Underwater Active Oneshot Scan with Static Wave Pattern and Bundle Adjustment

Hiroki Morinaga¹, Hirohisa Baba², Marco Visentini-Scarzanella¹, Hiroshi Kawasaki¹, Ryo Furukawa², and Ryusuke Sagawa³

¹ Kagoshima University, Japan
² Hiroshima City University, Japan
³ National Institute of Advanced Industrial Science and Technology, Japan

Abstract. Structured Light Systems (SLS) are widely used for various purposes. Recently, a strong demand to apply SLS to underwater applications has emerged. When SLS is used in an air medium, the stereo correspondence problem can be solved efficiently by epipolar geometry due to the co-planarity of the 3D point and its corresponding 2D points on camera/projector planes. However, in underwater environments, the camera and projector are usually set in special housings and refraction occurs at the interfaces between water/glass and glass/air, resulting in invalid conditions for epipolar geometry which strongly affect the correspondence search process. In this paper, we tackle the problem of underwater 3D shape acquisition with SLS. In this paper, we propose a method to perform 3D reconstruction by calibrating the system as if they are in the air at multiple depth. Since refraction cannot be completely described by a polynomial approximation of distortion model, grid based SLS method solve the problem. Finally, we propose a bundle adjustment method to refine the final result. We tested our method with an underwater SLS prototype, consisting of custom-made diffractive optical element (DOE) laser and underwater housings, showing the validity of the proposed approach.

Keywords: Underwater scan, Camera-Projector system, One-shot scan

1 Introduction

Structured Light Systems (SLS) are widely used in various applications such as augmented reality, medical examination, games, movies, etc. A typical SLS consists of a camera and a projector. Usually the projector projects an encoded pattern onto an object’s surface, and the images of the object captured by the camera can be easily decoded through knowledge of the projected pattern. Because the technique uses an active pattern projector, the correspondence searching process becomes much easier than passive stereo techniques due to the uniqueness and distinctiveness of the features in the pattern. Therefore, a higher reconstruction accuracy and density can be achieved with active stereo systems, and SLS has become one of the most important non-contact
3D shape measurement methods [18, 5]. In particular, one-shot SLS is one of the main techniques in active scanning of dynamic environments [15, 2].

The most critical factor affecting the accuracy of an SLS is calibration. Before SLS reconstruction of an object by using a fixed pattern, the intrinsic parameters of the camera and projector as well as the extrinsic parameters relating them should be estimated. After calibration, reconstruction is done from epipolar geometry: the epipolar lines corresponding to the feature points detected on the camera images can be drawn on the fixed pattern image which is projected by the projector. Finally the correspondences can be found by searching along these epipolar lines, and 3D reconstruction performed by triangulating the corresponding points [12].

When the SLS operates within an air medium, the correspondence problem can be solved efficiently by the valid epipolar geometry due to the co-planarity of the 3D point and its corresponding 2D points on camera/projector planes as shown in Fig. 1(a). However, in an underwater environment, the camera and projector (regarded as an inverse camera) are usually set in special housings [17]. Since refraction occurs at the interfaces between water/glass and glass/air, the co-planarity condition is not enforced anymore, as shown in Fig. 1(b). Fig. 2(a) and (b) show the line of sight of the camera in the air and underwater, respectively. And thus, it is necessary to find an efficient way to calibrate an underwater SLS, and to ensure that the epipolar assumptions can hold.
This paper proposes three approaches to jointly tackle the aforementioned issues. First, we introduce a depth-dependent calibration method that uses a polynomial approximation model for the SLS for underwater environment. Second, to solve the problem that the epipolar geometry is only approximately valid underwater, we introduce a grid-based active scanning method (specifically, a wave grid pattern) which allows to find correspondences that stray away from the epipolar line while still maintaining a good matching performance. Since the results are based on approximate model, we also introduce a refinement algorithm based on bundle adjustment which uses the wave reconstruction results as the initial parameters to achieve high accuracy. We demonstrate the effectiveness of the proposed approach with simulation as well as a real system with a special housing of camera and pattern projector placed underwater in a pool tank.

