Two-View Camera Housing Parameters Calibration for Multi-Layer Flat Refractive Interface

Xida Chen   Yee-Hong Yang

Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada

{xida, yang}@cs.ualberta.ca

Abstract

In this paper, we present a novel refractive calibration method for an underwater stereo camera system where both cameras are looking through multiple parallel flat refractive interfaces. At the heart of our method is an important finding that the thickness of the interface can be estimated from a set of pixel correspondences in the stereo images when the refractive axis is given. To our best knowledge, such a finding has not been studied or reported. Moreover, by exploring the search space for the refractive axis and using reprojection error as a measure, both the refractive axis and the thickness of the interface can be recovered simultaneously. Our method does not require any calibration target such as a checkerboard pattern which may be difficult to manipulate when the cameras are deployed deep undersea. The implementation of our method is simple. In particular, it only requires solving a set of linear equations of the form $Ax = b$ and applies sparse bundle adjustment to refine the initial estimated results. Extensive experiments have been carried out which include simulations with and without outliers to verify the correctness of our method as well as to test its robustness to noise and outliers. The results of real experiments are also provided. The accuracy of our results is comparable to that of a state-of-the-art method that requires known 3D geometry of a scene.

1. Introduction

Camera calibration is one of the most extensively researched topics in computer vision. It is a required step for 3D geometry in order to extract metric information from 2D images. Despite the remarkable success [17] for land-based camera systems, underwater camera calibration has been unaddressed in computer vision community until recently. The main difficulty is the refraction which occurs when light travels through different media, which results in distortion in the image that depends on the scene distance and cannot be simply modeled as lens radial distortion [11].

Therefore, calibrating an underwater camera system with multiple refractive layers with an unknown refractive axis and layer thickness is still an open problem.

In this paper, we focus on the problem of recovering the refractive axis and layer thickness simultaneously for a two-view underwater stereo camera system. The key idea of our method is an important finding that the layer thickness can be estimated once the refractive axis is known. To the best of our knowledge, such a relationship is unknown in previous work. Moreover, we show that the refractive axis is constrained within a search space and can be specified once the stereo cameras are calibrated in air. Combining the above two findings, both the axis and layer thickness can be computed. Our main contributions are as follows.

• A new theory to demonstrate that the layer thickness can be estimated once the refractive axis is given.

• A new method to recover both the axis and the thickness by analyzing the search space.

• Our method only requires a set of pixel correspondences in the stereo images of any arbitrary 3D scene. The calibration pattern is used only once to calibrate the relative pose between the cameras as a pre-processing step.

• The implementation is simple, which can be easily adapted to most existing underwater stereo camera systems.

2. Related Work

The refraction effects are either ignored in the early works of underwater computer vision [12], or approximated, such as using focal length adjustment [4, 6, 9]. However, it is a known fact that the refraction effects are highly non-linear and depends on the scene geometry. Therefore, approximation methods usually produce errors in the results. It is demonstrated by Treibitz *et al.* [15] that errors exist in underwater camera calibration by using a single viewpoint (SVP) model. Moreover, they develop a calibration technique to recover the distance from the camera center to the interface assuming that there is only one refraction. This method uses a planar checkerboard pattern, and assumes that the depth of the pattern is known. It also

applies the constraint that the interface must be parallel to the image plane which is not practical.

Applying a physically correct model has attracted the attention of researchers since it is able to improve the 3D reconstruction results. Chari and Sturm [3] provides theoretical analysis of refraction to demonstrate that the refractive fundamental matrix exists between two cameras viewing a scene through the same refractive interface. However, only theoretical results are provided instead of practical implementation. A multi-view 3D underwater reconstruction method is proposed in [2] where all the cameras share the same refractive interface. In this method, the effect of refraction is modeled as a function of scene depth. Nonetheless, it assumes that the normal of the refractive interface is known and also requires an additional device called inertial measurement unit (IMU) to measure the roll and pitch angles of the camera. A flexible method is presented in [14] to calibrate the housing parameters for underwater stereo rigs. The method does not require a calibration object and can account for two refractions by assuming that the glass thickness of the housing is known. A major limitation is that its optimization process is very time consuming and the result of real data is not evaluated. Kang *et al.* [7] study two-view structure and motion using cameras that do not share the same interface. A limitation of this technique is the assumption that the interface is parallel to the image plane. Agrawal *et al.* [1] show that the flat refractive geometry corresponds to an axial camera. Such a finding leads to a general theory of camera calibration by assuming that the 3D geometry of the calibration target is known. As a result, a checkerboard pattern is usually used as the calibration target when implementing the method, which may not be practical when the camera system is deployed deep undersea. Yau *et al.* [16] extend the work of [1] by considering the dispersion of light which improves the calibration accuracy. However, it requires a heavy custom built submersible light box. More recently, [5] present a structure-from-motion method for underwater images, which performs calibration for both the housing parameters and the relative pose. However, the method requires a good initial estimate for the housing parameters. Our method does not have this limitation. In particular, our method obtains a good initial estimate from the stereo correspondences and refines it by sparse bundle adjustment.

We can see that there are certain requirements in most existing calibration methods that make them impractical. For example, a planar pattern or a light box is difficult to use when the camera system is deployed deep undersea. Others assume that the normal of the refractive interface is known or parallel to that of the image plane, or assume that the dimension of the calibration target is known [8]. In contrast to previous methods, our method makes no assumption on the configuration of the camera system or the refractive interface, and works on arbitrary unknown 3D scenes.

