A Pixel-wise Varifocal Camera Model for Efficient Forward Projection and Linear Extrinsic Calibration of Underwater Cameras with Flat Housings

Abstract

This paper is aimed at presenting a new virtual camera model which can efficiently model refraction through flat housings in underwater photography. The key idea is to employ a pixel-wise virtual focal length concept to encode the refractive projection inside the flat housing. The radially-symmetric structure of the varifocal length around the normal of the housing surface allows us to encode the refractive projection with a compact representation. We show that this model realizes an efficient forward projection computation and a linear extrinsic calibration in water. Evaluations using synthesized and real data demonstrate the performance quantitatively and qualitatively.

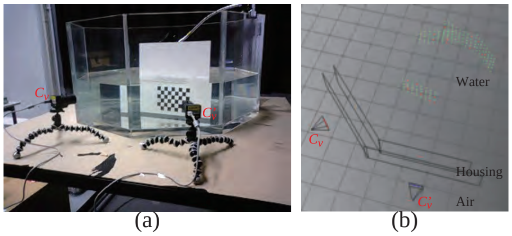
1. Introduction

The successful development of image-based 3D sensing techniques in computer vision is based on the well-studied perspective camera model and the multiple-view geometry in which light rays are supposed to be straight lines [9].

However, this modeling is not valid for environments with more than one media such as underwater photography. In particular, the forward projection via flat housings which computes the projection of 3D points in water to the image is known to be a time-consuming process involving highly non-linear computations [1]. This fact makes applying conventional vision techniques into underwater scenario difficult, since such inefficiency makes all the algorithms on top of 3D-to-2D projections impractical.

To solve this problem, this paper proposes a new virtual camera model which encodes the refractive projection process inside a flat housing by a simple representation, and realizes an efficient forward (3D-to-2D) refractive projection computation and a linear extrinsic calibration. We believe it will open possibilities for applying computer vision techniques into underwater scene, and its applications include education and entertainment such as free-viewpoint 3D visualization of underwater scenes for digital aquariums, and 3D analysis of underwater objects and events such as fertilized eggs and their development.

The key idea on modeling the refraction by flat housing is to employ a pixel-wise virtual focal length which encodes the refraction process inside the flat housing. By exploiting a radially symmetric structure of the pixel-wise focal lengths, we can describe them with a compact representation.



(a) (b)

Figure 1. Overview of the setup. (a) Two cameras observing a chessboard in an octagonal water tank. (b) Calibration result by our method.

The contribution of this paper is twofold. Firstly, our pixel-wise varifocal camera model realizes a compact and efficient representation of the refractive projection via flat housings. Secondly our model realizes a linear extrinsic calibration of cameras in water. To the best of our knowledge, this is the first paper which proposes a linear extrinsic calibration of cameras with flat housings in water.

The rest of this paper is organized as follows. Section 2 reviews related studies. Section 3 defines our measurement model, and Section 4 introduces our pixel-wise varifocal camera model and a linear extrinsic calibration algorithm. Section 5 provides qualitative and quantitative evaluations to demonstrate the advantage of our method. Section 6 concludes this paper with discussions on future work.

2. Related work

While many studies have been proposed for underwater vision [2–4, 7, 12, 16], most of them do not explicitly model refractions by housings. This is mainly because such refractive distortions can be compensated by using dome-shaped housings carefully tailored for each of the cameras. However, flat surface housings are also popular because of the cost, and also because of the fact that regular cameras capturing objects in a water tank via its flat surface are equivalent to underwater cameras with flat surface housings.

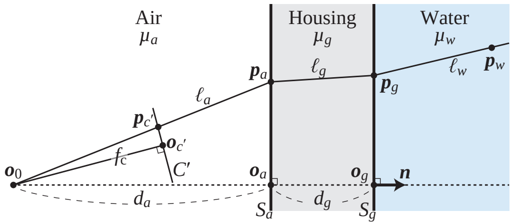


Figure 2. Measurement model

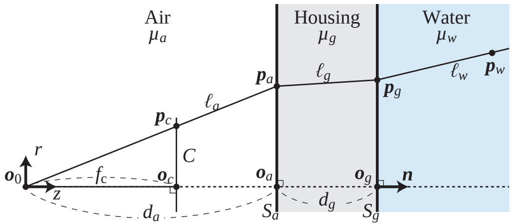


Figure 3. Axial measurement model

In the context of refractions by flat surfaces [1, 6, 10], Agrawal *et al*. [1] have proposed a novel calibration technique based on the axial camera model which estimates the exact model parameters of the refraction such as the thickness of the refractive surface and its refractive indices w.r.t. water and the air, *etc*. By knowing these parameters, Snell's law allows computing the light path passing through the refractive media. However, projecting a 3D point in water to the image involves highly non-linear computations, and hence can be intractable if used for 3D sensing in water in practice. We solve this problem by introducing a new virtual camera model utilizing a pixel-wise varifocal length concept to improve computational efficiency.

