

A Closed-Form Solution to Single Underwater Camera Calibration Using Triple Wavelength Dispersion and Its Application to Single Camera 3D Reconstruction



Xida Chen and Yee-Hong Yang, *Senior Member, IEEE*

Abstract—In this paper, we present a new method to estimate the housing parameters of an underwater camera by making full use of triple wavelength dispersion. Our method is based on an important finding that there is a closed-form solution to the distance from the camera center to the refractive interface once the refractive normal is known. The correctness of this finding is mathematically proved in this paper. To the best of our knowledge, such a finding has not been studied or reported, and hence is never proved theoretically. As well, the refractive normal can be estimated by solving a set of linear equations using wavelength dispersion. Our method does not require any calibration target, such as a checkerboard pattern, which may be difficult to manipulate when the camera is deployed deep undersea. Extensive experiments have been carried out which include simulations to verify the correctness and robustness to noise of our method and real experiments. The results of real experiments show that our method works as expected. The accuracy of our results is evaluated against the ground truth in both simulated and real experiments. Finally, we also show how we can apply dispersion to compute the 3D shape of an object using one single camera.

Index Terms—Underwater camera calibration, triple wavelength dispersion, 3D reconstruction.

I. INTRODUCTION

COMPUTER vision has been a popular method to explore the underwater environment used by several groups of researchers such as biologists and geologists. It has many applications such as scientific exploration of the structure of the sea floor, and the monitoring of the undersea habitat. Hence, many researchers are interested in underwater computer vision. A fundamental and important topic in this area is underwater camera calibration which includes estimating the camera housing parameters. Camera calibration is a required step for recovering the 3D geometry of a scene using 2D images. Despite the remarkable success [1] for land-based camera systems, underwater camera calibration has not

Manuscript received October 4, 2015; revised October 27, 2016; accepted June 7, 2017. Date of publication June 16, 2017; date of current version July 11, 2017. This work was supported in part by NSERC, in part by AITF, in part by CIF, and in part by the University of Alberta. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Paul Rodriguez. (*Corresponding author: Xida Chen.*)

The authors are with the Department of Computing Science, University of Alberta, Edmonton, AB T6G 2E8, Canada (e-mail: xida@cs.ualberta.ca; yang@cs.ualberta.ca).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2017.2716194

been addressed until recently. In a typical underwater camera system, the camera is placed inside a watertight housing, and views the scene through a flat piece of glass. As a result, the light that travels into the camera undergoes two refractions and its path is not a straight line, the result of which causes distortion in the captured images. The distortion depends on the scene depth and cannot be simply modeled as lens radial distortion [2]. Therefore, calibrating underwater cameras remains a challenging problem in the computer vision area.

In most previous work, the dispersion of light, which is a common phenomenon that light refracts at a different angle depending on its wavelength, was ignored. In contrast, Yau *et al.* [3] study the characteristic of double wavelength dispersion, and exploit dispersion to compute the camera housing parameters. Motivated by their method, we have made an exciting discovery that with triple wavelength dispersion, the interface distance has a closed-form solution once the refractive normal is known. The major advantage of our method is that it does not require any calibration object such as a checkerboard pattern or a light box, and is able to find a unique solution of the camera housing parameters and provides a unique 3D object reconstruction using a single camera. To our best knowledge, all the previous works either require a calibration object, or impose constraints that we do not need. For example, the refractive interface is required to be parallel to the image plane [4]. As well, instead of a closed-form solution, previous methods often initialize the housing parameters with either non-linear optimization [5] or a manual guess [6]. The main contributions of our paper include:

- A new mathematical proof and derivation of a closed-form solution of the interface distance using triple wavelength dispersion after the refractive normal is determined.
- A new method to calibrate the camera housing parameters without using any calibration objects.
- A new method to perform 3D reconstruction using a single camera by exploiting triple wavelength dispersion.

II. RELATED WORK

In underwater computer vision, the refraction effect is sometimes ignored [7], or approximated by [8]–[10]. However, since the refraction effects are highly non-linear and depend on scene geometry, these methods usually produce errors.

Treibitz *et al.* [11] show that the bundle of rays imaged by a perspective camera through a refractive interface does not correspond to a single-viewpoint (SVP) camera. They develop a technique to recover the distance from the camera center to the interface under the assumption that there is only one refraction. The method requires a planar checkerboard pattern with known distance from the camera. Another requirement which is not practical is to require the refractive interface parallel to the image plane.

Chari and Sturm [12] derive a 12×12 refractive fundamental matrix analogous to the ordinary fundamental matrix. However, there is no demonstrated practical application as far as we know. A 3D reconstruction method is proposed in [13] which models the refraction effect as a function of scene depth. The main limitation is that it requires an Inertial Measurement Unit (IMU) to provide the vertical direction of each view. Another limitation is that all the cameras must share the same interface. A calibration method is presented in [6] which does not require a calibration target. It can account for two refractions by assuming that the glass thickness is known. The limitation is that its nonlinear optimization takes hours and the results of real data are not compared to the ground truth. Kang *et al.* [4] develop an optimization procedure with the limitation that the refractive interface is parallel to the image plane. Agrawal *et al.* [14] propose an efficient calibration method by showing that the flat refractive geometry corresponds to an axial camera. With this finding, all the refraction parameters can be computed by solving a set of linear equations. Non-linear optimization is still required to refine the results. Nevertheless, the method assumes the 3D geometry of the calibration target to be known. Therefore, a checkerboard pattern is typically used which may not be practical when the cameras are deployed deep underwater. More recently, Yau *et al.* [3] extend the work of [14] by accounting for the dispersion of light which improves the calibration accuracy. The main limitation is that it requires a heavy custom-built submersible light box which weighs over 60 lbs. Utilizing different wavelengths has been applied in [15] for underwater image enhancement. In particular, this method is based on the fact that different wavelengths of light are attenuated at different rates in water. The method proposed in [5] uses a “virtual camera” error function where each 3D point is projected using an imaginary perspective camera (the “virtual camera”), and applies an iterative nonlinear optimization to minimize the reprojection error. It is claimed that high accuracy is achieved in calibrating the refraction parameters. However, the results of real experiments are not evaluated against the ground truth. A virtual camera model is presented in [16] to model refraction as pinhole cameras with a specific focal length for each image pixel. The model can be used for estimating the camera pose when the housing parameters are provided. A new method is presented in [17] to calibrate the housing parameters with the limitation that the camera pose is provided.

