the corresponding residual vector of refractions, reflections and projections is minimized with respect to the help points, end points and planes. The estimated planes are the results of calibration.

In the experimental evaluation (section V-B), the refractive scene calibration was performed using a checkerboard submerged in a water tank and the checkerboard points were found using the same checkerboard-detection algorithms as for the pre-calibration. In this case, heuristics for deciding the refractive/reflective configuration that each ray undergoes could be easily decided due to the scene-geometry of this particular application. A close-enough initial calibration was attained by manually selecting the points in the corners of the tank for which the geometry was known.

V. Experiments

The performance of the system is evaluated on both real data and synthetic data. The accuracy and robustness in estimates of image points, scene points and refractive and reflective

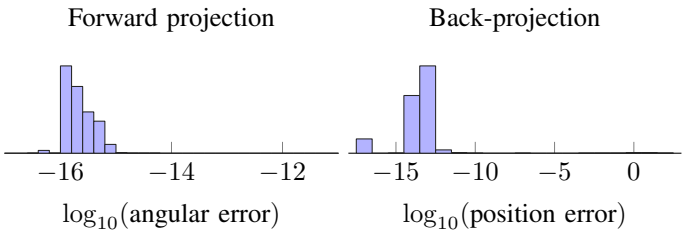

Fig. 4. Left: the distribution of angular errors with respect to Snell's law for forward-projection in 1000 random synthetic problem instances with no added noise. Right: The distribution of error in position, defined as the euclidean distance from an estimate to ground truth, for back-projection in 1000 random synthetic problem instances with no added noise. Points were simulated at distances approximately 200 mm from the camera focal points.

planes is analyzed in Section V-A. Real images are analyzed in Section V-B, where a checkerboard is submerged in an aquarium.

A. Synthetic experiments

The method for forward-projection is evaluated by simulating 1000 random scenes, each consisting of a refractive plane, a reflective plane and two camera poses. Scene points are generated and subsequently forward-projected to image points in the cameras, as explained in Section III-D. The difference in θ_2 given by solving $\|\vec{n}\|\|\vec{v}\|\cos(\theta_2) = \vec{n}^\top\vec{v}$ and Eq (1) for θ_2, using the estimated θ_1, is evaluated and presented in Fig 4. This shows that the forward projections of the synthetic data are accurate, thereby proving the validity of the synthetic data that the following analyses are based on.

The numerical accuracy of the method for back-projection (introduced in Section III-D) is tested by back-projecting simulated image points and estimating the 3d positions that corresponding pairs of image points are observations of. The estimated 3d points are compared to the ground truth 3d points, and the distribution of distances between the points is presented in Fig 4. Sensitivity to noise was tested by adding zero mean normal distribution noise with increasing variance to the image points, resulting in the distribution of distances shown in Fig 5.

The numerical accuracy of the method for estimating refractive and reflective planes, as introduced in Section IV, is tested on 1000 random simulated problem instances. Each instance consists of the observed image coordinates of 3 scene points (the minimal case) in a fixed, simulated scene with two cameras and planes similar to the real scene analyzed in Section VI. The performance of the plane estimation method is tested by perturbing the initial plane parameters slightly and using the image points to estimate the planes. Fig 5 presents the results of plane estimation with respect to estimating the position of the refractive plane, where the position of the plane is defined as the point at the center of the plane, computed using the ground truth planes for all other planes. Additionally, sensitivity to noise in image coordinates is tested by adding zero-mean normal distributed noise to the image coordinates and subsequently estimating the planes. The result of estimating the refractive plane position as a function of added noise variance is presented in Fig 6.

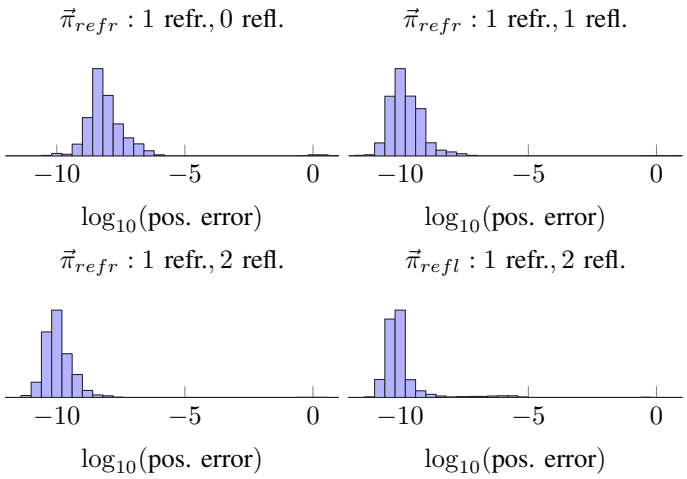

Fig. 5. Distribution of errors in estimated plane positions compared to ground truth for three scene settings, given slightly perturbed initial plane parameters. Note that the distributions in the top row and the left column shows the errors in position of the refractive plane, and the bottom right shows the distribution of errors in position of both reflective planes.