Our pixel-wise varifocal length concept defines an incident ray direction for each pixel. Hence it can be seen as a special case of the *raxel* concept [8] in general. In this sense, our contribution is to provide (1) a computationally efficient forward projection algorithm and (2) a linear extrinsic calibration on top of the raxel concept by specializing it as the pixel-wise focal length.

Our extrinsic calibration also allows the flat housing surfaces of the cameras to be located arbitrary in water, while conventional multi-camera systems with flat refractive surfaces [5] assume the cameras to share a single flat surface.

3. Measurement model

Figure 2 illustrates the measurement model of this paper. A pinhole camera C' at $\boldsymbol{o}_0$ observes the underwater scene via a flat housing surface (*e.g*. glass) of d_g thick at d_a distance from $\boldsymbol{o}_0$. A point $\boldsymbol{p}_w$ in water is projected to $\boldsymbol{p}_g$ on the refraction boundary S_g along the segment ℓ_w. $\boldsymbol{p}_g$ is projected to $\boldsymbol{p}_a$ on the refraction boundary S_a along the segment ℓ_g, and then $\boldsymbol{p}_a$ is imaged by the pixel at $\boldsymbol{p}_{c'}$ along the ray ℓ_a connecting $\boldsymbol{p}_a$ and $\boldsymbol{o}_0$.

Notice that we assume the two surfaces S_a and S_g of the housing are flat and parallel, but the camera is not necessarily front-parallel to them. Instead, we employ the axial camera concept proposed by Agrawal *et al*. [1] to simplify this model without loss of generality.

Consider a virtual camera C such that its projection center is placed at $\boldsymbol{o}_0$ and its optical axis is directed along $\boldsymbol{n}$, the normal vector of S_a and S_g (Figure 3). Also let C and C' share a same intrinsic parameter A calibrated beforehand. Then the relationship between 2D pixels of these two cameras is expressed by a homography matrix which projects pixels from C' to C, and the projected point is on the line connecting $\boldsymbol{o}_0$ and $\boldsymbol{p}_{c'}$ by definition.

Since this homography is bijective, we can use C instead of C' without loss of generality. In addition the light paths described using C have a radially symmetric structure about the Z-axis by definition. Hence we utilize the $(r,z)^\top$ coordinate system hereafter.

Let r_α and z_α be the r and z elements of $\boldsymbol{\alpha}$. Also let $\boldsymbol{v}_X = (r_{v_X}, z_{v_X})^\top$ denote the direction vector of line ℓ_X towards the water from the camera. Since Snell's law is expressed as $\mu_a r_{v_a} = \mu_g r_{v_g} = \mu_w r_{v_w}$ by using the refractive indices μ_a, μ_g, and μ_w of the air, housing and water, we can trace the light path $\ell_a - \ell_g - \ell_w$ as

$$\boldsymbol{v}_a = \left(r_{p_a}/\sqrt{r_{p_a}^2 + d_a^2},\ f_c/\sqrt{r_{p_a}^2 + d_a^2} \right)^\top, \quad \boldsymbol{p}_a = \frac{d_a}{f_c}\boldsymbol{v}_a, \tag{1}$$

$$\boldsymbol{v}_g = \left(\frac{\mu_a}{\mu_g} r_{v_a},\ \sqrt{1 - r_{v_g}^2} \right)^\top, \quad \boldsymbol{p}_g = \boldsymbol{p}_a + \frac{d_g}{z_{v_g}}\boldsymbol{v}_g, \tag{2}$$

$$\boldsymbol{v}_w = \left(\frac{\mu_g}{\mu_w} r_{v_g},\ \sqrt{1 - r_{v_w}^2} \right)^\top, \tag{3}$$

where f_c is the focal length of the camera.

