2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Two-stage pose optimization algorithm using color information for underwater SLAM with light-sectioning-based 3D scanning method

Takaki Ikeda[1], Takafumi Iwaguchi[1], Diego Thomas[1], Hiroshi Kawasaki[1]

Abstract— **The demand for 3D shape measurement of underwater scene is increasing in various applications. Especially, simultaneous localization and mapping (SLAM) technique utilizing remotely operated vehicle (ROV) attached with 3D sensors has been intensively researched. This paper focuses on solving pose optimization problem for underwater robots with camera/multiple-line-lasers setup, especially for the scene with some textures (color information). To this end, a two-stage pose optimization technique is proposed. In the first stage, due to the sparse nature of the reconstructed shape in the light-sectioning method consisting of several 3D curves, we bundle 10 to 20 consecutive frames to form a block shape, refining significant errors in the initial sensor poses using a novel bundle adjustment algorithm. In the second stage, remaining pose errors are corrected by a block-based matching algorithm utilizing iterative closest point (ICP) algorithm with color information. Through experiments in underwater environment with a real system, it was validated that the proposed method demonstrates superior performance compared to past underwater SLAM techniques.**

I. Introduction

Recently, there has been an increasing demand for high-density, high-precision and large-scale 3D shape measurement of underwater scene, which are utilized for applications such as a seabed mapping, marine resource and biodiversity surveys, and port facility inspections. While acoustic sensors like sonar are commonly used to measure 3D information of underwater scene, they are generally recognized for their low accuracy, low density, and high cost. On the other hand, image sensors, although highly accurate on land, experience a significant decrease in accuracy underwater when the distance between the sensor and the target object increases, mainly due to attenuation and scattering effects. Recently, Simultaneous Localization And Mapping (SLAM) techniques employing image sensors mounted on Remotely Operated Vehicles (ROVs) to scan wide areas of the seabed by crawling near its surface have drawn wide attention. Among wide variety of 3D reconstruction algorithms utilizing image sensors, the light-sectioning method, comprising of line lasers and cameras, is frequently used due to its simple configuration and robustness [1], [2], [3], [4] (details of the light-sectioning method will be explained in Sec.II-B).

In these systems, the reconstructed point cloud obtained through light-sectioning method is integrated using poses estimated by acoustic sensors like the Doppler Velocity Log (DVL). However, due to the lack of an optimization process ensuring consistency between these two sensors, there are no means to correct errors in the estimated poses. Additionally, light refraction in underwater environments complicates camera modeling, leading to the development of several approximation models, such as those based on fisheye distortion. However, the point clouds reconstructed from these models inevitably contain errors [4]. These factors lead to a decrease in accuracy of the reconstructed shape and the risk of residual misalignments in textures. In this paper, we propose a two-stage pose optimization method that utilizes a physically accurate camera model as well as texture (color) information of the underwater scene.

In the first stage, since the reconstructed shape of the light-sectioning method consists of several 3D curves, which are sparse, we first bundle 10 to 20 consecutive frames to form a block shape, which is dense, ensuring the extraction of image features. Then, bundle adjustment is applied within a block, where image features as well as geometric constraints are used. Note that both 3D shape and color information cannot be retrieved simultaneously, since the laser pattern is projected onto the object, and thus, a novel bundle adjustment algorithm specialized for the light-sectioning method is proposed. In the second stage, Iterative Closest Point (ICP) algorithm [5] using color information [6] is applied to eliminate remaining pose errors. The Color-ICP [6] is employed between each block and the entire shape to accomplish global optimization while preserving the color of each point cloud. However, the color of each point will change when the camera poses change, which process is necessary for actual optimization. In addition, as the color recalculation process cannot be differentiated, we alternate the process of acquiring colors of reconstructed points and the Color-ICP until convergence.

Our contributions are as follows:

- A two-stage optimization method to refine sensor poses and 3D shapes by using both color and shape information specialized for light-sectioning method was proposed.
- A physically accurate camera model has been introduced to underwater SLAM to achieve higher accuracy.
- Comprehensive experiment in simulated and actual underwater environments were conducted to compare our proposed method with previous methods.

II. Related Works

A. Underwater Camera Calibration

In aerial settings, cameras are typically modeled by a perspective projection (or central projection) model, incorporating lens distortion. However, in an underwater environment,

[1] Department of Advanced Information Technology, Kyushu University

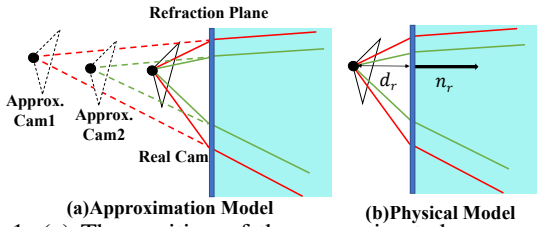


Fig. 1: (a) The position of the approximated camera cannot be determined unless manually chosen. Once fixed, it can only represent a limited area in a space. (b) Physical camera model can represent all of the space correctly.