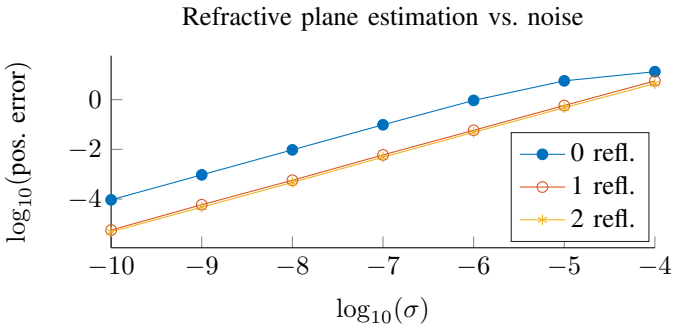

Fig. 6. The median of error in position of the refractive plane for three scene settings as a function of the noise variance σ^2. For each σ, zero-mean Gaussian noise with variance σ^2 is added to the normalized image coordinates, and the planes are estimated given slightly perturbed initial plane parameters. Note that the difference between one and two reflective planes is small.

B. Real checkerboard experiments

The setup presented in Section VI was used for testing the performance of the method on real data. A checkerboard of known dimensions was submerged in the water in an aquarium overlooked by four cameras, as presented in Fig 8, in order to evaluate the performance of the system. The checkerboard was moved in a range of approximately 200 mm to 400 mm from the cameras. The cameras have been pre-calibrated and initial estimates of the planes of the aquarium are used. The method of optimizing the refractive and reflective planes, presented in Section IV and synthetically evaluated in Section V-A, was applied to observed image points from the scene.

The re-projection errors obtained during the estimation are presented in Fig 7. In addition, the relative positions of the estimated scene points are analyzed as follows. Since the absolute pose of the checkerboard is unknown, the estimated scene points are transformed by an euclidean transformation to the local coordinate system of the checkerboard. The distribution

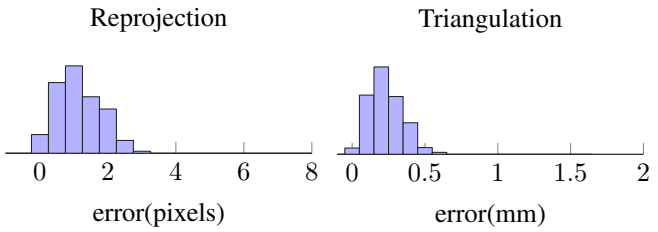

Fig. 7. The distribution of re-projection errors (left) and estimated scene points position errors (right) for a real experiment using a checkerboard.

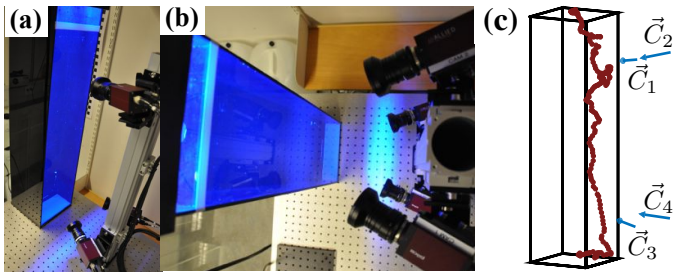

Fig. 8. (a) and (b): Side and top view of the experimental setup, composed of an aquarium and four cameras arranged as vertical stacked stereo-pairs that overlook the entire aquarium from the side. (c): The track of one *Daphnia*, the model of the aquarium and the four cameras.

of distances between the transformed estimated scene points and the checkerboard points is presented in Fig 7.

VI. Application - Tracking Small Aquatic Organisms

The proposed system is applied for a biological experiment on *Daphnias* (small aquatic organisms), where the behavioral response to e.g. ultraviolet radiation is studied. A setup consisting of an aquarium overlooked by four cameras arranged as vertical stacked stereo-pairs was used to acquire videos, as depicted in Fig 8. The cameras were pre-calibrated and the refractive plane was estimated, as described in Section IV. Note that since the cameras approximately lie in a plane which is approximately parallel to the refractive surface, a good initial estimate of the refractive surface normal can be made. Measuring the distance from the refractive plane to the cameras by ruler enables initiation of the fourth parameter of the refractive plane. Since the aquarium is rectangular and distances between parallel surfaces are known, initial values of all planes can be made with sufficient precision.