These equations allow computing ℓ_w, *i.e.* $\boldsymbol{v}_w$ and $\boldsymbol{p}_g$, from $\boldsymbol{p}_c$. Similarly, computing $\boldsymbol{p}_c$ from ℓ_w can be done by applying Snell's law inversely. Moreover, since only ℓ_w can be the line of the backprojection of ℓ_a, the principle of reversibility of light and the definition of the pinhole imaging ensure that only ℓ_w can be imaged by C among other rays incident at p_w on S_g with different angles.

This suggests that knowing the correct direction of projection is crucial in computing the projection of a point $\boldsymbol{q}_w$ in water. If $\boldsymbol{v}_w$ is available, Snell's law simply provides the analytical solution to find $\boldsymbol{p}_g$, $\boldsymbol{p}_a$, and $\boldsymbol{p}_c$. Otherwise, *i.e.*, if $\boldsymbol{v}_w$ is not given, it requires solving a 12th degree equation, and becomes a time-consuming process [1]. Our goal is to provide a new virtual camera model which realizes an efficient computation of the forward projection of the latter case.

4. The pixel-wise varifocal camera model

Suppose all the model parameters in the previous section including the homography between C' and C have been calibrated beforehand by conventional methods [1]. The goal

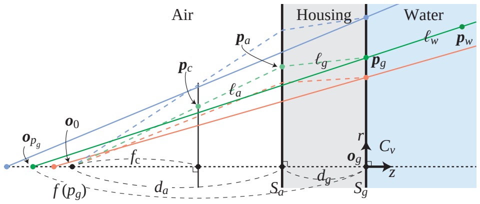


Figure 4. Pixel-wise varifocal camera model. The dashed lines illustrate the correct refractive paths while the straight lines illustrate the perspective projections. In order to represent the correct incident angles of rays in water by a perspective manner, the projection center $\boldsymbol{o}_{p_g}$ moves on Z-axis per pixel ($\boldsymbol{p}_g$) basis.

of this section is to introduce a new virtual camera model which realizes a simple and efficient computation scheme of the refractive forward and backward projections by compiling the calibrated parameters of Figure 3 into another representation.

To this end, we employ *pixel-wise virtual focal lengths* and introduce a virtual camera C_v such that the image screen coincides with S_g and the focal length changes per pixel basis as shown in Figure 4. That is, we make the projection center move by $f(\boldsymbol{p}_{c_v})$ along Z-axis according to the position of each pixel $\boldsymbol{p}_{c_v}$ so that the ray ℓ_w in water passing through a pixel $\boldsymbol{p}_g$ of C_v, $\boldsymbol{p}_a$ and $\boldsymbol{p}_c$ (Figure 4, the green dashed line) can be represented simply by connecting $\boldsymbol{p}_g$ in question and the pixel-wise projection center (Figure 4, the green straight line).

4.1. The pixel-wise focal length

Given a pixel $\boldsymbol{p}_g = (r_{p_g}, d_a + d_g)^\top$ of the virtual camera C_v, consider representing the ray ℓ_w incident at $\boldsymbol{p}_g$ as if C_v is a pinhole camera and its projection center is on Z-axis. Obviously its projection center $\boldsymbol{o}_{p_g} = (0, -f(\boldsymbol{p}_g))^\top$ is given as the intersection of the Z-axis and the line ℓ_w as illustrated in Figure 4. Hence by solving $\boldsymbol{o}_{p_g} = t\boldsymbol{v}_w + \boldsymbol{p}_g$ using Eq (3), we have

$$\begin{pmatrix} 0 \\ -f(\boldsymbol{p}_g) \end{pmatrix} = t \begin{pmatrix} \frac{\mu_g}{\mu_w} r_{v_g} \\ \sqrt{1 - r_{v_w}^2} \end{pmatrix} + \begin{pmatrix} r_{p_g} \\ 0 \end{pmatrix}, \tag{4}$$

$$t = -\frac{\mu_w}{\mu_g} \frac{r_{p_g}}{r_{v_g}}, \tag{5}$$

$$f(\boldsymbol{p}_g) = \frac{\mu_w}{\mu_g} \frac{r_{p_g}}{r_{v_g}} \sqrt{1 - r_{v_w}^2}, \tag{6}$$

$$= \frac{\mu_w}{\mu_g} \frac{r_{p_g}}{r_{v_g}} \sqrt{1 - (\frac{\mu_g}{\mu_w} r_{v_g})^2}, \tag{7}$$

$$= \frac{\mu_w}{\mu_a} \frac{r_{p_g}}{r_{v_a}} \sqrt{1 - (\frac{\mu_a}{\mu_w} r_{v_a})^2}. \tag{8}$$

Once obtained $f(\boldsymbol{p}_g)$ for each radial distance of the virtual camera C_v, we can compute ℓ_w for each r_g without tracing the refraction inside the housing, and can compute the forward (3D-to-2D) and backward (2D-to-3D) projection computations (details are given later). Hence we can consider C_v as a virtual camera which models the refractions inside the housing without loss of generality.