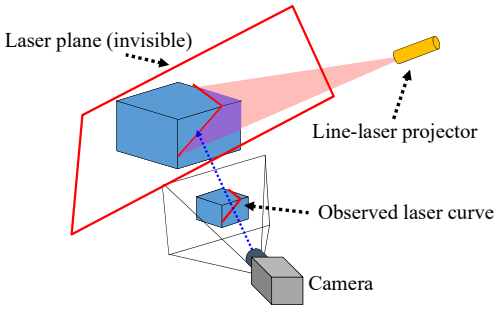


Fig. 2: Light-sectioning method

representing cameras by this model becomes impractical due to the refraction between water and air. Previous research mainly utilized two models for underwater camera representation: an approximation model and a physical model as shown in Fig.1. The approximation model [4] mitigates refractive effects between water and air by incorporating lens distortion, approximating the optical model of the camera within the water-housing as a perspective projection model as shown in Fig.1(a). Its advantage lies in its simplicity, allowing the direct application of most 3D shape measurement methods used in the air without modification. However, the drawback is evident when compensating for the refraction phenomenon using a common lens distortion model only near a predetermined depth. This limitation results in a significant degradation of 3D shape measurement accuracy when the scene depth deviates from the predefined depth [2]. On the other hand, a physical model elucidates the refraction effect of planar housing glass between air and water based on Snell's law, providing a rigorous representation of all optical phenomena through mathematical calculations as shown in Fig.1(b). A calibration method based 4D equation, which is thus not practical, has also been proposed [7].

B. Underwater 3D shape measurement using lasers

Major 3D shape measurement techniques for underwater scenes have predominantly been acoustic-based, using methods like sonar [8], [9]. Nonetheless, due to the inherent limitations in sonar accuracy, laser-based techniques have garnered substantial research attention and proposals [10], [1], [11], [12], [2]. Fig.2 shows the basic principle of the method, where 3D positions of the observed laser curves are reconstructed by triangulation. Recently, they have been extended to be self-calibrated or multiple line-laser projections [13], [14], [15], [16], [17]. Since there is strong

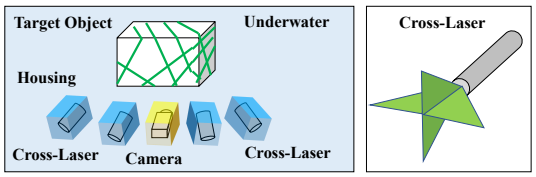


Fig. 3: System configuration of the proposed system. Multiple-line lasers and a camera are installed in a waterproof housing and are tightly fixed together.

attenuation in the water, most of the techniques based on a light sectioning method adopted a single line-laser projector to concentrate energy along a single line, which results in sparse shape reconstruction, such as a single curve for each frame [1], [12]. Several methods have been proposed to increse the density, such as parallel lines [11] or grid lines [2]. Recently, in pursuit of achieving high-density and wide-area 3D shape measurements, a SLAM technique utilizing Remote Operative Vehicle (ROV) to move and accumulate the scanning results has been introduced [1], [12], [3], [4]. In the method [4], the light sectioning approach is applied to each cross-line laser, and the shapes of laser-illuminated areas for each frame are integrated using the camera pose estimated via Direct Sparse Odometry (DSO) [18].

C. Camera Pose Estimation

Camera pose estimation without explicit calibration has been extensively researched since only an image set is necessary, making it applicable to many scenarios. Structure From Motion [19], [20], [21] and Visual Odometry (VO) [22], [23], [24] are well-known methods for achieving this goal using image sequences. However, their robustness diminishes when only a few feature points are detected. Recently, Direct Sparse Odometry (DSO) [18] was proposed, which demonstrates relative robustness in texture-less scenes. Nevertheless, it still struggles in underwater environments where limited texture is present due to murky water. We propose a method to optimize low-accuracy estimated poses using the reconstruction results obtained through light-sectioning.

III. Line-laser 3D scan with ROV System

A. 3D shape Measurement system configuration

As shown in Fig.3, the 3D shape measurement system in the proposed method consists of one RGB camera and multiple mono-color cross-line laser projectors, each of which is placed in a housing with a planar boundary surface. The relative pose between the camera and each planar laser is fixed, and camera's intrinsic parameters are calibrated in the air in advance. We use central projection camera model in the paper. Then, calibration of the refraction plane of the camera housing and the external parameters of each laser plane, and 3D reconstruction by light sectioning method are conducted.

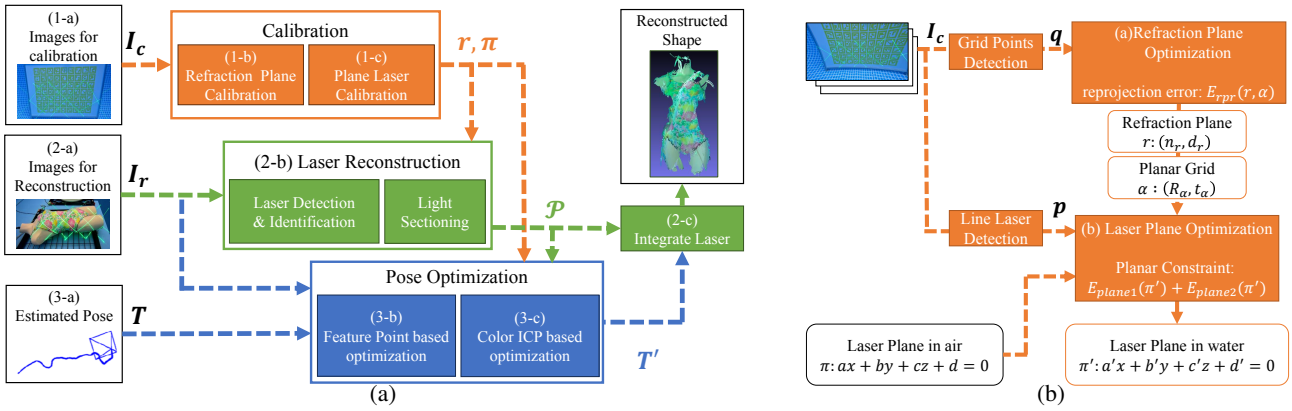


Fig. 4: (a) Algorithm of underwater 3D scan with SLAM. It consists of three processes: calibration, 3D shape reconstruction and pose optimization. Each process is explained in detail in the main text. (b) Overview of our calibration method.