3. Proposed Method

Without loss of generality, we assume that each camera is facing the refractive interface of its own underwater housing. Notice that the two cameras sharing the same interface is a special case. Moreover, we assume that there is one or multiple refractive interfaces and they are all parallel to each other. Both 3D and 2D diagrams for this scenario are shown in Fig. 1. The diagram shows two corresponding rays coming out from the camera centers, passing through multiple interfaces, and intersecting at an object point.
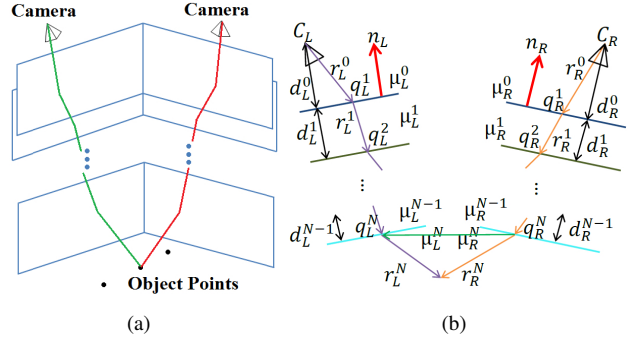

Figure 1. (a) Flat refractive geometry with multi-layer of refractive interface for a stereo camera system. (b) 2D diagram for (a).

We first describe the symbols shown in Fig. 1(b). The subscript "L" denotes the left camera view and "R" the right. Suppose there are N refractions in each camera view and the refractive indices μ_L^i and $\mu_R^i, i \in [0, N]$ are given. Here i is always an integer. The normals of the refractive interfaces are denoted as n_L and n_R. We assume that the orientations and positions of the two cameras are known by performing offline calibration in air. In other words, camera centers C_L and C_R are provided. Moreover, we assume that a set of corresponding rays coming out of the camera centers has been established. For example, $\{r_L^0, r_R^0\}$ is a pair of corresponding rays and there are M pairs. The directions of rays in each refractive interface are denoted as r_L^i and r_R^i, and the rays intersect each interface at points q_L^i and q_R^i. The thickness of the left layer is d_L^i, and of the right $d_R^i, i \in [0, N-1]$. To sum up, C_L, C_R and M pairs of corresponding rays such as $\{r_L^0, r_R^0\}$, and the refractive indices μ_L^i, μ_R^i are assumed to be known. Our method solves for n_L, n_R, d_L^i and $d_R^i, i \in [0, N-1]$ simultaneously. Once they are obtained, the 3D geometry of the scene can be reconstructed by ray-tracing from the correspondences.

3.1. Given n_L and n_R, estimate d_L^i and d_R^i

We first assume that n_L and n_R are given, which will be relaxed later in Section 3.2. We demonstrate that d_L^i and d_R^i can be obtained by solving a set of linear equations once

n_L and n_R are known. Our method is based on an important finding which is expressed as

$$(q_L^N - q_R^N) \cdot (r_L^N \times r_R^N) = 0. \qquad (1)$$

In the equation, $\cdot$ denotes dot product and $\times$ cross product. The equation states that $q_L^N q_R^N$ is on the plane formed by the two rays r_L^N and r_R^N. We can see that this condition is always true. It is known that $q_L^i = q_L^{i-1} - d_L^{i-1} \frac{r_L^{i-1}}{n_L \cdot r_L^{i-1}}, i \in [2, N]$, and q_R^i is obtained similarly. Furthermore, $q_L^1 = C_L - d_L^0 \frac{r_L^0}{n_L \cdot r_L^0}$ and a similar equation holds for q_R^1. By substituting these to Eq. 1, which can be written as

$$\Big(\big(C_L - d_L^0 \frac{r_L^0}{n_L \cdot r_L^0} - d_L^1 \frac{r_L^1}{n_L \cdot r_L^1} - \cdots - d_L^{N-1} \frac{r_L^{N-1}}{n_L \cdot r_L^{N-1}} \big)$$
$$- \big(C_R - d_R^0 \frac{r_R^0}{n_R \cdot r_R^0} - d_R^1 \frac{r_R^1}{n_R \cdot r_R^1} - \cdots$$
$$- d_R^{N-1} \frac{r_R^{N-1}}{n_R \cdot r_R^{N-1}} \big) \Big) \cdot (r_L^N \times r_R^N) = 0. \qquad (2)$$

Let $\mathcal{R} = (r_L^N \times r_R^N)$, $\mathcal{R}_L^i = \frac{r_L^i}{n_L \cdot r_L^i}$ and $\mathcal{R}_R^i = \frac{r_R^i}{n_R \cdot r_R^i}$. Eq. 2 can be further expanded as

$$\big(C_L - C_R - d_L^0 \mathcal{R}_L^0 - \cdots - d_L^{N-1} \mathcal{R}_L^{N-1}$$
$$+ d_R^0 \mathcal{R}_R^0 + \cdots + d_R^{N-1} \mathcal{R}_R^{N-1} \big) \cdot \mathcal{R} = 0. \quad (3)$$

Re-arrange the order of terms produces

$$- d_L^0 (\mathcal{R}_L^0 \cdot \mathcal{R}) - \cdots - d_L^{N-1}(\mathcal{R}_L^{N-1} \cdot \mathcal{R}) + d_R^0 (\mathcal{R}_R^0 \cdot \mathcal{R})$$
$$+ \cdots + d_R^{N-1}(\mathcal{R}_R^{N-1} \cdot \mathcal{R}) = (C_R - C_L) \cdot \mathcal{R}. \qquad (4)$$