2 Related Work

Calibration models for underwater camera have been proposed extensively \[14, 3, 13, 1, 17, 10, 7, 6, 8\]. However, none of them gives an entire calibration and reconstruction procedure for an SLS. Because of the correspondence matching problem in SLS, some of the proposed models becomes invalid since the formulated models do not offer a practical strategy for matching and reconstruction. Besides, projector calibration underwater is also a slightly different issue than camera calibration due to the "blindness" of the projector [4].

There are some early works for underwater 3D reconstruction based on approximation model [14, 3, 13]. Queiroz-Neto et al. proposed an underwater model which simply ignores the effects of the refraction, but earns results with low accuracy due to the non-linear refraction effect [14]. Some approximate methods also have been proposed, such as focal length adjustment [14], lens radial distortion approximation [3] and a combination of the two [13]. Unfortunately, the accuracy of these approximation models are also insufficient to an SLS system for correspondence search using epipolar geometry.

To improve the accuracy of underwater measurement, some physical models for camera calibration and reconstruction have been proposed \[1, 17, 10, 7, 6, 8\]. Agrawal et al. gives a general calibration method for underwater cameras, based on a physical refractive model [1]. They consider that all refractive planes are parallel to each other, and they derive front-projection and back-projection equations for their refractive model. However, it is necessary to solve 4th degree equations even for one refractive plane’s case, and 12th degree equations in the 2 plane case in a forward projection situation, and thus, it is difficult to use this method directly for SLS. Sedlazeck et al. focus on the underwater light rays which are projected as a curved surface: after learning this surface, perspective projection can be done [17]. According to this method, it is also difficult to tackle the forward projection problem due to the complicated learning phase. Kang et al. and Sedlazeck also consider the underwater reconstruction with Structure from Motion (SfM) [7, 6]. SfM is a passive way to recover 3D shape of objects, and it is difficult to achieve a dense reconstruction result due to the difficulty of the correspondence searching. Kawahara et.al. proposed pixel-wise varifocal camera model, where the focal length of the projection varies pixel-by-pixel, for modeling non-central projection of an underwater camera, and a calibration method for the cameras [8]. They
also proposed an active-stereo system composed of a projector and two cameras, where projection of the cameras and the projector is based on their model [9]. Since image-based correspondence search using epipolar lines are not valid for underwater cameras, they applied space carving method, where only photo-consistency is needed.

In terms of SLS for underwater, Campos et al. proposed an underwater active stereo system that uses a DOE-based pattern projector [11]. They used a pattern of parallel lines and each line is not coded into local features. Their decoding method (i.e., the method for solving correspondences between the captured image and the projected pattern) relies on the order between the detected lines on the camera image, thus, ambiguity may occur if only a small region of the pattern is detected.

3 Overview

3.1 System configuration

In this research, we set up a camera-projector system. The camera and projector are set into housings, respectively. The actual configuration is shown in Fig. 3. We made a waterproof housing as shown in Fig. 4(a). Left and right housings are for the cameras, while the center housing is for the laser projector with a diffractive optical element (DOE) of a wave pattern (Fig. 4(b)). Our choice of using two cameras stems from the following reasons:

1. With two cameras and the appropriate baseline it is possible to reconstruct areas occluded in one view, thereby reconstructing a much wider area than with conventional monocular active sensing.
2. By using multiple cameras, our system is equivalent to multi-view stereo, so its accuracy can be further improved with Bundle Adjustment.

![Fig. 3. Set up a camera-projector.](image-url)
3.2 Algorithm

We adopt a coarse to fine approach for reconstruction. First, the approximated model is used to perform the wave grid reconstruction to retrieve coarse shape. Then, the estimated 3D points are used as initial values for Bundle Adjustment refinement using an accurate non-central projection camera model, which takes into account the refractive environment. The reason why we need the approximation model for the coarse level is that a central projection model does not work in the underwater environment, that means epipolar constraint does not work, however, the epipolar constraint is a key to efficiently find the correspondences with active stereo techniques. Certainly approximation errors inevitably occur at the coarse level, however, those are corrected during the refinement process. Furthermore, there is no practical problem if the deviation of the initial model from the actual model is within the tolerance of the epipolar matching to still produce the correct match.