B. Algorithm overview

Fig.4(a) shows a overview of the entire process of underwater shape reconstruction algorithm using ROV equipped with multi-line-lasers and a camera. As shown in the figure, the algorithm consists of mainly three parts including a calibration process, a 3D reconstruction process and a pose optimization process.

1) Calibration process: In our method, camera intrinsic parameters are assumed to be known (precalibrated in the air). Calibration begins with the utilization of images depicting laser projections on a planar grid (Fig.4(a)(1-a)), where the initial step involves estimating the refraction plane of the camera (Fig.4(a)(1-b)), followed by the estimation of each individual laser plane (Fig.4(a)(1-c)). Details are described in SecIV.

2) 3D reconstruction process: The 3D reconstruction is based on [4], and done by reconstructing the shape along laser curves in each image using the light sectioning method (Fig.4(a)(2-b)). They are integrated with the poses estimated by Visual SLAM and acoustic sensors (Fig.4(a)(2-c)). By using the refraction plane and the plane lasers parameters by calibration, the laser curves in the image I_r^i are reconstructed as a 3D point cloud $\mathcal{P}^i$ by the light sectioning method:

$$\mathcal{P}^i = \sum_k P^{i,k}, \tag{1}$$

$$P^{i,k} = IP(\pi^j, BP(p^{i,k}, n_r, d_r, \mu)), \tag{2}$$

where $p^{i,k}$ represents he 2D point of the line laser on the image I_r^i, μ is refractive index, IP is intersection point of a plane and a line and BP is backward projection with refraction based on the Snell's law. In order to integrate the reconstruction results in multiple frames and reconstruct the overall shape, the self-posture of the shape measurement system $\mathcal{P}$ is used:

$$\mathcal{P} = \sum_i \{R_c^i p^k + t_c^i | p^k \in \mathcal{P}^i\}, \tag{3}$$

where R_c, t_c are 3x3 rotation matrix and a 3-dimensional translation vector of camera respectively.

In the proposed method, when the posture of the 3D shape measurement system is estimated, it becomes possible to acquire the colors of the reconstructed point cloud. In the light-sectioning method, only the parts of the image where the laser is projected are reconstructed, so it is not possible to obtain the original colors of the unreconstructed points from that image. However, when the posture of the system is known, the reconstructed points can be projected onto images from other frames [7]. Therefore, if the projected points do not overlap with the laser, it is possible to acquire the colors of the reconstructed points from other frames. Therefore, it should be noted that the colors of the reconstructed point cloud depend on the poses of the sensor in each frame and should be recalculated once they change.

3) Pose Optimization process: The accuracy of the final reconstructed shape depends not only on the accuracy of the light-sectioning method but also on the accuracy of pose estimation. In the proposed method, posture optimization aims not only to improve the accuracy of the reconstructed shape but also to resolve texture errors. In SLAM algorithms, the same area is reconstructed several times in different frames as the 3D shape measurement system moves. The basic policy of the proposed method is to adjust the posture so that these overlapped point clouds match both in shape and texture (Fig.4(a)(3-b)). Additionally, a process is introduced to correct large errors using feature points to provide good initial values for Color-ICP (Fig.4(a)(3-c)).

IV. Underwater 3D Scanner Calibration

A. Underwater Camera Refractive Plane Calibration

Since intrinsic parameters of the camera is known, we calibrate the refractive plane parameters of the camera housing. The overview of the calibration process is shown in Fig.4(b). Input images I_c^i (i: number of images) are captured by projecting lasers onto a plane with grid points, such as a checkerboard and the intrinsic parameters of the camera in the air (fx, fy, cx, cy) and distortion coefficients (k_1, k_2, p_1, p_2, k_3) are known. First, the grid points on the image are detected, and the refraction plane of the camera housing, r, is estimated to minimize the reprojection error (Figure

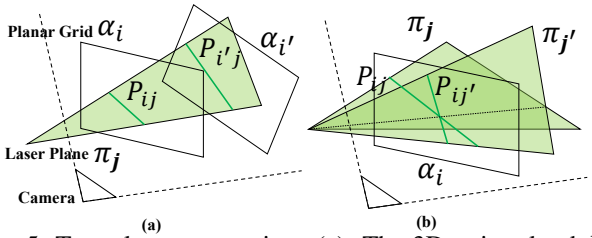


Fig. 5: Two planar constraints. (a): The 3D point cloud P_{ij}, which is reconstructed as the intersection of the planar grid α^i and the back-projected line of the laser on the image, lies entirely on the plane laser π^j, regardless of i, if j is the same. (b): $P^{i,j}$ lies entirely on planar grid α^i, regardless of j, if i is the same.

4(a)). By calibrating the normal n_r and the distance to the camera lens center d_r of r, the camera is represented as a perspective projection model with refraction(Fig.1). This calibration is performed using a planar plate α with grid points q^l(l: number of grid points), and it is obtained by minimizing the reprojection error expressed by the following equation.

$$Q^{i,l} = IP(\alpha^i, BP(q^{i,l}, n_r, d_r, \mu)) \quad (4)$$

$$e_{rpr} = FP(Q^{i,l}, n_r, d_r, \mu) - q^{i,l} \quad (5)$$

$$E_{rpr} = \sum_{i,l} e_{rpr} \quad (6)$$

where IP represents the 3D intersection coordinates of the line of sight of $q^{i,l}$ and the refraction plane α^i and FP is forward projection with refraction (more details are shown in [7]).