To sum up, there are still limitations in existing methods. Motivated by that, we present a new method that requires no assumption on the configuration of the camera system or the refractive interface, and no calibration target, and yet our

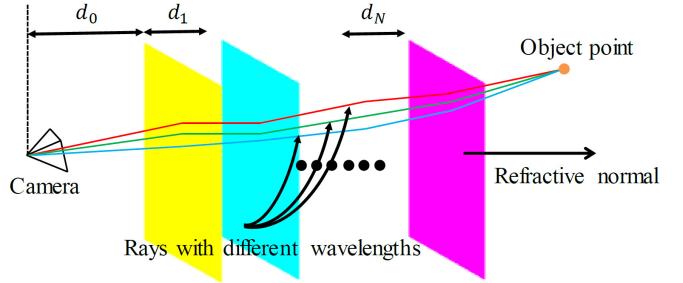


Fig. 1. Refractive model.

method is able to estimate the housing parameters and perform 3D object reconstruction using one single camera with very good accuracy. To our best knowledge, there is no previous method that can map a point from an arbitrary-shaped 3D object to multiple points in one camera view, and perform 3D construction from these multiple points.

III. PROPOSED METHOD

Fig. 1 shows the refraction model used in our method. A pinhole perspective camera is viewing the object through multiple flat refractive layers where all layers are parallel to each other, which means that they have the same refractive interface normal n . The number of layers is N and each layer has a thickness of d_i . The distance from the camera center to the first refractive interface is denoted as the “interface distance”, d_0 . Assume that the object can emit three different wavelengths of light (red, green and blue), then the dispersion can be observed due to different refractive indices for different wavelengths. Our goal is to compute both the refractive normal n and the interface distance d_0 .

We focus on a specific refraction case which is the most common and practical one. In particular, we assume that there are two refractions, which are air \rightarrow glass \rightarrow water. This is the most common scenario for a camera system deployed underwater and for the lab environment. Our method assumes that the thickness of the glass is known because it can be easily measured.

A. Interface Normal Computation

Our method of normal estimation is the same as the one described in [3]. In particular, when observed by the camera, an object point reaches different pixel locations in the camera through different rays due to different wavelengths. Denote v_a, v_b as the two rays with different wavelengths that travel in air and reach the camera, then these two rays must lie on the same plane as the refractive normal n . As a result, a constraint for n can be written as follows.

$$(v_a \times v_b)^T n = 0. \quad (1)$$

Due to different wavelengths, $v_a \neq v_b$ and the above constraint describes that n is on the same plane formed by v_a and v_b . Assume that there are multiple object points described above, then by stacking multiple linear equations yields a linear system. Thus, the refractive normal n can be computed by solving this linear system.

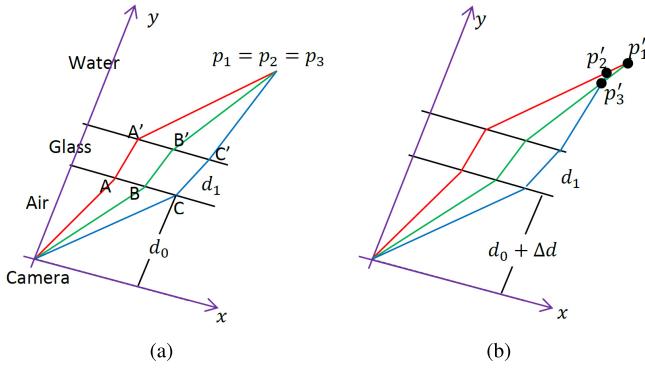


Fig. 2. Computing the interface distance d_0 . See text for details.

B. Interface Distance Computation

The most important novelty of our proposed method is in estimating the interface distance d_0 . In particular, we provide a mathematical proof and derivation to show that there is a closed-form solution to d_0 when the interface normal n is known based on the assumption that triple wavelength dispersion is observed. Because of this finding, our method does not require any calibration target.

Fig. 2(a) depicts a typical case of triple wavelength dispersion. Due to the dispersion of the red, green and blue, an object point is observed at three different pixel locations in the captured images. The direction of the three rays in air can be computed once the camera intrinsic parameters are known. Since the interface normal n is obtained by the method in Section III-A, the direction of the rays traveling in glass and in water can be computed as well. When d_0 is the ground truth, the three rays in water intersect at three coincident points p_1 , p_2 and p_3 . With reference to Fig. 2(b), in the following, we prove that when d_0 is changed to $d_0 + \Delta d$ with any $\Delta d \neq 0$, the three rays in water will intersect at three non-coincident points p'_1 , p'_2 and p'_3 , and vice versa. Using this important finding, d_0 corresponds to the point at which the three points p_1 , p_2 and p_3 are coincident. It is noteworthy that using triple wavelength dispersion is critically important in our method and cannot be replaced using two wavelengths such as the method in [3]. The reason is as follows. Assume that there are only two wavelengths (red and blue), and hence there are two rays in water. We can see that with any d_0 , these two rays in water will always intersect at one point. Because we do not use any calibration target, we cannot determine which d_0 corresponds to the ground truth point in water. In contrast, our method can determine the correct d_0 which corresponds to the point where p_1 , p_2 and p_3 are coincident. In general, our method can be extended to more than 3 wavelengths.

Let's denote the directions of the three rays in air as v_a^r, v_a^g, v_a^b , where the subscript “a” represents air and the superscript color, namely, red(r), green(g) and blue(b). Using this notation, the directions of rays are denoted as v_g^r, v_g^g, v_g^b in glass, and v_w^r, v_w^g, v_w^b in water.

Lemma 1: $v_a^r, v_a^g, v_a^b, v_g^r, v_g^g, v_g^b, v_w^r, v_w^g, v_w^b$ are always on the same plane even if $\Delta d \neq 0$ as shown in Fig. 2(b).