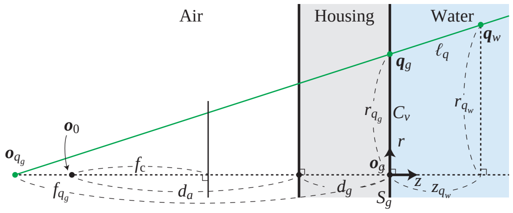


Figure 5. Forward projection by quadratically and globally convergent optimization

4.2. Backward projection using pixel-wise focal length

The backward projection using our varifocal camera model can be done straightforwardly. If a point $\boldsymbol{q}_g$ on S_g is the projection of a 3D point $\boldsymbol{q}_w$ in water, then the viewing ray ℓ_w connecting $\boldsymbol{q}_g$ and $\boldsymbol{q}_w$ is given as

$$\ell_w : (0, -f(\boldsymbol{q}_g))^\top + t\boldsymbol{v}_w, \tag{9}$$

using a parameter t.

4.3. Forward projection using pixel-wise focal length

Consider a 3D point $\boldsymbol{q}_w$ in water, and a 3D line ℓ_q passing through $\boldsymbol{q}_w$ and intersecting with S_g and Z-axis at $\boldsymbol{q}_g$ and $\boldsymbol{o}_{q_g} = (0, -f_{q_g})^\top$ as illustrated in Figure 5. Then the following proposition holds.

Proposition 4.1. *$f(\boldsymbol{q}_g)$, the pixel-wise focal length stored at $\boldsymbol{q}_g$, is equal to f_{q_g} if and only if ℓ_q is identical to the ray imaged by the camera C.*

Proof. The definition of the varifocal camera model ensures that the line passing through $\boldsymbol{p}_g$ on S_g and $\boldsymbol{o}_{p_g} = (0, -f(\boldsymbol{p}_g))^\top$ represents a 3D ray which is projected onto a single pixel of the camera C. On the other hand, the principle of reversibility of light and the definition of the pinhole imaging ensure that there exists only a ray incident at $\boldsymbol{p}_g$ which can be imaged by the camera C. Hence it is only the case making $f(\boldsymbol{q}_g) = f_{q_g}$ that ℓ_q is identical to the ray imaged by C. □

This proposition indicates that we can obtain the projection of $\boldsymbol{q}_w$ on S_g, the image screen of the varifocal camera, by seeking $\boldsymbol{q}_g$ which minimizes the difference between f_q and $f(\boldsymbol{q}_g)$.

4.3.1 Forward projection by a recurrence relation

As illustrated in Figure 6, suppose the 3D point $\boldsymbol{q}_w$ in question is first projected perspectively to $\boldsymbol{q}_0$ on S_g, the

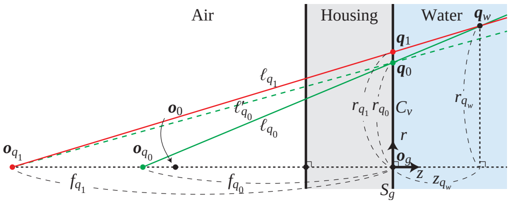


Figure 6. Forward projection by a recurrence relation

virtual screen of C_v, by using an initial (or tentative) focal length f_{q_0} (the green line). By the definition of the pixel-wise varifocal model, the pixel $\boldsymbol{q}_0$ stores its own focal length $f_{q_1} = f(\boldsymbol{q}_0)$ given at the calibration stage. That is, $\boldsymbol{o}_{q_1} = (0, -f_{q_1})^\top$ is the correct virtual projection center for $\boldsymbol{q}_0$ instead of $\boldsymbol{o}_{q_0} = (0, -f_{q_0})^\top$. By iteratively applying perspective projections using $(0, -f_{q_0}), (0, -f_{q_1}), \ldots$, we have

$$r_{q_{k+1}} = r_{q_w} f(\boldsymbol{q}_k) / (z_{q_w} + f(\boldsymbol{q}_k)) \,. \tag{10}$$