3.3 Polynomial Approximation of Refraction

Problem statement Before introducing our polynomial approximation model for refraction, let us consider the problem when we perform underwater reconstruction with a full physical refraction model. To simplify the model, we only consider the forward-projection considering one refractive layer introduced in [1]. We suppose that a camera and a projector are all set into housings respectively, and assume that the housings’ thicknesses can be ignored. Fig. 5(a) shows the camera model. Coordinate $x$ shows the refractive plane, and the refractive indices of the media above and below this plane are $\mu_1$ and $\mu_2$ respectively. The blue line shows a ray coming from a 3D point $b$, and refraction will occur on the intersection with the plane at point $p_1 = (x_1, 0)$. $d$ is the distance between the $b$ and camera plane, $x_b$ is the distance between $b$ and the optical axis $z$. $y_c$ is the focus of camera. The angle of the incidence is supposed as $\alpha$ and the
refracting as \( \beta \). Based on Snell’s law, the following equations are obtained.

\[
\frac{\sin \alpha}{\sin \beta} = \frac{\mu_2}{\mu_1} = n, \tag{1}
\]

\[
\sin \alpha = \frac{x_1}{\sqrt{x_1^2 + y_1^2}}, \sin \beta = \frac{x_b - x_1}{\sqrt{(x_b - x_1)^2 + d^2}}. \tag{2}
\]

After some manipulation, the next equation can be obtained,

\[
(n^2 - 1)x_1^4 + (-2x_b n + 2x_b)x_1^3 \\
+ (x_b^2 n + y_c^2 n - x_b^2 - d^2)x_1^2 - 2x_b y_c^2 n x_1 + x_b^2 y_c^2 n^2 = 0. \tag{3}
\]

By solving this 4th order equation, the corresponding epipolar line on the plane of projector pattern with predefined depth can be calculated for each feature points.

**Polynomial Approximation Model** We propose a polynomial approximation of the full physical refraction model. As shown in Fig. 5(b), we consider two kinds of light paths. The blue arrows show the light paths which are outgoing from a 3D point, going through a water-air interface, reflected on the surface of an object, and finally going into a camera. The red ones show the same light paths as if through an air medium. Blue and red light paths are outgoing from \( b \), and incoming to the center of a camera \( a \). \( p_1 \) is the intersection point of the blue ray and camera plane, and \( p_2 \) is the intersection point of the red ray. The most important factor in the polynomial approximation model is the distance between these \( p_1 \) and \( p_2 \), which is defined as our approximation error. The relationship between the error and \( p_1 \) is defined as the following equation.

\[
E(p_1) = \alpha_1 r^2 + \alpha_2 r^4 \tag{4}
\]

Although only the x-z plane is drawn in the Fig. 5(b), the same applies to the y-z plane. The \( r \) in Equation 4 represents the 2-dimensional Euclidean distance between the center of the camera and \( p_1 \) in the 3D coordinate axis \( xyz \). During the calibration phase, not only the extrinsic parameters, but also polynomial approximation parameters \( \alpha_1 \) and \( \alpha_2 \) are estimated. The pinhole projection can then be represented with the equation (4) with the approximation model.

Using the calibration parameters which are calibrated under water, reprojection of 3D points to captured image plane can be done by the following manner. First, 3D points are converted to the camera coordinate system by using the extrinsic parameters, which contains the rotation and translation information, and then, reprojected to the camera plane by using the intrinsic parameter of the camera. Then, the 2D coordinates are further distorted by the equation (4). Note that the process is completely same as the ordinary camera reprojection process other than the parameters are estimated under water, and, unlike the air environment, it contains an inevitable error derived from the approximation model, and thus, it should be solved in the further step.
4 Depth dependent calibration

4.1 Overview of the calibration process

First, the camera and projector are put into their respective housings, and placed into a pool filled with water. After that, the intrinsic parameters of the camera are estimated with a checkerboard [19]. Then, the intrinsic parameters of the projector and the extrinsic parameters between them are estimated by a second calibration using a sphere of known size, described in the next section.

Since the effect of refraction is depth dependent, we conduct the calibration at the multiple depth in the paper. From the multiple calibration results, it is possible to represent the refraction effect with several hyper parameters. However, we take another solution to cope with a depth dependent effect in the paper for simplicity and leave the hyper parameter estimation approach for our future task. In order to retrieve a discrete set of depth-dependent calibration parameters, we put the calibration objects, i.e., checker board planes and sphere, at multiple depth and conduct calibrations independently.