Proof: According to Snell's Law, all the incoming rays and the refracted rays are on the same plane as the refractive

normal. Moreover, in Section III-A, we show that v_a^r, v_a^g, v_a^b are always on the same plane because they are due to the dispersion effects. In other words, all the 9 rays in Fig. 2(a) are on the same plane. Notice that once the interface normal n is known, v_g^r, v_g^g, v_g^b and v_w^r, v_w^g, v_w^b are fixed because they depend on v_a^r, v_a^g, v_a^b and n only. Therefore, even if d_0 is changed to $d_0 + \Delta d$, the ray directions do not change. Hence, the 9 rays are always on the same plane. \square

Since all the rays are on the same plane, we can simplify our problem from 3D to 2D. In particular, we assume that the z plane is aligned with the plane that is formed by the above 9 rays. As shown in Fig. 2(a), we select the camera center as the origin of our new coordinate system. The y axis is perpendicular to the refractive interface, and the x axis is set so that all the 9 rays are in the first quadrant. We denote the slopes of the 9 rays as $m_a^r, m_a^g, m_a^b, m_g^r, m_g^g, m_g^b, m_w^r, m_w^g, m_w^b$.

Theorem 1: $p'_1 \neq p'_2, p'_1 \neq p'_3$ and $p'_2 \neq p'_3 \Leftrightarrow \Delta d \neq 0$.

Proof: We only prove $\Delta d \neq 0 \Rightarrow p'_1 \neq p'_2$, other cases are similar.

The three rays in air pass through the origin. Therefore their equations are $y = m_a^r x$, $y = m_a^g x$ and $y = m_a^b x$. These three rays intersect the inner glass interface at three different points whose coordinates are $(\frac{d_0}{m_a^r}, d_0)$, $(\frac{d_0}{m_a^g}, d_0)$ and $(\frac{d_0}{m_a^b}, d_0)$. Similarly, the equations for the rays in glass and in water, as well as their intersections with the outer glass interface, can be computed. The coordinates of p_1, p_2 and p_3 are provided here. Consider only the x components of these three points, which are written as:

$$\begin{aligned} p_1(x) &= \frac{\frac{d_1 m_w^r}{m_g^r} - \frac{d_1 m_w^g}{m_g^g} + \frac{d_0 m_w^r}{m_a^r} - \frac{d_0 m_w^g}{m_a^g}}{m_w^r - m_w^g} \\ p_2(x) &= \frac{\frac{d_1 m_w^r}{m_g^r} - \frac{d_1 m_w^b}{m_g^b} + \frac{d_0 m_w^r}{m_a^r} - \frac{d_0 m_w^b}{m_a^b}}{m_w^r - m_w^b} \\ p_3(x) &= \frac{\frac{d_1 m_w^g}{m_g^g} - \frac{d_1 m_w^b}{m_g^b} + \frac{d_0 m_w^g}{m_a^g} - \frac{d_0 m_w^b}{m_a^b}}{m_w^g - m_w^b} \end{aligned} \quad (2)$$

Since $p_1 = p_2 = p_3$, $p_1(x) = p_2(x) = p_3(x)$.

The slopes of all the rays in Fig. 2(b) are the same as those shown in Fig. 2(a). Based on the above results, when d_0 is changed to $d_0 + \Delta d$, $p'_1(x), p'_2(x), p'_3(x)$ can be written as follows:

$$\begin{aligned} p'_1(x) &= p_1(x) + \frac{\frac{\Delta d m_w^r}{m_a^r} - \frac{\Delta d m_w^g}{m_a^g}}{m_w^r - m_w^g} \\ p'_2(x) &= p_2(x) + \frac{\frac{\Delta d m_w^r}{m_a^r} - \frac{\Delta d m_w^b}{m_a^b}}{m_w^r - m_w^b} \\ p'_3(x) &= p_3(x) + \frac{\frac{\Delta d m_w^g}{m_a^g} - \frac{\Delta d m_w^b}{m_a^b}}{m_w^g - m_w^b}. \end{aligned} \quad (3)$$

Combining Eq. 2 and 3, we can see that in order to prove $p'_1(x) \neq p'_2(x)$, we need to prove that

$$\left(\frac{m_w^r}{m_a^r} - \frac{m_w^g}{m_a^g} \right) (m_w^r - m_w^b) \neq \left(\frac{m_w^r}{m_a^r} - \frac{m_w^b}{m_a^b} \right) (m_w^r - m_w^g). \quad (4)$$

With careful derivation, Eq. 4 can be simplified to $\frac{|A'C'|}{|AC|} \neq \frac{|B'C'|}{|BC|}$ where $|AC|$ denotes the length of line segment AC . Denote the angle between the y axis and the red ray in air as θ_a^r , and we use a similar notation such as $\theta_a^g, \theta_a^b, \theta_r^g, \theta_r^b, \theta_g^b$. With this notation, $\frac{|A'C'|}{|AC|} \neq \frac{|B'C'|}{|BC|}$ is equivalent to Eq. 5

$$\begin{aligned} & \frac{\sin \theta_a^r}{\sqrt{1-(\sin \theta_a^r)^2}} - \frac{\sin \theta_g^b}{\sqrt{1-(\sin \theta_g^b)^2}} \\ & \frac{\mu_g^r \sin \theta_a^r}{\sqrt{1-(\mu_g^r \sin \theta_a^r)^2}} - \frac{\mu_g^b \sin \theta_g^b}{\sqrt{1-(\mu_g^b \sin \theta_g^b)^2}} \\ & \neq \frac{\frac{\sin \theta_a^r}{\sqrt{1-(\sin \theta_a^r)^2}} - \frac{\sin \theta_g^g}{\sqrt{1-(\sin \theta_g^g)^2}}}{\frac{\mu_g^r \sin \theta_a^r}{\sqrt{1-(\mu_g^r \sin \theta_a^r)^2}} - \frac{\mu_g^g \sin \theta_g^g}{\sqrt{1-(\mu_g^g \sin \theta_g^g)^2}}} \quad (5) \end{aligned}$$

In Eq. 5, μ_g^r denotes the refractive index for the red light in glass. We can see that the left-hand-side of Eq. 5 is a function of μ_g^r and μ_g^b . Let's denote it as $f(\mu_g^r, \mu_g^b)$. The right-hand-side can be denoted as $g(\mu_g^r, \mu_g^g)$. We will prove Eq. 5 by contradiction. Assume that $f(\mu_g^r, \mu_g^b) = g(\mu_g^r, \mu_g^g)$, we can see that this is possible only when both sides are equal to the same constant. If $f(\mu_g^r, \mu_g^b)$ is a constant, it implies that μ_g^r and μ_g^b are related, i.e. they are not independent which is a contradiction.