By Snell's law and the fact $r_{\boldsymbol{q}_k} > r_{\boldsymbol{q}_{k'}} \Leftrightarrow f(\boldsymbol{q}_k) > f(\boldsymbol{q}_{k'})$, the following monotonicity conditions hold:

$$r_{q_1} > r_{q_0} \Rightarrow r_{q_{k+1}} \geq r_{q_k}, \quad r_{q_1} < r_{q_0} \Rightarrow r_{q_{k+1}} \leq r_{q_k}. \tag{11}$$

Also, the definition of the pixel-wise varifocal model ensures

$$r_{q_{k+1}} = r_{q_k} \Leftrightarrow f(\boldsymbol{q}_{k+1}) = f_{q_k}, \tag{12}$$

and, since $\mu_a < \mu_g$ and $\mu_a < \mu_w$,

$$d_a + d_g \leq {}^{\forall} f_{q_k}. \tag{13}$$

Since Proposition 4.1 ensures that there exists only one r_{q_k} which satisfies Eq (12), starting the recurrence from $f_{q_0} = d_a + d_g$ always converges to the correct value satisfying Eq (12) as shown in Figure 8.

However, the rate of the convergence becomes slower and slower by iteration, because the lines by $\boldsymbol{o}_{q_k}$ and $\boldsymbol{o}_{q_{k+1}}$ (the green and the red lines of Figure 6) become nearly parallel. To solve this problem, we propose a method based on the Newton's algorithm which utilizes this recurrence relation.

4.3.2 Forward projection by a quadratically and globally convergent optimization

Using the 3D point $\boldsymbol{q}_w$ in question, we can describe r_{q_g} as

$$r_{q_g} = E_f(r_{v_w}) = r_q - \frac{r_{v_w}}{z_{v_w}} z_q = r_q - \frac{r_{v_w}}{\sqrt{1 - r_{v_w}^2}} z_q, \tag{14}$$

as shown in Figure 5. On the other hand, the back-projection of the original corresponding pixel in C gives

$$\begin{aligned} r_{q_g} = E_b(r_{v_w}) &= \frac{r_{v_a}}{\sqrt{1 - r_{v_a}^2}} d_a + \frac{r_{v_g}}{\sqrt{1 - r_{v_g}^2}} d_g \\ &= \mu_w r_{v_w} \left(d_a \Big/ \sqrt{\mu_a^2 - \mu_w^2 r_{v_w}^2} + d_g \Big/ \sqrt{\mu_g^2 - \mu_w^2 r_{v_w}^2} \right), \end{aligned} \tag{15}$$

as shown in Figure 3. Since these two r_{q_g} should be equal to each other, we can formulate this as a problem finding r_{v_w} which makes the following $E(r_{v_w})$ be zero.

$$E(r_{v_w}) = E_b(r_{v_w}) - E_f(r_{v_w}). \tag{16}$$

The best r_{v_w} which makes $E(r_{v_w}) = 0$ can be computed by the Newton's method efficiently, and moreover, it converges globally regardless of the initial value.

Proof. The theorem on Newton's method for a convex function ensures that if a function is twice continuously differentiable, increasing, convex and has a zero, then the zero is unique, and the Newton's method will converge to it from any initial value [11].

In case of Eq (16), the first and the second derivatives of $E(r_{v_w})$ are given as

$$\begin{aligned} \frac{dE(r_{v_w})}{dr_{v_w}} =& z_q E_1 + z_q r_{v_w}^2 E_1^3 + d_g \mu_w E_2 \\ &+ d_g \mu_w^3 r_{v_w}^2 E_2^3 + d_a \mu_w E_3 + d_a \mu_w^3 r_{v_w}^2 E_3^3, \end{aligned} \tag{17}$$

$$\begin{aligned} \frac{d^2E}{dr_{v_w}^2} =& 3 z_q r_{v_w} E_1^3 + 3 z_q r_{v_w}^3 E_1^5 + 3 d_g \mu_w^3 r_{v_w} E_2^3 \\ &+ 3 d_g \mu_w^5 r_{v_w}^3 E_2^5 + 3 d_a \mu_w^3 r_{v_w} E_3^3 + 3 d_a \mu_w^5 r_{v_w}^3 E_3^5, \end{aligned} \tag{18}$$