For the selection of the best parameters, the residual errors of epipolar constraints are used. To achieve this, the 3D reconstruction process is conducted for all the parameter sets independently. The sum of residual errors of the correspondences, which are normally errors of epipolar constraints, is calculated and used for the selection of the best result.

4.2 Sphere based projector calibration

For sphere-based calibration, images are captured with the pattern projected onto the sphere as shown in Fig. 6(a). The radius of the spherical surface is known. From the image, points on the spherical contour are sampled. Also, the correspondences between the grid points on the camera image and the grid points on the projected pattern are assigned manually.
For the calibration process, we minimize reprojection errors between the imaged grid points on the sphere and the simulated grid positions, with respect to the extrinsic parameters, the intrinsic parameters of the projector, and the position of the calibration sphere. Fig. 6(b) shows how the simulated grid positions are calculated. From a grid point (for example, \( g_{p1} \) in Fig. 6(b)) of the projector, the grid projection on the sphere \( (g_{c1}) \) is calculated by ray-tracing, and is projected to the camera \( (g_{i1}) \). If the ray of the grid point does not intersect with the sphere (for example, \( g_{p2} \)), we use intersection of the ray with an auxiliary plane \( (g_{c2}) \) that is fronto-parallel and includes the sphere center.

Other than the reprojection errors, points on the spherical contour are also used for the optimization. The line of sight of a contour point \( (\tilde{s} \ c) \) should be tangent to the sphere in the 3D space; thus, the distance between the spherical center \( (c) \) and the line should equal to the sphere radius \( (r) \). Thus, the difference between the distance from the spherical object and the radius \( (\sqrt{\|c\|^2 - (\tilde{s} \cdot c) - r}) \) is also considered to be an error. Thus, the sum of squares of these errors is minimized by using Levenberg-Marquardt method.

### 5 3D Reconstruction

#### 5.1 Wave grid reconstruction

For 3D reconstruction, it is necessary to find matches between points on the image plane and the known projector pattern. In our method, we use a “wave pattern” because of the distinctiveness and uniqueness of its features and its reconstruction density [16]. Fig. 7 is an example of the pattern. The correspondences are found through an epipolar search. During the search, the impact of our polynomial approximation on accuracy is limited since the interval between intersections in the wave grid is much larger than the pixel width, and an error of a few pixels does not affect the correspondence search. This feature is important for our underwater scanning method because the polynomial approximation inevitably will create some errors on the epipolar lines, and depth dependent calibration parameters is conducted only with a sparse set of depth values.
Since the reconstructed results have some errors because of approximation model and inconsistent shapes because of depth dependent calibration parameters, those errors are effectively solved in the refinement process.

Fig. 7. (a) Corresponding point, (b) Epipolar line for (a).

5.2 Refinement with bundle adjustment

Refinement of 3D shape as well as camera and projector parameters will be conducted by the following way. We set 3D points and a position of the glass between air and water as parameters to be estimated with bundle adjustment. In terms of the position of the glass, it is described with four parameters consisting of a surface normal and a distance between camera center and surface of the glass. Since we can retrieve a hundred of corresponding points between camera and projector image through the wave reconstruction process, we can calculate the reprojection error by simply solving the fourth order polynomial equation (3). Leaven-Marquardt algorithm is used to minimize the error.

The main differences from the ordinary bundle adjustment algorithm, which is used for structure from motion or multiple view stereo method, and ours are two folds. First, we use fourth order polynomial equation to calculate 2D coordinates on the image plane back-projected from 3D points considering a refraction between water and air. Second, we include the rigid transformation parameter of the interface plane between water and air to be estimated in the bundle adjustment process.