Finally Eq. 4 can be written as

$$\big[-(\mathcal{R}_L^0 \cdot \mathcal{R}), \ldots, -(\mathcal{R}_L^{N-1} \cdot \mathcal{R}), (\mathcal{R}_R^0 \cdot \mathcal{R}), \ldots, (\mathcal{R}_R^{N-1} \cdot \mathcal{R}) \big]$$
$$\big[d_L^0, \ldots, d_L^{N-1}, d_R^0, \ldots, d_R^{N-1} \big]^T = (C_R - C_L) \cdot \mathcal{R}. \quad (5)$$

Once n_L and n_R are known, then r_L^i and $r_R^i, i \in [1, N]$ can be obtained since r_L^0 and r_R^0 are known, and the refractive indices are assumed to be given. Therefore, $\mathcal{R}_L^i, i \in [0, N-1]$ which is $\frac{r_L^i}{n_L \cdot r_L^i}$, can be calculated. Similarly, $\mathcal{R}_R^i$ and $\mathcal{R}$ can be computed as well. Up to this point, we can see that $\big[d_L^0, \ldots, d_L^{N-1}, d_R^0, \ldots, d_R^{N-1} \big]^T$ are the unknowns in Eq. 5 and all the others can be calculated. Notice that Eq. 5 has the same form as $Ax = b$ which can be easily solved. Given M pair of correspondences, the dimension of A is $M \times 2N$, b is $2N \times 1$ and x is $M \times 1$. By solving the set of linear equations, d_L^i and $d_R^i, i \in [0, N-1]$ are obtained.

3.2. Search space

The above section demonstrates that once n_L and n_R are provided, d_L^i and d_R^i can be computed by solving a set of

linear equations. After that, the 3D geometry of a scene can be obtained as well. This method can be readily incorporated into an existing underwater stereo camera system where the normals of the refractive interfaces are known. Unfortunately, some systems do not have information of the normal and this section is designed to address this problem.

Although we do not know accurately n_L and n_R, we do know that they are both inside a search space which is the hemisphere shown in Fig. 2. That is, the z component of the normal direction is always negative in the camera's coordinate system. Denote $n_L = [n_L(x), n_L(y), n_L(z)]$. Then,
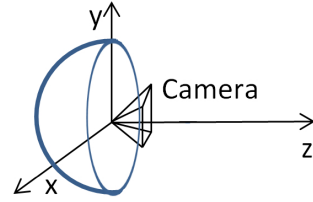

Figure 2. The search space for the normal of the refractive interfaces.

$n_L(x) \in [-1, 1], n_L(y) \in [-1, 1], n_L(x)^2 + n_L(y)^2 \leq 1$ and $n_L(z) = -\sqrt{1 - x^2 - y^2}$. Both $n_L(x)$ and $n_L(y)$ are within $[-1, 1]$ because n_L represents the normal direction and its length is 1. The search space for n_R can be defined in the same way. Since the two cameras are calibrated in air, the relationship between their coordinate systems can be obtained. As a result, the search space for both n_L and n_R can be defined.

There are many hypotheses for n_L and n_R in their search spaces, and the reprojection error is used to measure whether or not the result produced by a particular hypothesis is appropriate. To be more specific, for a certain hypothesis of n_L and n_R, we compute d_L^i and $d_R^i, i \in [0, N-1]$ by using our method in Section 3.1. After that, M pairs of corresponding rays generate M 3D points by ray-tracing, and these 3D points are projected back to the images to compute the reprojection error. It is demonstrated in [1] that by solving a 4th degree equation the reprojected pixels can be obtained for a single refraction, and a 12th degree equation for two refractions. Assuming that the measured corresponding pixels in the two images are $\hat{p}_L$ and $\hat{p}_R$, and the corresponding reprojected pixels are p_L and p_R, then the reprojection error is defined as the root mean square (RMS) error $\mathcal{J} = \sqrt{\frac{1}{M} \sum_{i=1}^{M} \big((p_L(i) - \hat{p}_L(i))^2 + (p_R(i) - \hat{p}_R(i))^2 \big)}$. We select the hypothesis of n_L and n_R that minimizes $\mathcal{J}$.

4. Implementation Details

A naive implementation of the method in Section 3.2 is a brute force search in the search space for the normals. For example, one can use a step size of 0.01 when searching the entire space for n_L and n_R. However, this implementation

is very time consuming and the results may not be accurate enough because the step size is 0.01. Instead, we use binary search in the defined search space. In particular, our implementation terminates at the 10th iteration and the step size at the ith iteration is 0.5^i. At each iteration, there are 25 hypotheses each for n_L and n_R, which result in a total of $(25)^2$ hypotheses. Therefore, our implementation explores $(25)^2 \times 10$ hypotheses and the final step size is 0.5^{10}, which is much finer than 0.01.

Since the established corresponding rays such as $\{r_L^0, r_R^0\}$ can be corrupted by noise, a post-processing step is performed. We use the sparse bundle adjustment [10] to refine the parameters n_L, n_R, d_L^i and $d_R^i, i \in [0, N-1]$.