Hence, $p'_1 \neq p'_2$ for any $\Delta d \neq 0$. \square

Denote the distance between p'_1 and p'_2 as d'_{12} . In the following, we prove that it varies linearly with Δd .

Theorem 2: The distance d'_{ij} where $i \neq j, i = \{1, 2\}, j = \{2, 3\}$, varies linearly with Δd .

Proof: We give the proof for the case $i = 1, j = 2$. Other cases are similar. We only focus on the x component of d'_{12} because the y component is a linear function of x . From Eq. 3, we can derive that

$$\begin{aligned} p'_1(x) - p'_2(x) \\ = p_1(x) - p_2(x) + \Delta d \left(\frac{\frac{m_w^r}{m_a^r} - \frac{m_w^g}{m_a^g}}{m_w^r - m_w^g} - \frac{\frac{m_w^r}{m_a^r} - \frac{m_w^b}{m_a^b}}{m_w^r - m_w^b} \right). \quad (6) \end{aligned}$$

Because $p_1(x) = p_2(x)$ and from Theorem 1 we have $\frac{m_w^r}{m_a^r} - \frac{m_w^g}{m_a^g} \neq \frac{m_w^r}{m_a^r} - \frac{m_w^b}{m_a^b}$, we conclude that $p'_1(x) - p'_2(x) \neq 0$ and varies linearly with Δd . \square

Theorem 3: There is a closed-form solution to d_0 .

Proof: From Eq. 2 together with $p_1(x) = p_2(x)$, we can derive Eq. 7, which concludes that d_0 has a closed-form solution:

$$\begin{aligned} d_0 \\ = \frac{\left(\frac{d_1 m_w^r}{m_g^r} - \frac{d_1 m_w^g}{m_g^g} \right) (m_w^r - m_w^b) - \left(\frac{d_1 m_w^r}{m_g^r} - \frac{d_1 m_w^b}{m_g^b} \right) (m_w^r - m_w^g)}{\left(\frac{m_w^r}{m_a^r} - \frac{m_w^b}{m_a^b} \right) (m_w^r - m_w^g) - \left(\frac{m_w^r}{m_a^r} - \frac{m_w^g}{m_a^g} \right) (m_w^r - m_w^b)}. \quad (7) \end{aligned}$$

\square

We can see that d_0 is easy to compute once the interface normal n is estimated. In here, all the rays from three wavelengths are needed to compute d_0 .

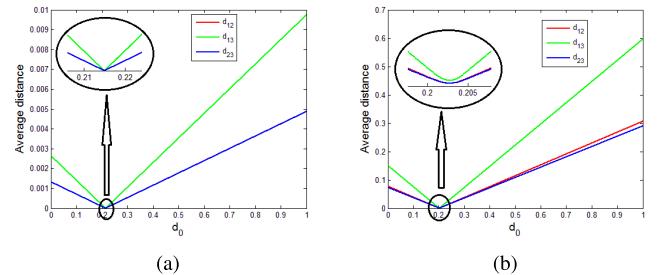


Fig. 3. Plot of d_{12}, d_{13}, d_{23} when the observed pixel is (a) Noise free (b). Corrupted by Gaussian noise with variance $\sigma = 0.5$ pixels.

C. 3D Reconstruction

Because an object point is observed at three different pixel locations due to dispersion, we can backward project from these three pixel locations to compute the 3D coordinates of this object point once the interface normal n and distance d_0 are known. Due to noise, we compute the barycenter of p_1, p_2 and p_3 as the final 3D point. One possibility is to compute the intersection of the red and blue rays in water as the final 3D point, which utilizes two wavelengths only. However, we found that using three wavelengths generates better results in general.

D. Implementation Details

Our method depends on triple wavelength dispersion, which means that the object needs to emit three different kinds of colors (red, green and blue). Our implementation to achieve this is simple and very practical. We use an LED projector (TI LightCrafter) to project three colors to illuminate the object. In this case, the object changes its original color to red, green and blue, to create the required triple wavelength dispersion.

When the observed pixel location is corrupted by noise, the three points p_1, p_2, p_3 may not be the same, in which case, we compute d_0 by minimizing the distances among these three points, which vary linearly with Δd as proved in Theorem 2. Fig. 3 demonstrates the procedure. In particular, we select two datasets in our simulated experiment where one is noise free (3(a)) and the other one corrupted with Gaussian noise (3(b)). The x axis is the value for d_0 and the y axis shows the average of d_{12}, d_{13} and d_{23} of all the object points. We can see that the distance varies linearly with Δd even in the presence of Gaussian noise. In the case when it is noise free, all three distances are zero at the minimum. Notice that the housing parameters are randomly generated in our simulated experiments, therefore the ground truth for d_0 is different for different datasets.

After the interface normal n and the distance d_0 are computed, we apply a nonlinear refinement process to further optimize the results. In particular, all the object points are reprojected to the image using all three wavelengths, and the reprojection error $\mathcal{J} = \sqrt{\frac{1}{M} \sum_{i=1}^M \sum_{j=R,G,B} (x_i(j) - \hat{x}_i(j))^2}$ is minimized. M is the number of correspondences, x is the measured pixel location where \hat{x} the reprojected. The Matlab function *fmincon* is applied in this step.

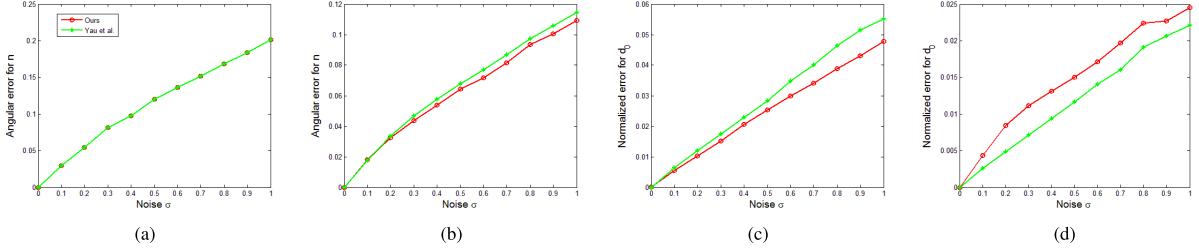


Fig. 4. Results for simulated experiments. (a, b) Angular error before and after nonlinear refinement. (c, d) Normalized error for d_0 before and after nonlinear refinement.