B. Underwater Multiple-Line Lasers Calibration

Next, the external parameters of each planar laser in water are optimized using two planar constraints(Fig.4(b)(b)). The first constraint is that the 3D reconstructed points by the same laser plane are on the same plane (the laser plane) in all images, as shown in Fig.5 (a). The equation is expressed as follows:

$$P^{i,k} = IP(\alpha^i, BP(p^{i,k}, n_r, d_r, \mu)) \quad (7)$$

$$e^{i,k}_{plane1} = DS(P^{i,k}, \pi^j) \quad (8)$$

$$E_{plane1} = \sum_{i,j,k} e^{i,l}_{plane1}, \quad (9)$$

where DS are the distance between the 3D point $P(x, y, z)$ and plane $\pi(ax + by + cz + d = 0)$ calculated as follows:

$$DS(P, \pi) = \frac{|ax + by + cz + d|}{\sqrt{a^2 + b^2 + c^2}}. \quad (10)$$

In order to determine which laser is the same in multiple images, it is necessary to identify the laser, which is done using the epipolar constraint proposed in [4]. The second coplanar constraint is the constraint that all 3D restoration points by laser planes are on the same plane (marker board) in the same image, as shown in Fig.5 (b), and is expressed as follows:

$$e^{i,k}_{plane2} = DS(P^{i,k}, \alpha^i) \quad (11)$$

$$E_{plane2} = \sum_{i,j,k} e^{i,k}_{plane2}. \quad (12)$$

The parameters of the laser plane in the water are estimated by optimizing the sum of the two coplanar constraints:

$$E_{plane} = E_{plane1} + E_{plane2}. \quad (13)$$

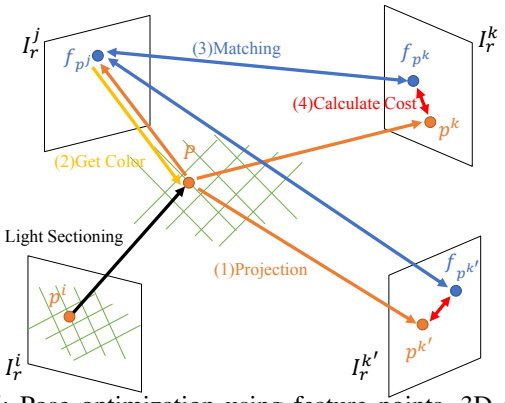


Fig. 6: Pose optimization using feature points. 3D position of the detected feature point on the target frame is back-projected to other frames to calculate the errors between corresponding points which are extracted by SuperGlue [25].

V. Camera and Laser System Pose Optimization

The purpose of the pose optimization process is to correct the discrepancy between shape and texture caused by errors in the initially estimated poses. In the proposed method, the optimization begins by correcting significant errors using image-based feature points, followed by optimizing the pose using Color-ICP to ensure global consistency between shape and texture.

A. Pose Optimization using Feature Point

To initially correct significant texture misalignments, we utilize image-based feature points to generate costs related to the colors of reconstructed points and perform local pose optimization within neighboring frames. Fig.6 illustrates the method for calculating the cost of 3D point P, which is reconstructed from p^i in frame i using light-sectioning method. First, P is reprojected onto the other frames. Subsequently, frames where the projected point does not overlap with the projected laser patterns are retained for further processing. Among the remaining frames, the distance between the reprojected point p^* and the closest image feature point f_{p^*} in the same frame is calculated, and only the frames where the distance is below the threshold are retained, such as $k, k', \cdots$. Finally, from the frames $k, k', \cdots$, the frame closest to i is selected as the base frame, denoted as j in this example. Finally, the cost for the bundle adjustment for P is as follows:

$$e = \sum_k \{||p^k - f_{p^k}|| + \delta C(p^k, f_{p^k})\}, \quad (14)$$

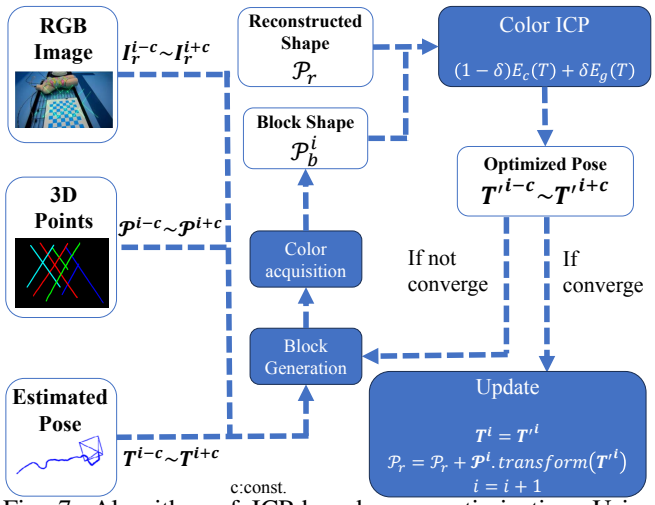


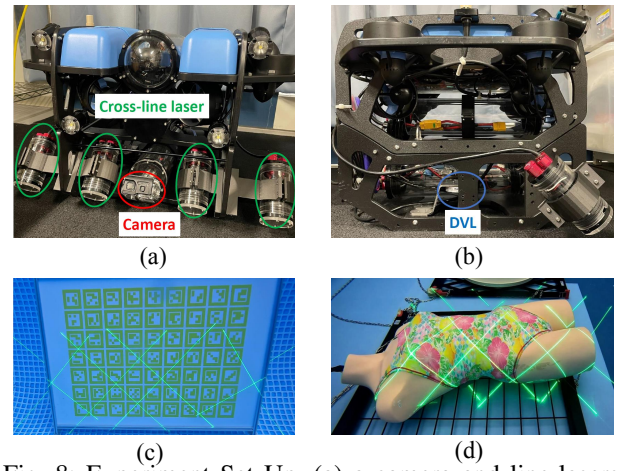
Fig. 7: Algorithm of ICP-based pose optimization. Using the correspondences extracted by Color-ICP were used to minimize the cost function (14), which achieves global optimization of refined pose parameters.

where $C(\cdot,\cdot)$ represents the sum of the squared distance (SSD) of $n \times n$ pixel patch around the two points and δ is a weight. This cost is computed for all the laser-reconstructed points and used in optimization process. Note that as the pose is updated, the base frame may change, therefore, the process of the base frame selection and the optimization process are alternated until convergence in the proposed method. SuperGlue [25] is used for feature point extraction and matching.

B. Pose Optimization using ICP with color information

After obtaining good initial values using the feature-based optimization method, pose optimization is performed with Color-ICP to ensure consistency of textures. Color-ICP is an extension of ICP widely used for aligning two point clouds by incorporating color information, making it suitable for both pose and texture optimization in our system. However, there are two challenges when using Color-ICP in our system. Firstly, the light-sectioning method can only reconstruct sparse shapes in a single frame due to multiple thin laser lines being projected, which hinders effective functioning of ICP. To address this, small blocks integrating the reconstructed shapes of multiple adjacent frames are constructed. Secondly, in the proposed system, the color of reconstructed points depends on the pose, necessitating simultaneous optimization. However, in basic Color-ICP, the color of each point is constant and does not change. To tackle this issue, the proposed method alternates between pose adjustment using Color-ICP and color recalculation until convergence, achieving pseudo-simultaneous optimization.

Fig.7 illustrates the process of our pose optimization (stage two) at frame i. $\mathcal{P}_r$ is the reconstructed shape using all the frames with the refined poses in the stage one. Initially, colored blocks $\mathcal{P}_b^i$ is generated using neighboring frames $[i-c : i+c]$ as same as the stage one. In the next step, Color-ICP is employed to adjust the pose of the block to align it with $\mathcal{P}_r$, ensuring global consistency. Color-ICP estimates



(a) (b) (c) (d)

Fig. 8: Experiment Set Up. (a) a camera and line-lasers on the ROV, (b) DVL attached to the ROV, (c) calibration board for refractive plane and multi-laser calibration and (d) target objects for scan and evaluation.

the relative pose between two point clouds $\mathbf{T}$, by minimizing the following costs.

$$E(\mathbf{T}) = E_c(\mathbf{T}) + \gamma \mathbf{E_g}(\mathbf{T}), \quad (15)$$

$$E_g(\mathbf{T}) = \sum_{(p,q)\in\mathcal{K}} ||p - \mathbf{T}q||_2 * n_p, \quad (16)$$

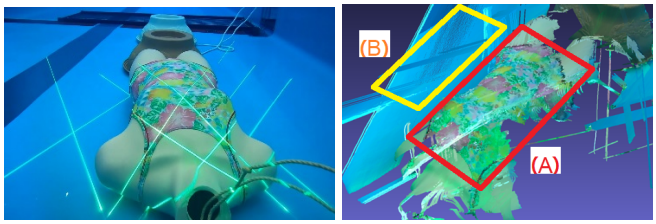
$$E_c(\mathbf{T}) = \sum_{(p,q)\in\mathcal{K}} ||F_p(\mathbf{T}q) - C(q)||_2, \quad (17)$$

where $\mathcal{K}$ represents corresponding point set, E_g is the geometric terms same as the basic ICP objective [5] and E_c is the color term to measure the difference between the color of point q (denoted as $C(\cdot)$) and the color of its projection on the tangent plane of p (denoted as $F_p(\cdot)$). After Color-ICP has been conducted, color information for point clouds are updated by using the updated pose. As this process reaches the final frame, it cycles back to the first frame. By iteratively applying this process until the Color-ICP cost falls below a threshold, camera poses converge to optimal values, provided the initial poses are reasonably accurate. In our experiments, convergence typically occurs within approximately three to five iterations of the entire frame sequence.

VI. Underwater Experiments in Swimming Pool

A. Experimental setup

To show the effectiveness of the method, we conducted an experiment in a swimming pool using underwater ROV attached with a camera, four cross-line lasers and acoustic odometry sensor as shown in Fig.8. While testing in various environments is crucial to confirm the method's generalizability and effectiveness, underwater experiments are challenging, and many studies have been conducted in limited conditions. In this paper, we have performed experiments under similar conditions, and exploring more diverse conditions will be a key focus for our future work. The RGB images were captured by GoPro HERO10, which mounted on the middle of the four cross-line lasers. The



(a) (b)

Fig. 9: (a): Capturing scene containing a floor of the pool and a mannequin. (b): The initial shape reconstruction results using camera poses estimated by image features [4]. (A) indicates the shape of the mannequin and (B) indicates the bottom of the pool.