5. Experimental Results

Extensive experiments have been performed and the results are reported here. The refractive index for air is 1.0, for water 1.33 and for glass 1.50. We assume that the light path from the camera to the object is air $\rightarrow$ glass $\rightarrow$ water. This is the most common scenario for both the lab environment and for camera systems deployed undersea. For example, water is usually contained in a tank made of glass in the lab environment. Cameras are normally placed in their own housing equipment made of glass before they are deployed undersea. From now on, we use the term "single approximation" to represent that we use a single refraction of water to approximate the refraction of glass + water. Under this circumstance, the glass thickness is not estimated.

5.1. Simulations

We test the following five cases in the simulated experiment. **Case 1**: Known refractive normal + single approximation. **Case 2**: Known refractive normal + estimate all layers' thickness. **Case 3**: Unknown refractive normal + known glass thickness. **Case 4**: Unknown refractive normal + single approximation. **Case 5**: Unknown refractive normal + estimate all layers' thickness. Our results for **Case 4** and **5** are compared with that produced by [1]. The reason for comparing only these two cases is that the method of [1] assumes that both the refractive normal and glass thickness are unknown. **Case 3** is designed because all the housing equipment is custom built and the thickness of the glass is normally known. The positions of the cameras, the normals of the refractive interfaces, the distances from the camera centers to the glass are randomly generated. Since the thickness of the glass varies for different systems, it is set to be a factor of the distance from the camera center to the glass. The factor varies from 0.1 to 1.9. The object is set to be a plane in order to easily inspect the quality of the reconstruction. It is noteworthy that our method does not require the scene to be a planar object. In our real experiments, arbitrary 3D objects are used. The size of the plane is 0.5×0.5 units. Throughout our simulated experiments, the image

resolution is 2048×1536 pixels. The number of pixel correspondences established from the object is $M = 2500$. The correspondences are generated with 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 $\hat{n}_L$ and the ground truth is n_L, then the angle between them is computed by $\arccos(\hat{n}_L \cdot n_L) \times 180°$ and termed "angular error." It is applied to n_R as well. Moreover, assume that the layer thickness recovered by our method is $\hat{d}_L^i, i \in [0, N-1]$ and the ground truth is d_L^i, then the normalized error is computed by $\frac{|\hat{d}_L^i - d_L^i|}{d_L^i}$ where $|\cdot|$ denotes absolute value. The error of d_R^i is also computed. Beside the above parameters, we also compute a 3D error as follows. Two objects are used to generate noise free correspondences for each test case. Combining the estimated parameters and the noise free correspondences the 3D models of the objects can be reconstructed. They are compared with the actual objects used. One of the objects is the plane that we used to generate the M correspondences, and the other object is the Stanford Bunny. The size of the Stanford Bunny is scaled to be around $0.5 \times 0.5 \times 0.5$ units. We denote the coordinates of each object point as P, and the coordinates computed by using our estimated parameters as $\hat{P}$. The distance between these two points is obtained and averaged for the entire object. We use D_p to denote the averaged distance for the plane, and D_b for the Stanford Bunny. We can see that all the measures indicate errors in the results. Therefore, a lower value in the curves indicates better results.

Due to the space limit, only the results of **Case 3** and **5** are shown in Fig. 4 and 5, the rest is placed in the supplemental material. The legend for these curves are the same and shown in Fig. 3. In particular, the legend shows the factor of the glass thickness relative to the distance from the camera center to the glass, which varies from 0.1 to 1.9. A 3D object reconstruction result is shown on the left of Fig. 3. In particular, the ground truth is shown in red which is overlapped with our result shown in green. The result without accounting for the refraction effects is shown in blue. It is the result for **Case 2** when there is no noise at all and the glass thickness factor is 0.1, which verifies the correctness of our method by showing that our method can recover the ground truth when the dataset is noise free. The object is a plane, and the reconstruction result is curved when the refraction is not accounted for. We compare our method with that in [1] for **Case 4** and **5**. In particular, the top two rows in Fig. 5 show our results and the bottom two show those produced by [1]. The comparison indicates that our method is comparable with that of [1], although not better than it. The major advantage of our method is that we do not have any requirement on the scene, while [1] requires the 3D geometry of the calibration object to be known.