We run our Matlab code on a desktop PC with Intel Core i7. Solving the refractive normal takes about 1 second, the interface distance about 10 seconds, and the nonlinear optimization less than 2 minutes.

IV. EXPERIMENTAL RESULTS

Simulations and real experiments are designed to demonstrate the accuracy and robustness of our method. In the experiments, we assume that the camera is mounted in a housing with a flat glass interface, and hence the light path from the camera to the object is air \rightarrow glass \rightarrow water, which is the most common scenario for the lab environment and for camera systems deployed underwater. The refractive index for air is 1.0 for all three wavelengths. In water, it is 1.343 for red, 1.337 for green and 1.332 for blue. In glass, it is 1.516 for red, 1.502 for green and 1.488 for blue.

A. Simulations

In our simulated experiment, the focal length of the camera is set to be 5600 pixels. The resolution of the image is 5472×3648 with the principal point at the center of the image. The parameters are based on the Canon 6D camera used in our real experiments. We assume that there is no distortion in the image. In the experimental setup, the refractive normal n for each camera is randomly generated within a range of $[10^\circ, 15^\circ]$ away from the camera optical axis. Similarly, the interface distance d_0 is also randomly generated to be within $0.2 \sim 0.25$ units. Because the glass thickness d_1 varies depending on the depth where a camera system is deployed, we set it to be a random number between 0.02 and 0.5. 100 object points are randomly generated and each placed at a random distance that is around 1.5 units from the camera. Each object point is observed at three different pixels and the pixel locations are corrupted by Gaussian noise (variance σ^2 pixels) and we perform 100 trials for each noise setting. We evaluate our results as follows. Suppose the refractive normal recovered by our method is \hat{n} and the ground truth is n , then the angle between them is measured in degrees and called "angular error." For interface distance d_0 , the normalized error is computed by $\frac{|\hat{d}_0 - d_0|}{d_0}$, where $|\cdot|$ denotes the absolute value. We can see that all the measures indicate errors. Hence, a lower value in the curves indicates a better result.

Fig. 4 shows the results of our simulated experiments. We compare our results with that of the method [3] that is

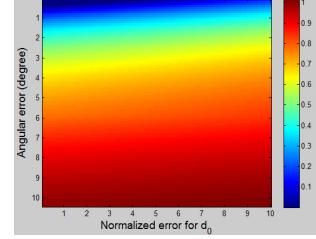


Fig. 5. Impact of the housing parameters to the final 3D reconstruction results. See text for details.

known to be the currently best underwater camera calibration method. The implementation of [3] requires that the 3D geometry of the object to be known. To achieve that, we assume that the 3D coordinates of the object points are known in the world coordinate system. We can see that this is impossible for an arbitrary 3D object. From the results we can see that comparing with [3], our method can achieve similar accuracy even though it does not require the 3D geometry of the object.

B. Importance of Calibration

We demonstrate the importance of the camera housing calibration by showing the impact of the housing parameters to the final 3D reconstruction. We use the same configuration and parameters as our previously mentioned simulated experiments. Using the observed pixel locations with triple wavelength dispersion along with the ground truth housing parameters, these 3D object points can be computed without any error. However, once the housing parameters are changed, errors would exist in the computed 3D object points. We denote the ground truth coordinates of each object point as P , and the coordinates computed by the changed parameters as \hat{P} . The distance between these two points is obtained and averaged for all the 3D points, which is denoted as D . We can see that D is the error when compared with the ground truth.

Fig. 5 shows the error D . In particular, the x axis denotes the normalized error for d_0 in percentage, and the y axis denotes the angular error of the interface normal. We compute the error D when the angular error is less than 10.5° and the error for d_0 is smaller than 10%. This figure is color coded where blue indicates a smaller error and red a larger error. The color bar is also attached to the right of this figure. The figure shows that when the interface normal is 10° off from the ground truth and d_0 is 10% off, the error D is around

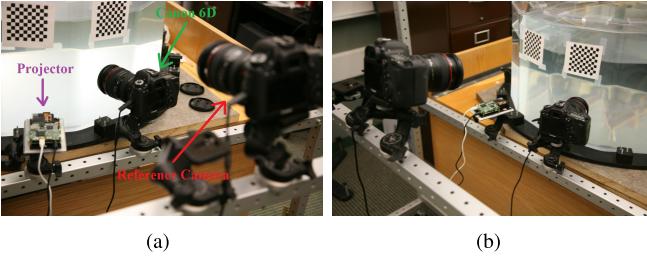


Fig. 6. Setup for our real experiments.

1 unit. This error is very large, because each of the 3D points in the ground truth is about 1.5 unit away from the camera. In other words, the housing parameters are critically important to the 3D reconstruction result.

C. Real Data

Real experiments are carried out in a lab environment and the setup is shown in Fig. 6. A Canon 6D camera (indicated by the green arrow), with a resolution of 5472×3648 , is placed in front of a plexiglass tank. A projector, which is indicated by the purple arrow, is used to project the red, green and blue lights onto the object. A reference camera (indicated by the red arrow), which is another Canon 6D camera, is used to obtain the ground truth of the refractive normal n and the interface distance d_0 . In particular, the reference camera focuses on the checkerboard pattern which is pasted on the interface. Therefore, the transformation between the coordinate system of the checkerboard and that of the reference camera can be computed through calibration. Once the two cameras are calibrated in air, the transformation between the coordinate systems of the checkerboard and of the cameras facing the interface can be computed. As a result, the ground truth of n and of d_0 can be obtained, and hence our results can be evaluated against them. The thickness of the glass is measured by a ruler. In addition to using a Canon 6D camera, we also perform the same experiment using a Point Grey Blackfly mono camera with a resolution of 2448×2048 . Since the method presented in [3] requires a light box as a calibration target while our experiments use an arbitrary scene, we do not compare with this method.