TABLE I: RMSE of reconstruction for each calibration method of mannequin (a) in Fig.9.

RMSE [mm]			Approx.	Physical
Before pose optimization			8.83	5.72
After pose optimization	Without Block shift	ICP	8.45	5.43
		Color-ICP	8.43	5.32
	With Block shift	ICP	8.52	5.24
		Color-ICP	8.37	**5.13**

TABLE II: RMSE with the estimated plane by plane fitting of the pool bottom (B) in Fig.9, reconstructed by each calibration method.

RMSE [mm] after plane fitting	Approx. model	Physical model
	12.42	3.27

acoustic pose data was acquired by Water Linked DVL A50, Blue Robotics Bar30 High-Resolution 300m Depth/Pressure Sensor, and PX4 Pixhawk IMU sensor. For calibration, a board with 7x9 AR markers printed on it was used as a planar grid as shown in Fig.8(c). Also, a mannequin and the bottom of a swimming pool were reconstructed and evaluated as measurement targets as shown in Fig.8(d).

For quantitative evaluation, we calculate RMSE applying ICP between an estimated shape and the ground truth, which was measured in the air using commercial products (Kinect [26]) with integration algorithm [27].

B. Evaluation of reconstruction accuracy by camera method

It is common for underwater environments that there is only a few texture or feature point as shown in Fig.9(a), the bottom of the pool. To integrate the reconstructed laser curves of each frame under such condition, we used the pose obtained from DSO [24], which is a famous Visual SLAM method for such a severe condition. The final reconstructed shape is shown in Fig.9(b). To verify the effectiveness of the proposed method, we evaluated the RMSEs under different camera models for the mannequin (A) and the bottom of the pool (B). Tab.I shows the reconstruction results of mannequin with the approximation camera model and the physical camera method under various conditions. It can be seen that the accuracy of the physical method is significantly improved compared to that of the calibration with the approximation camera model, especially Color-ICP method was applied.

We also evaluated the bottom of the pool, which has a larger distance from a camera than the mannequin. For quantitative evaluation, plane fitting was performed and the RMSE with the estimated plane was calculated. The results are shown in Tab.II, proving that the physical camera method is able to reconstruct the shapes with higher accuracy than approximation camera model, even if the measurement distances are largely different from the calibrated condition.

TABLE III: Changes in RMSE with respect to GT in Fig.10.

σ[mm]	0	3	5	10	30
(a)before	5.95	8.56	10.62	17.64	32.43
(b)Sec.V-A	5.87	6.21	6.56	18.21	31.56
(c)Sec.V-B	5.67	5.92	6.13	15.34	39.37

C. Evaluation of pose and shape optimization of SLAM

Next, the effectiveness of the pose and shape optimization technique was verified using the same underwater data acquired at real swimming pool. To confirm the effectiveness of our method, Gaussian noises are added to the DSO's pose estimation results, since DSO algorithm usually does not work in real sea environment because of strong scattering, light attenuation and turbidity, and initial poses of ROV are usually retrieved by DVL and IMU sensor, which are much worse than DSO's estimation.

The initial reconstructed shapes by light-sectioning method are shown in Fig.10(a). The results after stage one are shown in Fig.10(b). It is confirmed that shapes are improved when initial noises are not more than σ =5mm. When the noise surpasses σ =5mm, we mainly observe a smoothing effect, primarily due to our method's incorporation of a pseudo-optimization process, as detailed in the methodology. Improving results under such bad condition is our future work.

Finally, the stage two optimization technique is applied to the results of the stage one, which is shown in Fig.10(c). It is clearly confirmed that all the shapes are improved from the first stage visually, especially significant improvement in the alignment of textures, which were misaligned during the reconstruction process. The zoom-up view of the improvement of σ=5mm is shown in Fig.11. Quantitative evaluations are summarized in Table.III, where it is clearly observed that the proposed method is effective when the noises are below σ=5mm after applying both the feature-point-based method (stage one) and the Color-ICP-based method (stage two).

D. Compared to previous methods

Furthermore, we compared with previous methods, such as Colmap [28], DSO [18] and the technique without pose optimization [4] under the condition of no initial noise for the camera pose. Results are shown in Fig.12. Since Colmap and DSO are image based method, reconstructed shape is inevitably sparse, such as body bard of the object. Laser-based previous method [4] achieved dense reconstruction, however, its accuracy is lower than ours even if it was reconstructed with noise=0mm. To the contrary, our method which was reconstructed even with initial noise with $\sigma = 5mm$, RMSE was the lowest, proving superiority of our method.