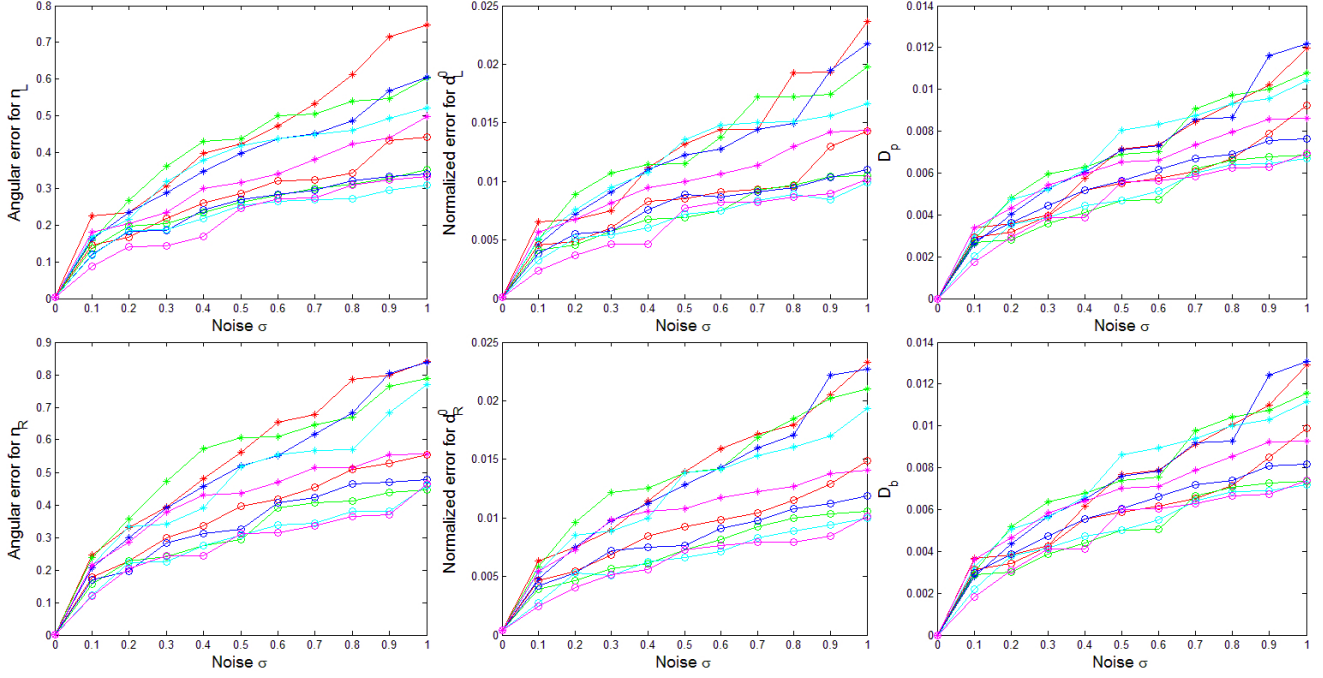Comparing all the five cases, we discovered that the av-

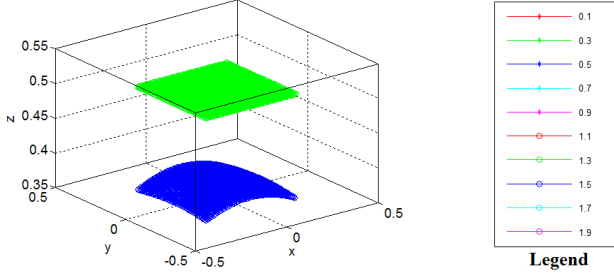Figure 4. Error for **Case 3** in the simulated experiments.

Figure 3. Left: 3D reconstruction results for a simulated experiment. Right: legend for all figures.

eraged error in the 3D model is increased from **Case 3** $\rightarrow$ **2** $\rightarrow$ **1** $\rightarrow$ **4** $\rightarrow$ **5**. Therefore, we recommend researchers to measure the thickness of the housing equipment before the camera systems are deployed underwater. Moreover, one should try to avoid the cases when both the refractive normal and the glass thickness are unknown. For example, when the glass thickness factor is 0.1 and the noise is 0.5 pixels, the error D_b is about 0.0075 unit in **Case 3** while 0.07 in **Case 5**. Considering that the size of the 3D object is about $0.5 \times 0.5 \times 0.5$, we can conclude that the error in **Case 3** is quite small (1.5%).

5.2. Simulations with Outliers

In our implementation, gray code structured light patterns [13] are projected onto the scene to establish correspondences. The reason of using structured light patterns

is that it usually produces better accuracy comparing with passive stereo methods. However, outliers could still exist in the established correspondences. RANSAC is used to detect outliers.

The experiments are designed as follows to demonstrate that our method is robust to outliers. We generate $M = 2500$ correspondences with outliers and the number of outliers varies from 0% to 30%. We select the noise variance to be 0.5 pixels and repeat the experiments for all the 5 cases described in Section 5.1. Only the result for **Case 2** is shown in Fig. 6. The curves for the other cases have a similar shape and is shown in supplemental material. Our method can handle up to 25% of outliers. In reality, the number of outliers in the established correspondences by using the gray code patterns is usually much less than 25%.
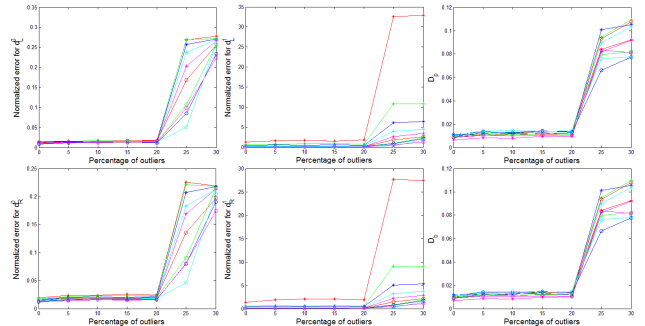