Videos of daphnias were recorded and the daphnias were subsequently detected. The daphnias' image points were input into the framework and their corresponding 3-d positions were estimated as explained in Section III-D. The track of one daphnia is shown in Fig 8.

VII. Conclusion

We have presented a method that can be used for the tracking of objects within a scene where both flat refractions and flat reflections are present. Evaluations on both synthetic data and real data have confirmed that the information that is found in the reflections can be used to improve the performance of estimating the refractive and reflective planes and the position of objects in the scene. Lastly, we have presented an application of the system that produces qualitatively good results.

References

[1] A. Jordt-Sedlazeck and R. Koch, "Refractive structure-from-motion on underwater images," in *The IEEE International Conference on Computer Vision (ICCV)*, December 2013.

[2] A. Sedlazeck and R. Koch, "Perspective and non-perspective camera models in underwater imaging–overview and error analysis," in *Outdoor and Large-Scale Real-World Scene Analysis*. Springer, 2012, pp. 212–242.

[3] A. Agrawal, S. Ramalingam, Y. Taguchi, and V. Chari, "A theory of multi-layer flat refractive geometry," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 3346–3353.

[4] S. Haner and K. Åström, "Absolute pose for cameras under flat refractive interfaces," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1428–1436.

[5] J. Gedge, M. Gong, and Y.-H. Yang, "Refractive epipolar geometry for underwater stereo matching," in *Computer and Robot Vision (CRV), 2011 Canadian Conference on*. IEEE, 2011, pp. 146–152.

[6] Y. Y. Schechner and N. Karpel, "Clear underwater vision," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1. IEEE, 2004, pp. I–536.

[7] ——, "Recovery of underwater visibility and structure by polarization analysis," *Oceanic Engineering, IEEE Journal of*, vol. 30, no. 3, pp. 570–587, 2005.

[8] E. Nascimento, M. Campos, and W. Barros, "Stereo based structure recovery of underwater scenes from automatically restored images," in *Computer Graphics and Image Processing (SIBGRAPI), 2009 XXII Brazilian Symposium on*. IEEE, 2009, pp. 330–337.

[9] A. Jordt, "Underwater 3d reconstruction based on physical models for refraction and underwater light propagation," Ph.D. dissertation, Universitätsbibliothek Kiel, 2013.

[10] K. N. Kutulakos and E. Steger, "A theory of refractive and specular 3d shape by light-path triangulation," *International Journal of Computer Vision*, vol. 76, no. 1, pp. 13–29, 2008.

[11] V. Chari and P. Sturm, "A theory of refractive photo-light-path triangulation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[12] P. Sturm and T. Bonfort, "How to compute the pose of an object without a direct view?" in *Computer Vision–ACCV 2006*. Springer, 2006, pp. 21–31.

[13] C. Geyer and K. Daniilidis, "Structure and motion from uncalibrated catadioptric views," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–279.

[14] ——, "Mirrors in motion: Epipolar geometry and motion estimation," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE, 2003, pp. 766–773.

[15] M. T. Ekvall, G. Bianco, S. Linse, H. Linke, J. Bäckman, and L.-A. Hansson, "Three-dimensional tracking of small aquatic organisms using fluorescent nanoparticles," *PloS one*, vol. 8, no. 11, p. e78498, 2013.

[16] G. Bianco, V. Botte, L. Dubroca, M. R. dAlcalà, and M. G. Mazzocchi, "Unexpected regularity in swimming behavior of clausocalanus furcatus revealed by a telecentric 3d computer vision system," *PloS one*, vol. 8, no. 6, p. e67640, 2013.

[17] J. Cachat, A. Stewart, E. Utterback, P. Hart, S. Gaikwad, K. Wong, E. Kyzar, N. Wu, and A. V. Kalueff, "Three-dimensional neurophenotyping of adult zebrafish behavior," *PloS one*, vol. 6, no. 3, p. e17597, 2011.

[18] J.-Y. Bouguet, "Camera calibration toolbox for matlab," 2004.

[19] Z. Zhang, "A flexible new technique for camera calibration," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 11, pp. 1330–1334, 2000.