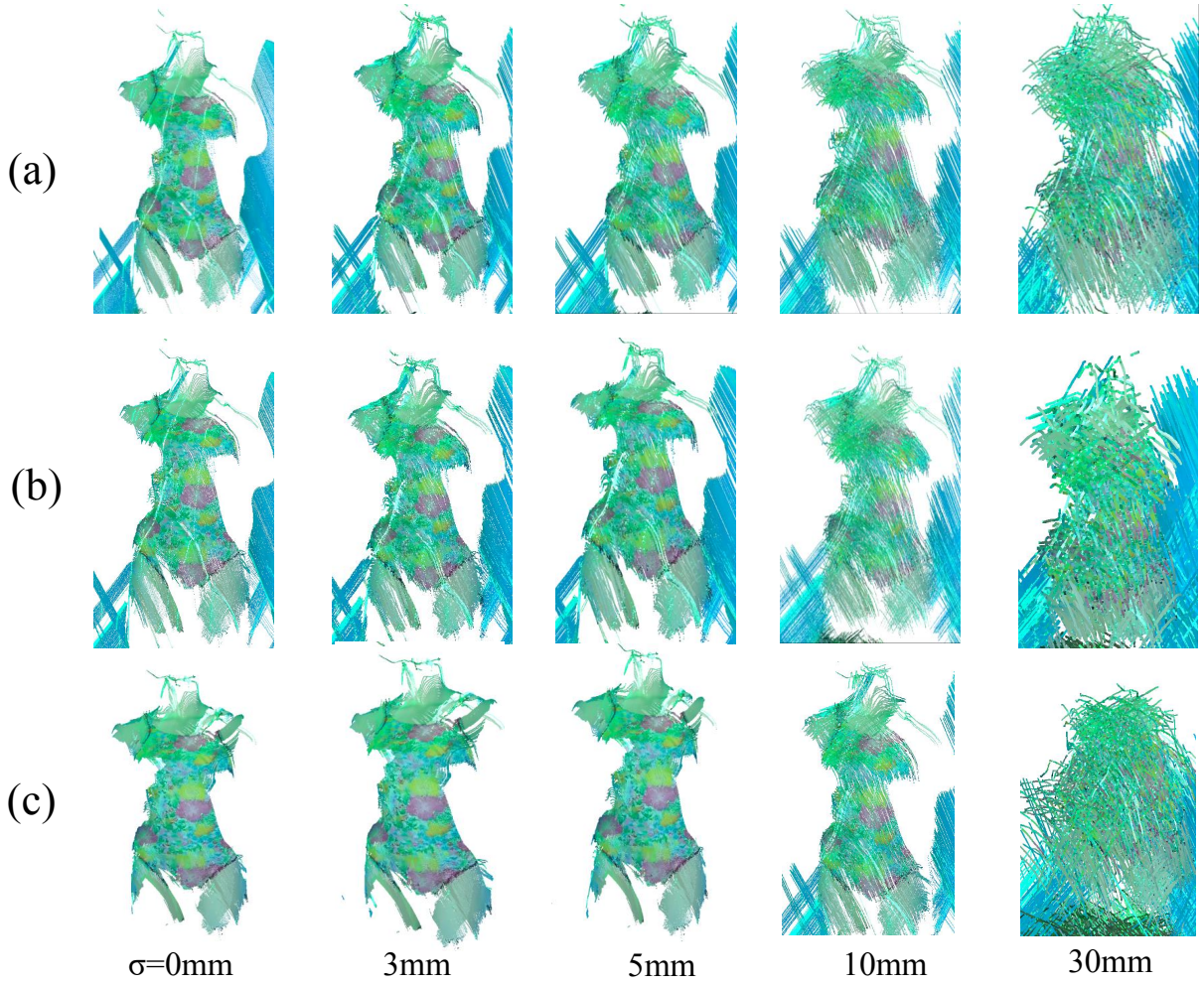


Fig. 10: Texture changes in the reconstructed shape due to pose optimization. (a): Reconstructed shape with added noise in the pose. (b): Result after applying the method of Sec.V-A to (a). (c): Further result after applying the method of Sec.V-B to (b).

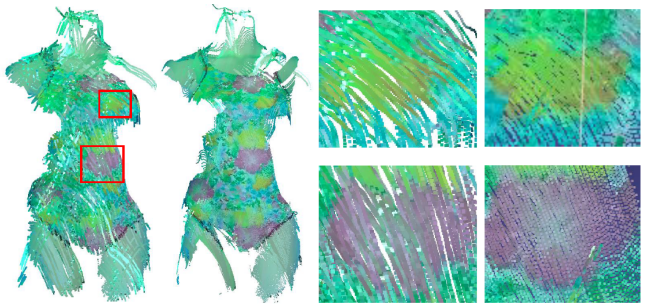


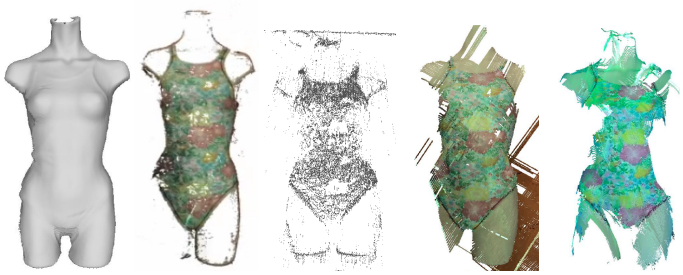
Fig.10(a) $\sigma = 5mm$ | Fig.10(c) $\sigma = 5mm$ | Close-up of (a) | Close-up of (c)

Fig. 11: Close up view of red boxes of Fig.10 (a) and (c), $\sigma = 5mm$. Improvement of texture is clearly observed.

VII. Conclusion

This paper focuses on an underwater 3D information of the scene measurement method using an RGB camera and multiple planar lasers, proposing a robust camera calibration method as well as pose optimization technique to align textures. The effectiveness of proposed physically-correct camera model was proved by comprehensive experiment compared with the previous common camera model. As for the sensor pose optimization, the proposed method consists of a two-step optimization process. Firstly, as the first optimization process, a method to correct significant errors in shape by associating image-based feature points in each frame with 3D reconstructed points by light-sectioning method was proposed. Additionally, as the second optimization process, a method to refine poses by aggregating the restored sparse curves of light-sectioning method into a block and alternating between color recalculation and Color-ICP was proposed. Experiments demonstrated that the proposed pose optimization method improves the accuracy of the restored shape even with highly added noise to the initial poses, confirming the effectiveness of the proposed method.