where $E_1 = 1/(1 - r_{v_w}^2)^{1/2}$, $E_2 = 1/({\mu_g}^2 - {\mu_w}^2 r_{v_w}^2)^{1/2}$, and $E_3 = 1/({\mu_a}^2 - {\mu_w}^2 r_{v_w}^2)^{1/2}$. Since r_{v_w} is non-negative by definition, $\frac{d^2E(r_{v_w})}{dr_{v_w}^2} \geq 0$ holds and $E(r_{v_w})$ is a convex function. Obviously $E(r_{v_w})$ is twice continuously differentiable, increasing, and has a zero for $r_{v_w} \geq 0$, then the Newton's method converges globally. □

In addition, while this global convergence allows us to start finding r_{v_w} from any value in $[0, \mu_w/\mu_g]$, the recurrence relation of Eq (10) can provide a reasonable initial guess of r_{v_w} by projecting first by a tentative virtual focal length with a smaller computational cost as shown in Table 1.

5. Linear Extrinsic Calibration Using 16 Points

Suppose we have two pixel-wise varifocal cameras C_v and C'_v. The goal of the extrinsic calibration is to estimate the relative pose R, T of these cameras from a set of corresponding points in their images. Since our virtual camera

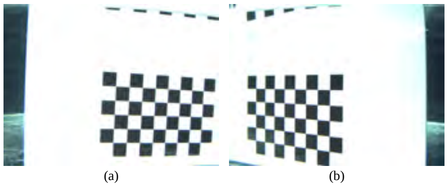


Figure 7. Input images captured by the two cameras in Figure 1(a).

is an axial model, its extrinsic calibration can be seen as a special form of the one for axial cameras [13].

Given a pixel $\boldsymbol{p}_{c'}$ in the real image, we can obtain the corresponding position $\boldsymbol{p}_g$ on S_g without loss of generality as illustrated by Figures 2 and 4. Therefore, given a pair of corresponding points, we can represent the 3D point in water as

$$\begin{aligned} \boldsymbol{q}_w = t_{q_w}\boldsymbol{v}_w + \boldsymbol{q}_g &= \lambda_{q_w}\boldsymbol{v}_w + \boldsymbol{o}_{q_g}, \\ &= R(\lambda'_{q_w}\boldsymbol{v}'_w + \boldsymbol{o}'_{q_g}) + T, \\ \Leftrightarrow \lambda_{q_w}\boldsymbol{v}_w - \lambda'_{q_w}R\boldsymbol{v}'_w &= R\boldsymbol{o}'_{q_g} + T - \boldsymbol{o}_{q_g}, \end{aligned} \tag{19}$$

where λ_{q_w} and λ'_{q_w} denote the unknown depths of the 3D point from $\boldsymbol{o}_{q_g}$ and $\boldsymbol{o}'_{q_g}$.

This equation indicates that $\boldsymbol{v}_w$, $R\boldsymbol{v}'_w$ and $R\boldsymbol{o}'_{q_g} + T - \boldsymbol{o}_{q_g}$ are on a single plane. In other words, they satisfy:

$$\boldsymbol{v}_w^\top \left(\left(R\boldsymbol{o}'_{q_g} + T - \boldsymbol{o}_{q_g} \right) \times \left(R\boldsymbol{v}'_w \right) \right) = 0. \tag{20}$$

By rewriting this as an element-wise formula, we have

$$\begin{aligned} & l_w^\top E_v l'_w = 0, \\ & l_w = \begin{pmatrix} x_{v_w} & y_{v_w} & z_{v_w} & f_{q_g}x_{v_w} & f_{q_g}y_{v_w} \end{pmatrix}^\top, \\ & l'_w = \begin{pmatrix} x'_{v_w} & y'_{v_w} & z'_{v_w} & f'_{q_g}x'_{v_w} & f'_{q_g}y'_{v_w} \end{pmatrix}^\top, \\ & E_v = \begin{pmatrix} r_{31}y_t - r_{21}z_t & r_{32}y_t - r_{22}z_t & r_{33}y_t - r_{23}z_t & -r_{12} & r_{11} \\ r_{11}z_t - r_{31}x_t & r_{12}z_t - r_{32}x_t & r_{13}z_t - r_{33}x_t & -r_{22} & r_{21} \\ r_{21}x_t - r_{11}y_t & r_{22}x_t - r_{12}y_t & r_{23}x_t - r_{13}y_t & -r_{32} & r_{31} \\ -r_{21} & -r_{22} & -r_{23} & 0 & 0 \\ r_{11} & r_{12} & r_{13} & 0 & 0 \end{pmatrix}, \end{aligned} \tag{21}$$

where r_{ij} is the (i, j) element of R and $T = (x_t, y_t, z_t)^\top$. $(x_{v_w}, y_{v_w}, z_{v_w})$ and $(x'_{v_w}, y'_{v_w}, z'_{v_w})$ represent x, y, z elements of $\boldsymbol{v}_w$ and $\boldsymbol{v}'_w$ expressed in their camera coordinate systems respectively.