Since we can start the optimization from the initial shape calculated by the approximated model, it converges quickly with Leaven-Marquardt algorithm with our implementation. It should be noted that, since the images are undistorted by the parameter of approximation model in underwater environment to retrieve the initial shape, the image is needed again distorted by the approximation parameters and undistorted by ordinary distortion parameters which are estimated by openCV in the air before the bundle adjustment.
6 Experiments

6.1 Depth dependent calibration

The experimental environment is shown in Fig. 8. Two Point Grey Research Grasshopper cameras and a DOE laser projector were used. Then, the camera-projector system was placed underwater, and calibrated several times with multiple depth with the proposed technique. Fig. 9(a) shows the example of captured image for our sphere calibration. Two depth positions are considered, as the near range, 1m from the camera 0 and as the far range, 1.5m. The reason why we made calibration with such few positions is that the assumed depth range was not so wide based on the measurement environment we applied. As wave grid reconstruction applied in our method works using the epipolar constraint, erroneous reconstruction occurs when the projection error turns to be above our matching tolerance. However, such a problem didn’t occur under our experimental environment in the paper, because the deviation from the assumed model was still within our tolerance. Note that possibly erroneous connection of grids didn’t occur because wave grid reconstruction has some effect to correct the errors by using grid pattern connections. After acquiring intrinsic and extrinsic parameters of the system for each depth, the 3D shape of the sphere for calibration is reconstructed to verify the result of the calibration as shown in 1st row of Fig. 9. We can confirm that sphere is correctly reconstructed with the approximation model. It can be observed that two reconstructed shapes from two cameras are apart because they are independently calibrated and reconstructed, and such inconsistency will be efficiently eliminated by our refinement algorithm.

![Fig. 8. Experimental environment of underwater scan.](image)

6.2 Wave oneshot reconstruction

Then, we captured and reconstructed the 3D shape of a mannequin using wave reconstruction. 2nd row of Fig. 9(a) shows the example of captured image and Fig. 9(b) and (c) shows a reconstruction results. We can confirm that the complicated shapes are correctly recovered with our technique. Since two cameras are used for our system and
both cameras are calibrated independently, reconstructed shapes for each camera do not coincide as we can see in the results. Such gap can be eliminated and merged with our refinement algorithm.

![Image of reconstructed shapes](image)

(a) Captured image  (b) Depth = 1.5m  (c) Depth = 0.8m

**Fig. 9.** Reconstruction results of the mannequin at different depth. White shape is reconstructed by left camera and red shape is reconstructed by right camera.

### 6.3 Evaluation of refinement algorithm

First, we checked the effectiveness of the optimization method with simulation data. We assume underwater environment and emit 7*10 points from a virtual projector to a board of 2m ahead. Then, we synthesize the image with virtual camera and conduct reconstruction with approximation model. Using the predefined parameters and the synthesized image, we conduct refinement algorithm and results are shown in Fig. 10 and 11. From the results, we can confirm that the reprojection error calculated by solving the fourth order polynomial equation considering a refraction decreases with our bundle adjustment algorithm and correct shapes are reconstructed.

Finally, we optimize the reconstruction result of the board which is captured by our underwater scanning system. We project the wave pattern on to the planer board, sphere and mannequin and first restored the shape with approximated model, and then, the shape was refined by our bundle adjustment algorithm. As shown in Fig. 12, we can confirm that the reprojection error is drastically decreased and Fig. 13 right two columns (green colored shapes) show that the our refinement algorithm successfully merged two shapes (left two columns, red and white colored shapes) into single consistent shape.

For quantitative evaluation, we calculate RMSE for the planer board by fitting the plane to the board and it was drastically decreased from 9.7mm to 0.7mm, confirming the effectiveness of our algorithm.
7 Conclusion and Future Work

In this paper, we propose a practical oneshot active 3D scanning method in underwater environment. To realize the system, we propose three solutions. First, we calibrate the camera and projector parameters with polynomial approximation for multiple depth. Then, shapes are reconstructed by wave reconstruction which allows inevitable errors in epipolar geometry. Finally, 3D shapes are refined by the bundle adjustment algorithm which calculates the actual 2D position on the image plane by solving the fourth order polynomial of physical model. Experiments are conducted with simulation and real environment showing the effectiveness of our method. Temporal constraint to recover moving object under water environment is our future work.

References

Fig. 12. Reprojection of camera and projector image after optimization. 1st row: Before optimization. 2nd row: After optimization. Red points are observed points and green points are reprojected points. Both errors in left and right camera are decreased from initial positions, so that the total error is drastically reduced by LM algorithm.


Fig. 13. Reconstructed 3D shape (Red: initial position for left camera, white: initial position for right camera and green: final shape after optimization). 1st row: Planer board. 2nd row: Ball. 3rd row: Mannequin.