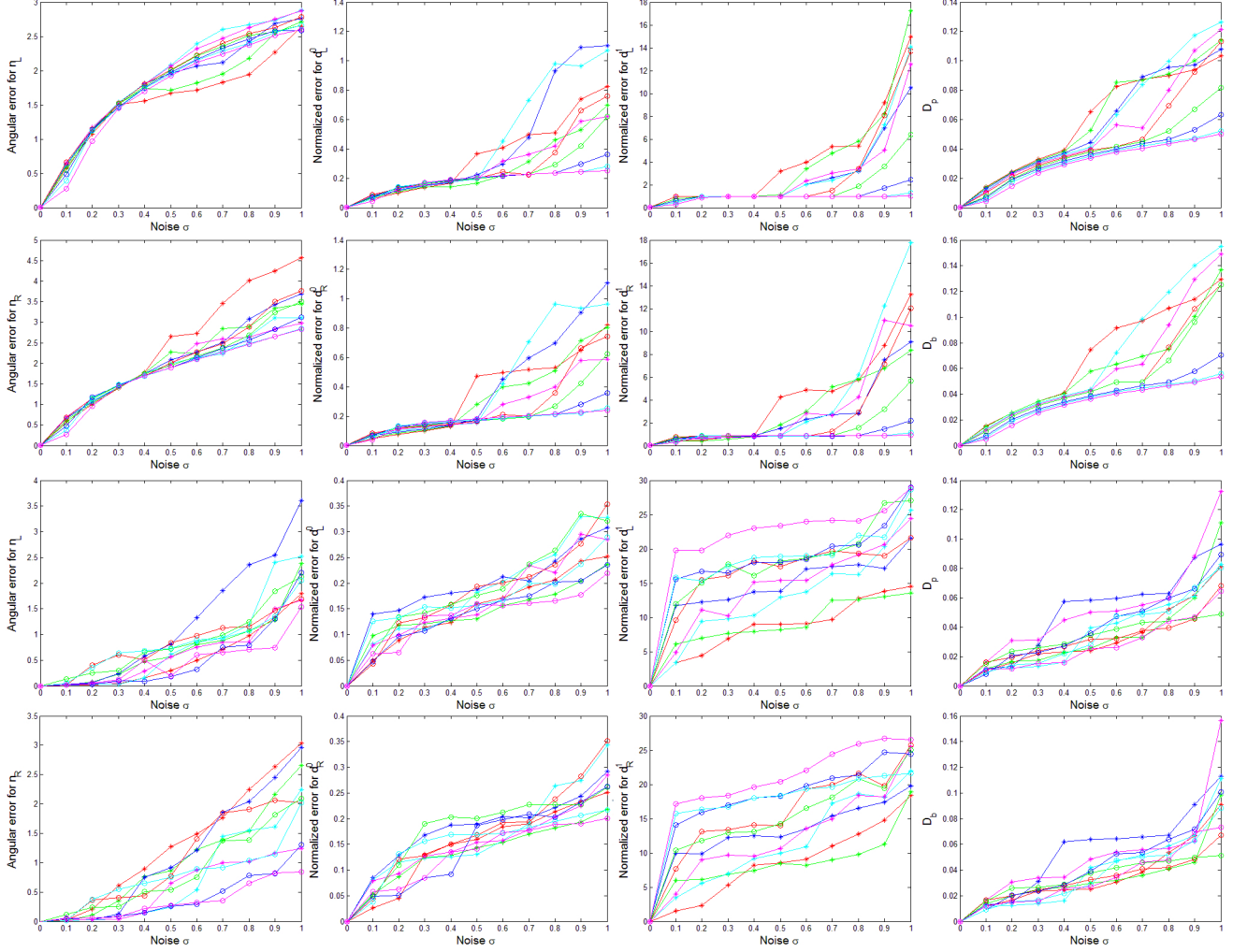Figure 6. Error for **Case 2** with outliers.

Figure 5. Error for **Case 5** in the simulated experiments. The top two rows show our results and the bottom two the results of [1].

5.3. Real Data

Besides the simulated experiments, real experiments are carried out in a lab environment and the setup is shown in Fig. 7. Two images are taken from different angles to show the details of our setup. In particular, two Canon 5D cam-
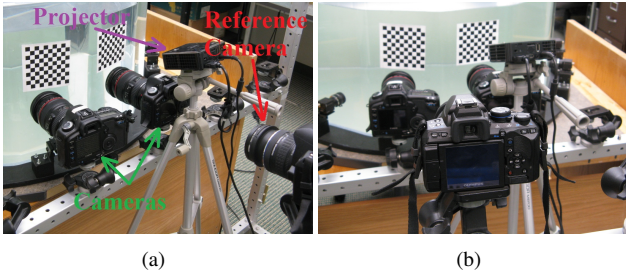

(a)   (b)

Figure 7. Setup for our real experiments.

eras (indicated by the green arrows) are placed in front of a

plexiglass tank and each camera has its own refractive interface. A projector, which is indicated by the purple arrow, is used to project gray code structured light patterns in order to establish dense correspondences between the cameras. The resolution of the cameras is 4368×2912 and of the projector is 1024×768. A reference camera (indicated by the red arrow) is used to obtain the ground truth for the refractive axis (n_L, n_R) and the distance from the two Canon cameras to the refractive interface (d_L^0, d_R^0). The reference camera focus on both checkerboard patterns. Therefore, the transformation between the coordinate systems of the checkerboard and the reference camera can be computed through calibration. After that, once the three cameras are calibrated in air, the transformation between the coordinate systems of the checkerboard and the two Canon cameras can be computed. As a result, the ground truth for the refractive axis and the distance can be obtained. The thickness of the glass (d_L^1, d_R^1) is measured by a ruler. Our results will be com-

pared with the ground truth to measure the accuracy.