Since l and l' are given by each of corresponding pairs, we can linearly estimate 17 unknown elements of E_v up to a scale, by using 16 or more corresponding pairs. Once E_v is given, R and T can be obtained linearly from E_v consequently.

6. Evaluation

Figures 1 and 7 show the evaluation setup and a pair of input images. We used two cameras (Pointgrey Chameleon) in front of an octagonal water tank, and observe the scene inside the tank via a flat acrylic surface tank of 35mm thick.

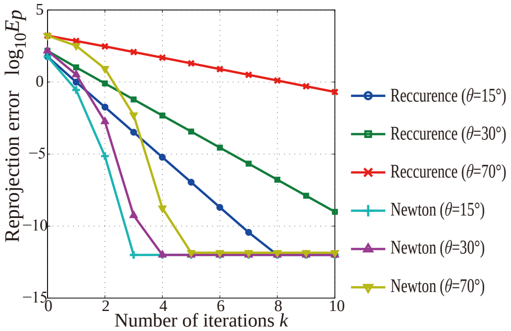


Figure 8. Comparison of the rate of convergence. Notice that errors are lower bounded by 10^{-12}, the default precision of the floating-point computations in our implementation.

Notice that this configuration is equivalent to having two cameras with housings of 35mm thick in water. The model parameters of Section 3 are calibrated by [1] beforehand, and we used the same parameters to synthesize data for quantitative evaluation[1].

6.1. Forward projection

The following evaluations focus on showing the efficiency of the model rather than comparing the accuracy with state-of-the-arts since the proposed model does not improve the accuracy by definition.

Rate of convergence To evaluate the rate of convergence of our iterative methods for the forward projection in Section 4.3, Figure 8 shows the projection error E_p against the number of iterations k. By using a synthesized data set, the reprojection error is defined as

$$E_p = \|P'(\hat{r}_q) - P'(r_{q_k})\|, \tag{22}$$

where $\hat{r}_q$ is the ground-truth and r_{q_k} is the value returned by the algorithm at the k-th iteration in C_v. $P(r_{q_k})$ denotes the pixel position in the original image of C' corresponding to r_{q_k} in C_v. Notice that $P(\cdot)$ is employed only for evaluating E_p in pixels, and is not required for the forward projection to C_v.

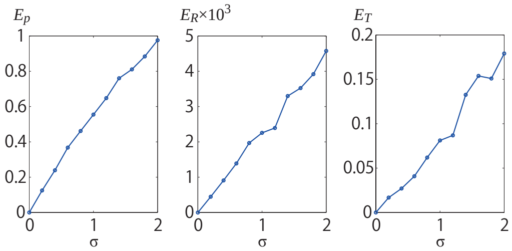
From these results, we can observe that (1) the rates of convergence of the recurrence relation and the Newton-based one are linear and quadratic respectively, and (2) our Newton-based algorithm with 3 times iterations achieve a sub-pixel accuracy.

Computational efficiency Table 1 reports computational costs of our methods computing up to the subpixel accuracy and the state-of-the-art solving the 12th degree of equation analytically [1]. They are the average values of 6400 forward projections run in Matlab on an Intel Core-i5 2.5GHz

[1] Our implementation is available online at `http://annonymous/`.

Table 1. Average computational costs of single forward projections

	Analytical [1]	By Recurrence	By Newton
Runtime	1.39 msec	0.14 msec	0.27 msec
FLOPS	1512	113	250



(a) Reprojection error (b) Rotation error (c) Translation error

Figure 9. Quantitative evaluations of our extrinsic calibration

PC. The FLOPS are counted using Lightspeed Matlab toolbox [14].

From these results, we can conclude that our method runs much faster than the analytical method while maintaining the sub-pixel accuracy.