We first demonstrate that the dispersion effect can be observed by both types of cameras. A checkerboard pattern is placed in water and two images are captured when it is illuminated by the red and blue lights from the projector. If the dispersion were not observed, then the edge of the grids would align perfectly with each other in these two images. We superimpose the blue image onto the red image and the results are shown in Fig. 7. We can see that the edges do not align well which indicates that the dispersion between red and blue can be observed.

The procedures of our experiment can be described as follows. (1) We first correct the lens chromatic aberration (CA) in air using the same method presented in [3]. Throughout our experiments, we found that the error in d_0 is about 10 times larger without correction than with CA correction. (2) Red, green and blue patterns are projected onto an arbitrary

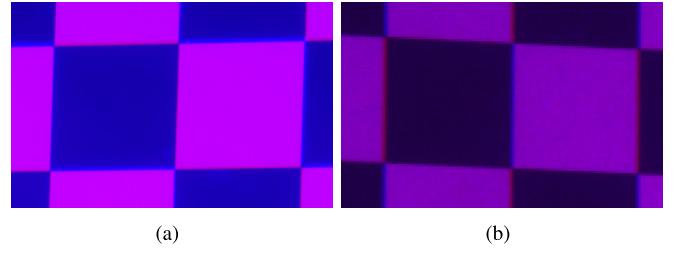


Fig. 7. The dispersion effect observed by (a) Canon 6D camera and (b) Blackfly camera.

TABLE I

EVALUATION OF OUR METHOD AGAINST THE GROUND TRUTH

	Err in d_0	Err in n (deg)	3D err (mm)
Canon	1.733%	0.835	1.773
Blackfly	2.318%	1.093	2.393

scene and for each pattern, one image is captured. (3) SIFT descriptors are compared among the three images to find SIFT matches. The refractive normal and the interface distance are estimated using the detected SIFT matches. (4) Gray code [18] patterns are projected using red, green and blue in order to establish dense correspondence. The dispersion effect from the projector can be safely ignored due to its low resolution. Fig. 8 shows the result for each type of camera. The top row is the result for the Canon camera and the bottom row for the Blackfly camera. The left image is captured under ambient light. The one in the bottom row has no color because it is a mono camera. The middle two are the reconstructed 3D point clouds using the dense correspondences established from gray code and the housing parameters estimated by our method. We can see that many details are preserved in the 3D model. It is worthy to note that even though the image taken by the Blackfly camera has no color, we can combine the images captured when the scene is illuminated under the red, green and blue light, to form a color image. We follow the method presented in [17] to compare the 3D point clouds of our method and that of the ground truth. That is, the right image shows the error maps by comparing the 3D point cloud reconstructed using our estimated parameters with that using the ground truth parameters. The error is larger when it is brighter. The evaluation of our result for these two scenes is shown in Table I. The errors for the parameters are very close to the ones reported in [3] using a calibration target, and are consistent with our simulated results. The results also indicate that our method is practical. Besides the parameters, we also compute the averaged distance between our 3D points and the ground truth, and is denoted as “3D err” in the table. The value of the 3D error is very small, which indicates that our method can generate accurate results. Our real experiments also demonstrate that a mono camera can be used because the dispersion effects can be observed. Moreover, by combining the three images captured when the scene is illuminated by the red, green and blue patterns, we can get the color information of every pixel.

We have also applied our method to an undersea monitoring system that is currently deployed at over 100m at Sannish

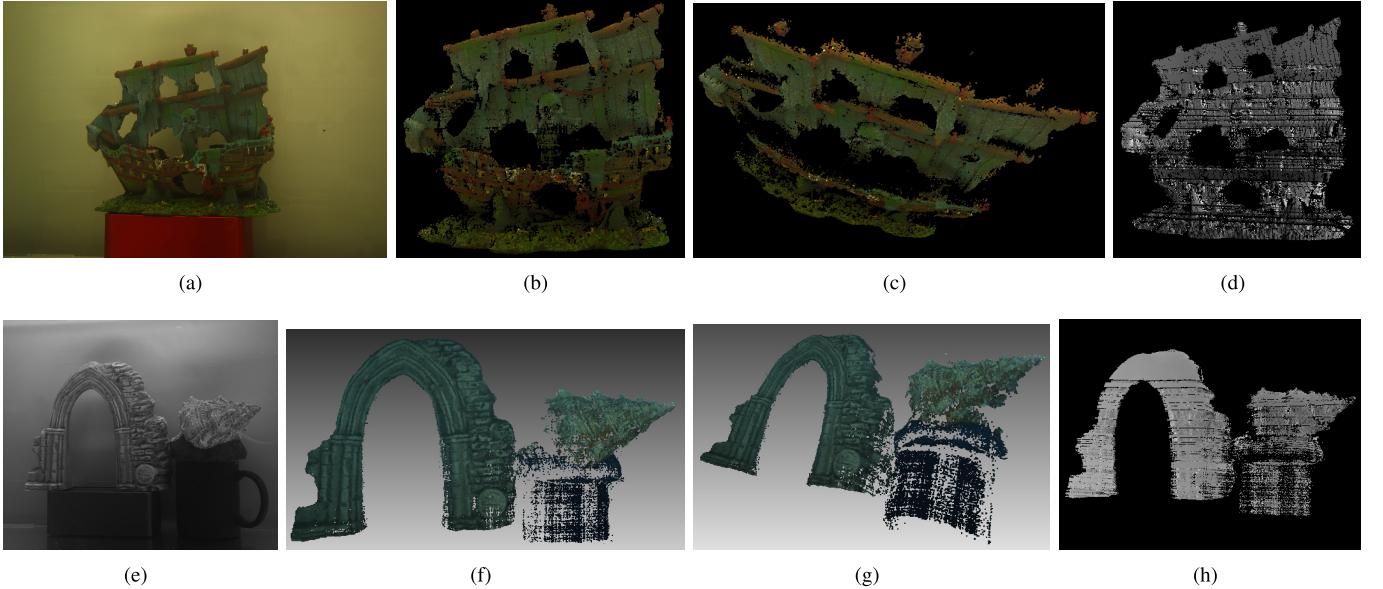


Fig. 8. (a, e) Images captured under ambient light. (b,c,f,g) Reconstructed 3D point cloud from two viewpoints. (d, h) Error map for 3D point cloud compared with the ground truth.