Three datasets are captured under the same stereo camera configuration and all the five cases are tested. In particular, the three datasets are named "Plane", "Cave" and "Boat". Correspondences are established for each dataset and the parameters are estimated from each set of correspondences. Once all the parameters (n, d) are estimated, the 3D reconstruction can be performed. The results of "Plane" scene are shown in Fig. 8. From left to right, Fig. 8 shows the captured image, the 3D reconstruction results when refraction is not accounted for, the results of **Case 3** and **5** by applying our method. In the supplemental material, the results of all the three datasets including all the five cases are shown. The "Plane" scene includes a piece of laminated paper pasted on a glass to show the impact of the refraction effects. Since we measure the ground truth for all the parameters (n, d), the "ground truth" of the 3D reconstruction can be obtained by ray-tracing using the established pixel correspondences. The ground truth are shown in red for comparison. To demonstrate that our measurement of ground truth is accurate, we mark two points on the milk carton. The distance between these two points from 3D reconstruction by using our measured ground truth is 99.0mm, and the distance measured by a ruler is 98.7mm. The above two numbers indicates that our measurement of the ground truth for the parameters is very accurate. From the second column of Fig. 8, we can see that the objects are larger in 3D reconstruction when the refraction effects are ignored which is the case in reality. In particular, the distance between the two points marked on the milk carton is 149.8mm when refraction is not accounted for. We can see that the error is quite large compared to the one measured by ruler. Moreover, the reconstruction of the laminated paper is not flat anymore. It demonstrates that the error can be significant when the refraction effects are not accommodated in 3D reconstruction. By visual comparison with the ground truth, the results are the best in **Case 3** and worst in **Case 5**, which is consistent with our conclusion in the simulated experiments. However, the results of all the five cases are better than that of without accounting for refraction.

Besides the visual comparison of the 3D reconstruction results between the ground truth and ours, the results for all the parameters are shown in Fig. 9. Both gray code dense correspondences and SURF matches are used in our method. Although different scenes generate different correspondences, the estimated parameters are very close. We only present the result of "Cave" scene due to the space limit. The unit for the thickness is mm. Some cells are filled in with "N/A" because it is not estimated in that case. For example, n_L and n_R are assumed to be known in **Case 1** and **2**. "D" represents the 3D reconstruction error with unit in mm. We can see that even though the number of SURF matches is only about 500 pairs, it generates results

that are close to that of using dense correspondences. The results indicate that most of the parameters are well estimated except the glass thickness where large errors still exist. Therefore, we strongly recommend researchers to measure the thickness of the refractive interface when building the underwater camera systems. Moreover, a checkerboard pattern is placed in the water and the images are taken by both cameras in order to apply the method proposed in [1]. Comparing with [1], our method is comparable in **Case 4**, while it is worse in **Case 5**. Nonetheless, [1] requires the 3D geometry of the scene to be known while our method does not.

	n_L	n_R	d_L^0	d_R^0	d_L^1	d_R^1	D
GT	[-0.1540;0.0204;-0.9879]	[0.0942;0.0466;-0.9945]	123.35	186.00	5.6	5.6	0
Case 1	N/A	N/A	129.49	195.54	N/A	N/A	1.76
Case 2	N/A	N/A	126.06	189.29	66.87	67.48	1.57
Case 3	[-0.1417;0.0386;-0.9892]	[0.1121;0.0611;-0.9918]	125.31	189.27	N/A	N/A	1.03
Case 4	[-0.1753;0.0405;-0.9837]	[0.1076;0.0732;-0.9915]	129.72	196.77	N/A	N/A	4.91
Case 5	[-0.1491;0.1060;-0.9831]	[0.1386;0.1158;-0.9836]	183.30	271.63	70.77	72.90	17.43
Estimated parameters for the "Cave" scene using dense correspondences							
	n_L	n_R	d_L^0	d_R^0	d_L^1	d_R^1	D
Case 1	N/A	N/A	128.93	196.11	N/A	N/A	1.85
Case 2	N/A	N/A	126.35	191.2	55.36	68.19	1.77
Case 3	[-0.1511;0.0398;-0.9877]	[0.1169;0.0481;-0.9920]	125.54	189.75	N/A	N/A	1.37
Case 4	[-0.1862;0.0324;-0.9820]	[0.0713;0.0550;-0.9959]	128.86	195.74	N/A	N/A	4.63
Case 5	[-0.1347;0.0916;-0.9866]	[0.1277;-0.0489;-0.9906]	180.47	273.59	69.23	81.97	16.22
Estimated parameters for the "Cave" scene using SURF correspondences							
	n_L	n_R	d_L^0	d_R^0	d_L^1	d_R^1	D
Case 4	[-0.1601;0.0274;-0.9867]	[0.0982;0.0548;-0.9937]	134.65	202.36	N/A	N/A	4.37
Case 5	[-0.1438;0.0299;-0.9892]	[0.1039;0.0561;-0.9930]	151.23	233.62	95.52	89.97	8.20
Estimated parameters by Agrawal et al. [1]							

Figure 9. Comparison between the estimated parameters by our method and the ground truth. "GT" denotes the ground truth.

6. Limitation

When solving Eq. 5 which has the same form as $Ax = b$, an important assumption is that A has to be full rank. Moreover, any two rays r_L^i and r_L^j cannot be parallel. Similar assumption is applied to r_R^i. The second assumption infers that any two layers cannot be the same material. The following two solutions are proposed to address such limitation. (1) If there are layers of the same material, the sum of the thicknesses for the same material layers can be computed. However, the thickness for each layer cannot be obtained separately. (2) If the layer material is not known, we can compute the rank of A. After that, the sum of the thicknesses for the layers of the same materials can be computed.