Linear extrinsic calibration Figure 9 shows calibration errors under different noise levels. Given a set of 160 points synthesized in water, we randomly select 16 points for each trial, and add Gaussian noise with zero-mean and standard deviation $\sigma = 0.1, 0.2, \ldots, 2.0$ to their 2D projections. The three plots report the average errors of 100 trials at each noise level. Here the estimation error of R is defined as the Riemannian distance to the ground truth [15], and the estimation error of T is defined as the RMS error normalized by $|T|$. These results indicate that our linear method performs robustly against observation noise.

Figure 1(b) shows the calibration result using the real images shown in Figure 7. In this calibration, the average reprojection error was 0.3522 pixels, and the angle of the refractive surfaces in front of the two cameras is estimated as 134.7° while it is designed to be 135.0° because of its octagonal structure. These numbers indicate that our calibration method performs reasonably well in practice.

7. Conclusion

In this paper, we proposed a new camera model which employs pixel-wise virtual focal length in order to encode the refraction compactly. Based on the proposed varifocal camera model, we proposed (1) an efficient algorithm for the efficient forward (3D-to-2D) projection, and (2) a linear extrinsic calibration for cameras in water. The evaluations by synthesized and real data demonstrate that (1) our forward computation requires only a few steps to achieve a sub-pixel accuracy and reasonably robust against noise, and (2) our extrinsic calibration well performs in practice.

We believe this method helps us to establish a robust and practical 3D sensing of objects in water that depend on forward (3D-to-2D) projections. Future work includes further studies on the extrinsic calibration, in particular about its degenerated cases, and also on the full 3D surface recovery by multiple cameras in water, etc.

References

[1] A. Agrawal, S. Ramalingam, Y. Taguchi, and V. Chari. A theory of multi-layer flat refractive geometry. In *Proc. of CVPR*, pages 3346–3353, 2012.

[2] M. Alterman, Y. Schechner, and Y. Swirski. Triangulation in random refractive distortions. In *Proc. of ICCP*, pages 1–10, 2013.

[3] M. Alterman, Y. Y. Schechner, P. Perona, and J. Shamir. Detecting motion through dynamic refraction. *IEEE TPAMI*, 35(1):245–251, 2013.

[4] C. Beall, F. Dellaert, I. Mahon, and S. Williams. Bundle adjustment in large-scale 3d reconstructions based on underwater robotic surveys. In *IEEE OCEANS*, pages 1–6, 2011.

[5] Y.-J. Chang and T. Chen. Multi-view 3d reconstruction for scenes under the refractive plane with known vertical direction. In *Proc. of ICCV*, pages 351–358, 2011.

[6] V. Chari and P. Sturm. Multiple-view geometry of the refractive plane. In *Proc. of BMVC*, 2009.

[7] P. Corke, C. Detweiler, M. Dunbabin, M. Hamilton, D. Rus, and I. Vasilescu. Experiments with underwater robot localization and tracking. In *Proc. of ICRA*, pages 4556–4561, 2007.

[8] M. D. Grossberg and S. K. Nayar. The raxel imaging model and ray-based calibration. *IJCV*, 61(2):119–137, Feb. 2005.

[9] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[10] L. Kang, L. Wu, and Y.-H. Yang. Two-view underwater structure and motion for cameras under flat refractive interfaces. In *Proc. of ECCV*, pages 303–316, 2012.

[11] D. Kincaid and W. Cheney. *Numerical Analysis: Mathematics of Scientific Computing*. Pure and Applied Undergraduate Texts Series. American Mathematical Society, 2002.

[12] D. M. Kocak, F. R. Dalgleish, F. M. Caimi, and Y. Y. Schechner. A Focus on Recent Developments and Trends in Underwater Imaging. *Marine Technology Society Journal*, 42:52–67, 2008.

[13] H. Li, R. Hartley, and J.-H. Kim. A linear approach to motion estimation using generalized camera models. In *Proc. of CVPR*, pages 1–8, 2008.

[14] T. Minka. Lightspeed matlab toolbox. `http://research.microsoft.com/en-us/um/people/minka/software/lightspeed/`.

[15] M. Moakher. Means and averaging in the group of rotations. *SIAM J. Matrix Anal. Appl.*, 24:1–16, January 2002.

[16] Y. Schechner and N. Karpel. Clear underwater vision. In *Proc. of CVPR*, volume 1, pages I–536–I–543 Vol.1, 2004.