Fig. 9. The undersea monitoring system (Courtesy Ocean Networks Canada).

Inlet, BC, by Ocean Networks Canada. The system is designed at the University of Alberta [19] to monitor the undersea habitat. Fig. 9 shows a picture of the monitoring system. The system consists of 8 5M pixels Point Grey Flea3 Gige color cameras, 3 TI fanless data projectors. The system has two

PCs, one of them controls the cameras and projectors, while the other one serves as back-up. Fig. 10 shows a recently captured dataset and its result. In particular, Fig. 10(a) shows an image that is captured by one undersea camera. The top row of Fig. 10(b) to 10(c) are the 3D reconstruction result by using triple wavelength dispersion, where the bottom row show the result by using correspondences from two camera views. Visual comparison shows that our method produces result that is similar in quality to that of using two camera views.

1) Handling Outliers: Our method can handle outliers easily. When examining a pair of matched SIFT descriptors, they are regarded as outliers if the distance of matched pixel locations in these three images is larger than a threshold. For the Canon 6D camera, the threshold is 10 pixels and 5 for the Blackfly camera. The reason is that a matched SIFT descriptor must be from the same physical point, even though it is observed at different pixel locations due to dispersion, the difference of pixel location should not be too large.

V. DISCUSSIONS AND LIMITATIONS

Most existing methods can be grouped into two categories. The methods in the first category require at least two cameras viewing the same scene, and estimate parameters from a set of feature matches. Two representative methods are [5] and [6]. The methods in the second category perform single camera calibration with a calibration target. Two representative methods are [3] and [14].

We compare our method with some recent methods in Table II and highlight our advantages. First, comparing with the methods in the first category, we do not need to estimate the relative camera pose which largely reduces the number of parameters to be optimized. Moreover, our method can use SIFT matches from the entire field of view of the camera because we use only one camera. Last but not least, our initialization is more accurate than using manual guess

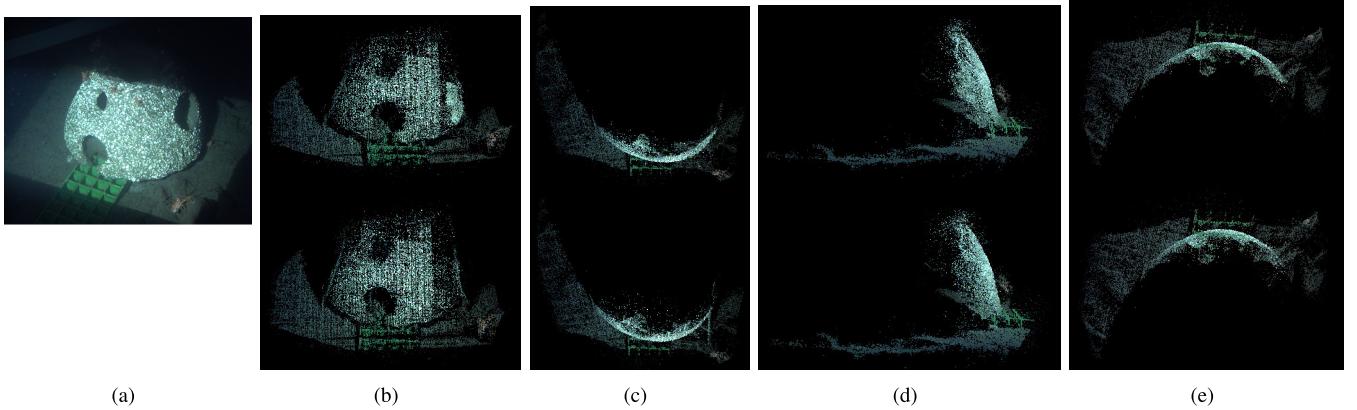


Fig. 10. (a) Captured image by a camera. (b-e) Comparison of 3D reconstruction result by using our method and by using two camera views. Top row: our result. Bottom row: result of using two camera views.

TABLE II
COMPARISON OF DIFFERENT METHODS

	Initialization	Refinement	Shortcomings
Category 1: [6], [5]	Manual guess [6], Initialization for the housing parameters is not shown [5]	Nonlinear optimization	Need to estimate relative camera pose, SIFT matches come from the regions viewed by both cameras only, and the results using real data are not compared to the ground truth.
Category 2: [14], [3]	Solving linear equations	Nonlinear optimization	Require calibration object.
Ours	Solving linear equations and closed-form solution	Nonlinear optimization	See text for details.

in [6]. Our nonlinear optimization is more efficient compared to that in [6]. The method in [5] claims that it can solve three different problems. However, the initialization for the housing parameters is not shown. As well, in their simulations, the results of the estimated the housing parameters are not shown or evaluated.

While our method is motivated by [3], it is not a trivial extension. Reference [3] uses two wavelengths dispersion to compute the interface normal while we use triple wavelengths to compute d_0 . We have shown earlier that using two wavelengths will fail to compute a unique d_0 . Our method produces results with similar accuracy comparing to that in [3], which means that it is more accurate than that in [14]. Comparing to [3], the major advantage of our method is that we do not require any calibration target. In particular, our novelty is to exploit triple wavelength dispersion, which not only requires no calibration target but also provides an easy to compute *closed-form* solution of the interface distance. Our method of projecting red, green and blue patterns to the object is also unique, which removes the requirement for an object of known geometry to emit light with three different wavelengths. Such an implementation is obviously better than the custom-built light box [3]. As well, our method can perform 3D reconstruction from a single camera where no previous methods can. In other words, our method does not require estimating the relative pose between cameras, which eliminates the error from such an estimation. To our best knowledge, all the existing 3D reconstruction methods either use multiple cameras or single moving cameras, which always require relative pose estimation.