In the future, the robust convergence technique under strong noises as well as deep sea exploration using real systems is planned.



(a) GT by Kinect	(b) Colmap [28]	(c) DSO [18] noise=0mm	(d) Prev [4] noise=5mm	(e) Ours
RMSE[mm]	4.34	11.08	8.46	6.13

Fig. 12: Comparison results. (a) Ground truth shape obtained by Kinect Fusion [27] (b) The reconstruction result of Colmap [28] which works accurately if there is rich texture, which is usually not the case for underwater scenario. (c) The reconstruction result of DSO [18] which is used to estimate the camera poses. Only sparse points are reconstructed. (d) The reconstruction result using the approximation model [4]. Even though no noise was added, RMSE was worse than ours. (e) The reconstruction result applying our optimization technique with Gaussian noise with $\sigma = 5mm$.

ACKNOWLEDGMENT

This work was supported by JST Startup JPMJSF23DR, ACT-X JPMJAX23C2 and JSPS/KAKENHI JP20H00611, JP23K28129 and JP23H03439 in Japan.

REFERENCES

[1] M. Bleier and A. Nüchter, "Low-cost 3d laser scanning in air orwater using self-calibratingstructured light," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 105–112, 2017.

[2] K. Ichimaru and H. Kawasaki, "Underwater stereo using refraction-free image synthesized from light field camera," in *IEEE International Conference on Image Processing ICIP*, 2019.

[3] H. Higuchi, H. Fujii, A. Taniguchi, M. Watanabe, A. Yamashita, and H. Asama, "Speckle-based pose estimation for 3D measurement of the featureless environment by two cameras," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 11515, June 2020, p. 115151P.

[4] G. Nagamatsu, J. Takamatsu, T. Iwaguchi, D. Thomas, and H. Kawasaki, "Self-calibrated dense 3d sensor using multiple cross line-lasers based on light sectioning method and visual odometry," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2021, Prague, Czech Republic, September 27 - Oct. 1, 2021*. IEEE, 2021, pp. 94–100.

[5] P. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[6] J. Park, Q.-Y. Zhou, and V. Koltun, "Colored point cloud registration revisited," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 143–152.

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

[8] Y. Xu, R. Zheng, S. Zhang, and M. Liu, "Robust inertial-aided underwater localization based on imaging sonar keyframes," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[9] J. McConnell, J. D. Martin, and B. Englot, "Fusing concurrent orthogonal wide-aperture sonar images for dense underwater 3d reconstruction," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 1653–1660.

[10] M. Massot-Campos and G. Oliver-Codina, "Optical sensors and methods for underwater 3d reconstruction," *Sensors*, vol. 15, no. 12, pp. 31 525–31 557, 2015. [Online]. Available: https://www.mdpi.com/1424-8220/15/12/29864

[11] ——, "Underwater laser-based structured light system for one-shot 3d reconstruction," in *SENSORS, 2014 IEEE*. IEEE, 2014, pp. 1138–1141.

[12] A. Bodenmann, B. Thornton, and T. Ura, "Generation of high-resolution 3d reconstructions of the sea floor in colour using a single camera and structured light," *Journal of Field Robotics*, 08 2016.

[13] R. Furukawa and H. Kawasaki, "Interactive shape acquisition using marker attached laser projector," in *Fourth International Conference on 3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings.*, 2003, pp. 491–498.

[14] R. Furukawa and H. Kawasaki, "Laser range scanner based on self-calibration techniques using coplanarities and metric constraints," *Computer Vision and Image Understanding*, vol. 113, no. 11, pp. 1118–1129, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S107731420900099X

[15] Chang Woo Chu, Sungjoo Hwang, and Soon Ki Jung, "Calibration-free approach to 3d reconstruction using light stripe projections on a cube frame," in *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*, 2001, pp. 13–19.

[16] J. Davis and X. Chen, "A laser range scanner designed for minimum calibration complexity," in *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*, 2001, pp. 91–98.

[17] B. D. Bradley, A. D. C. Chan, and M. J. D. Hayes, "A simple, low cost, 3d scanning system using the laser light-sectioning method," in *2008 IEEE Instrumentation and Measurement Technology Conference*, 2008, pp. 299–304.

[18] L. von Stumberg, V. Usenko, and D. Cremers, "Direct sparse visual-inertial odometry using dynamic marginalization," in *International Conference on Robotics and Automation (ICRA)*, May 2018.

[19] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *International Journal of Computer Vision*, vol. 9, pp. 137–54, 11 1992.

[20] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[21] J. L. Schönberger, E. Zheng, M. Pollefeys, and J.-M. Frahm, "Pixelwise view selection for unstructured multi-view stereo," in *European Conference on Computer Vision (ECCV)*, 2016.

[22] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[23] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[24] X. Gao, R. Wang, N. Demmel, and D. Cremers, "Ldso: Direct sparse odometry with loop closure," in *International Conference on Intelligent Robots and Systems (IROS)*, October 2018.

[25] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superglue: Learning feature matching with graph neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[26] Microsoft, "Kinect for Windows (TOF version)," 2013, http://www.microsoft.com/en-us/kinectforwindows.

[27] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon, "Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera," in *UIST '11 Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, October 2011, pp. 559–568.

[28] J. L. Schönberger and J.-M. Frahm, "Structure-from-Motion Revisited," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.