7. Conclusion

We have presented a new method to recover the normal and thickness of the refractive interface for an underwater stereo camera system. We first demonstrate that the thickness of the interface can be estimated when the normal is given. Then, we show that both the normal and the thickness of the interface can be estimated by exploring the search space for the normal and using the reprojection error as an evaluation metric. Simulated and real experiments are performed to evaluate the correctness and robustness of our method. All the experimental results are compared to
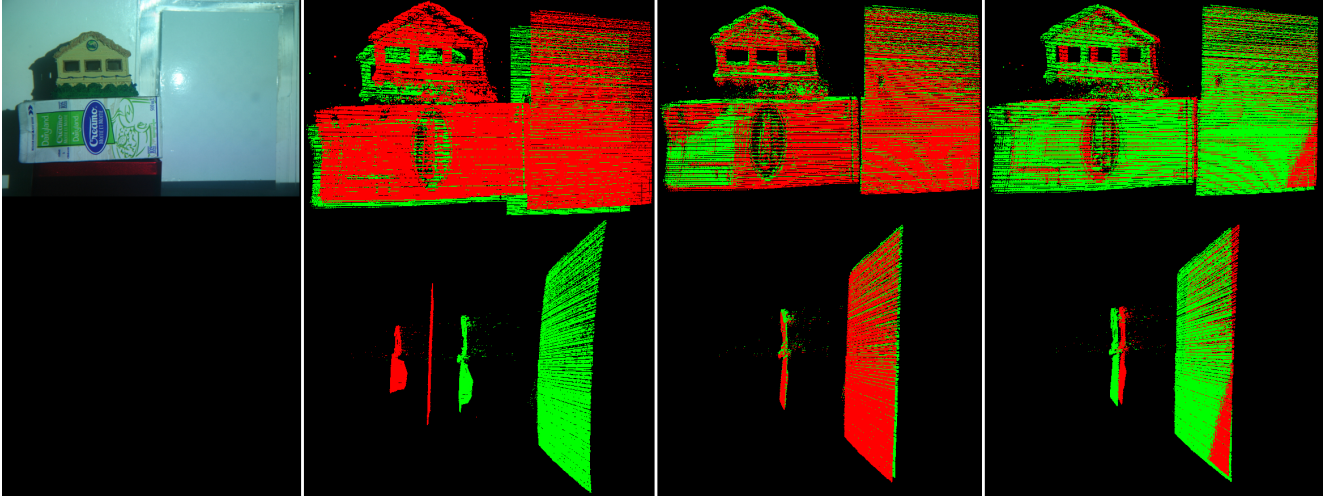
Figure 8. Results of the "Plane" scene. From left to right: captured image, 3D reconstruction results for the case when refraction is not accommodated, results of **Case 3** and **5** using the proposed method. The two rows shows the 3D reconstruction results in two different views. The ground truth is shown in red for comparison.

the ground truth, and some of them are compared with the state-of-the-art method [1]. The results indicates that the accuracy of our method is promising and that our method is practical since it works with any arbitrary 3D scene.

Future work includes three directions. First, we would like to exam the reprojection error in the normal search space to find a better optimization strategy instead of using the binary search which could be time consuming when extent to multiple view. Second, our method assumes that the relative pose between the two cameras is known by off-line calibration in air. We plan to relax such an assumption. Third, the outlier handling step can be further improved. Notice that the number of unknown parameters is only 2/4 for one/two refractions in Eq. 5. Therefore, the minimum number of correspondences required is 2 or 4. In this case, the RANSAC can run faster and handle more outliers.

Acknowledgements

References

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

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

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

[4] R. Ferreira, J. Costeira, and J. A. Santos. Stereo reconstruction of a submerged scene. In *IbPRIA*, pages 102–109, 2005.

[5] A. Jordt-Sedlazeck and R. Koch. Refractive structure-from-motion on underwater images. In *ICCV*, pages 57–64, 2013.

[6] 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. *Applied Optics*, 51(31):7591–7603, 2012.

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

[8] C. Kunz and H. Singh. Hemispherical refraction and camera calibration in underwater vision. In *OCEANS*, 2008.

[9] J.-M. Lavest, G. Rives, and J.-T. Lapreste. Underwater camera calibration. In *ECCV*, pages 654–668, 2000.

[10] M. I. A. Lourakis and A. A. Argyros. The design and implementation of a generic sparse bundle adjustment software package based on the levenberg-marquardt algorithm. Technical report, 2004.

[11] O. Pizarro, R. Eustice, and H. Singh. Relative pose estimation for instrumented, calibrated imaging platforms. In *Proc. of DICTA*, pages 601–612, 2003.

[12] J. P. Queiroz-Neto, R. L. Carceroni, W. Barros, and M. F. M. Campos. Underwater stereo. In *SIBGRAPI*, pages 170–177, 2004.

[13] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *CVPR*, pages 195–202, 2003.

[14] A. Sedlazeck and R. Koch. Calibration of housing parameters for underwater stereo-camera rigs. In *BMVC*, pages 118.1–118.11, 2011.

[15] T. Treibitz, Y. Y. Schechner, and H. Singh. Flat refractive geometry. In *CVPR*, 2008.

[16] T. Yau, M. Gong, and Y.-H. Yang. Underwater camera calibration using wavelength triangulation. In *CVPR*, 2013.

[17] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on PAMI*, 22(11):1330–1334, 2000.