There is a limitation of our method, which is demonstrated in Fig. 11. In particular, when the ray connecting the camera

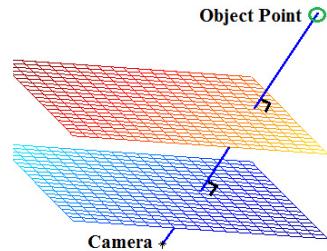


Fig. 11. Demonstration of our limitation.

center and the object point is perpendicular to the refractive interface, then this ray undergoes no refraction at all. In this case, there is no dispersion observed which means that our method does not apply. However, notice that there is only one point in the entire field of view (FOV) of the camera where this would happen. In Fig. 11, we assume that the dispersion of all the points within the green circle is less than one pixel in the image, and we will demonstrate that this green circle is very small comparing to the FOV of the camera. We use the configuration of our simulated experiments and assume that the z component of all the object points is 1.2 units. In this case, the FOV of the camera is about 0.9×0.6 units. We found that the radius of the green circle is 0.042 unit, which means that there is about 1.03% of the FOV of the camera where the dispersion is less than one pixel. Obviously this area is very small.

VI. CONCLUSION

We have presented a new method to compute the interface distance and a new 3D reconstruction method using a single

camera. We first demonstrate that the refractive normal can be estimated using dispersion. After that, we provide a mathematical proof and derivation of a closed-form solution of the interface distance by utilizing triple wavelength dispersion. Simulated and real experiments are performed to evaluate the correctness and robustness of our method. All the experimental results are evaluated against the ground truth, and the simulation results are compared with the state-of-the-art method [3]. The results indicate that the our method has a high accuracy and is practical since it works with any arbitrary 3D scene.

ACKNOWLEDGMENTS

The authors would like to thank Steve Sutphen for his help during the whole research. Finally, the undersea work would not be completed without the support of the staff of Ocean Networks Canada.

REFERENCES

- [1] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.
- [2] O. Pizarro, R. Eustice, and H. Singh, "Relative pose estimation for instrumented, calibrated imaging platforms," in *Proc. DICTA*, 2003, pp. 601–612.
- [3] T. Yau, M. Gong, and Y.-H. Yang, "Underwater camera calibration using wavelength triangulation," in *Proc. CVPR*, 2013, pp. 2499–2506.
- [4] L. Kang, L. Wu, and Y.-H. Yang, "Two-view underwater structure and motion for cameras under flat refractive interfaces," in *Proc. ECCV*, 2012, pp. 303–316.
- [5] A. Jordt-Sedlazeck and R. Koch, "Refractive structure-from-motion on underwater images," in *Proc. ICCV*, 2013, pp. 57–64.
- [6] A. Jordt-Sedlazeck and R. Koch, "Calibration of housing parameters for underwater stereo-camera rigs," in *Proc. BMVC*, 2011, pp. 118-1–118-11.
- [7] J. P. Queiroz-Neto, R. L. Carceroni, W. Barros, and M. Campos, "Underwater stereo," in *Proc. SIBGRAPI*, 2004, pp. 170–177.
- [8] R. Ferreira, J. Costeira, and J. A. Santos, "Stereo reconstruction of a submerged scene," in *Proc. IbPRIA*, 2005, pp. 102–109.
- [9] L. Kang, L. Wu, and Y.-H. Yang, "Experimental study of the influence of refraction on underwater three-dimensional reconstruction using the svp camera model," *Appl. Opt.*, vol. 51, no. 31, pp. 7591–7603, 2012.
- [10] J.-M. Lavest, G. Rives, and J.-T. Lapreste, "Underwater camera calibration," in *Proc. ECCV*, 2000, pp. 654–668.
- [11] T. Treibitz, Y. Y. Schechner, and H. Singh, "Flat refractive geometry," in *Proc. CVPR*, 2008, pp. 1–8.
- [12] V. Chari and P. Sturm, "Multiple-view geometry of the refractive plane," in *Proc. BMVC*, London, U.K., 2009, pp. 1–11.
- [13] Y.-J. Chang and T. Chen, "Multi-view 3D reconstruction for scenes under the refractive plane with known vertical direction," in *Proc. ICCV*, Nov. 2011, pp. 351–358.
- [14] A. Agrawal, S. Ramalingam, Y. Taguchi, and V. Chari, "A theory of multi-layer flat refractive geometry," in *Proc. CVPR*, 2012, pp. 3346–3353.
- [15] J. Y.-W. Chiang and Y.-C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1756–1769, Apr. 2012.
- [16] R. Kawahara, S. Nobuhara, and T. Matsuyama, "A pixel-wise varifocal camera model for efficient forward projection and linear extrinsic calibration of underwater cameras with flat housings," in *Proc. ICCV Workshop*, 2013, pp. 819–824.
- [17] X. Chen and Y.-H. Yang, "Two-view camera housing parameters calibration for multi-layer flat refractive interface," in *Proc. CVPR*, 2014, pp. 524–531.
- [18] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proc. CVPR*, 2003, pp. 195–202.
- [19] X. Chen, S. Sutphen, P. Macoun, and Y.-H. Yang, "Design and implementation of a 3D undersea camera system," *CoRR*, Apr. 2015. [Online]. Available: <http://arXiv:1504.01753>



Xida Chen received the Ph.D. degree from Computing Science Department, University of Alberta in 2015. His current research interests include computer vision and image processing. In particular, he was involved in structured light, underwater camera calibration and 3D reconstruction, and their related applications. He joins Google in 2015.



Yee-Hong Yang (SM'91) received the B.Sc. degree (Hons.) from The University of Hong Kong, Hong Kong, the M.Sc. degree from Simon Fraser University, and the M.S.E.E. and Ph.D. degrees from the University of Pittsburgh. He was a Faculty Member with the Department of Computer Science, University of Saskatchewan, from 1983 to 2001, and served as the Graduate Chair from 1999 to 2001. In addition to department level committees, he also served on many college and university level committees. Since 2001, he has been a Professor

with the Department of Computing Science, University of Alberta, where he served as Associate Chair (Graduate Studies), from 2003 to 2005. He has authored or co-authored over 140 papers in international journals and conference proceedings in the areas of computer vision and graphics. His research interests cover a wide range of topics from computer graphics to computer vision, which include physically based animation of Newtonian and non-Newtonian fluids, texture analysis and synthesis, human body motion analysis and synthesis, computational photography, stereo and multiple view computer vision, and underwater imaging. In addition to serving as reviewer to numerous international journals, conferences, and funding agencies, he has served on the program committees of many national and international conferences. In 2007, he was invited to serve on the expert review panel to evaluate computer science research in Finland. He serves on the Editorial Board of the journal *Pattern Recognition*.