Three-dimensional Underwater Environment Reconstruction with Graph Optimization Using Acoustic Camera

Yusheng Wang[1], Yonghoon Ji[2], Hanwool Woo[1], Yusuke Tamura[1], Atsushi Yamashita[1], and Hajime Asama[1]

***Abstract*— In this paper, a three-dimensional (3D) environment reconstruction framework based on graph optimization is proposed that uses acoustic images captured in an underwater environment. Underwater tasks such as unmanned construction using robots are becoming more and more important. In recent years, acoustic cameras which are forward-looking imaging sonars are being commonly used in underwater inspection. However, the loss of elevation angle information makes it difficult to get a better understanding of underwater environments. To cope with this, we apply 3D occupancy mapping method based on the acoustic camera rotating around the acoustic axis to generate 3D local maps. Next, from the local maps and a graph optimization scheme, we minimize the error of camera poses and build a global map. Experimental results demonstrate that our 3D mapping framework for the acoustic camera can reconstruct dense 3D models of underwater targets robustly and precisely.**

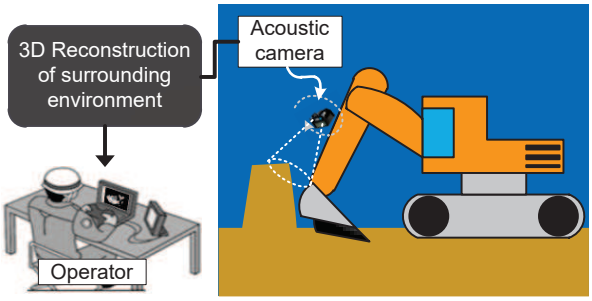


Fig. 1. Example of underwater construction by remote control of underwater robot based on 3D reconstruction of surrounding environment using acoustic camera.

I. Introduction

In recent years, waterfront development such as construction and reclamation projects of airports, ports and submarine tunnels has become much more important; however, hazards may prohibit human access and the limited field of vision due to turbidity and lack of illumination make it difficult for underwater operations. In order to fulfill tasks like inspection, removal of hazardous materials or excavation work, a remote control robot with a reconstruction system of a three-dimensional (3D) underwater environment is necessary (Fig. 1).

Recently, the development of acoustic cameras, such as dual frequency identification sonar (DIDSON) and adaptive resolution imaging sonar (ARIS) which can generate high-resolution and wide-range image, facilitates understanding of underwater situation [1]. To the best of our knowledge, most previous studies that achieve 3D reconstruction using acoustic cameras are feature-based methods because it is possible to calculate 3D coordinate values through matching of corresponding feature points extracted from multiple acoustic images, akin to stereo matching in optical images. Corners or lines are usually used as features for 3D reconstruction. Mai et al. achieved sparse 3D reconstruction from simultaneous localization and mapping (SLAM) system by tracing such features using extended Kalman filter (EKF) [2][3]. However, automatically detecting such features on the acoustic images is not easy due to noise and other factors. Moreover, this kind of sparse 3D model consisting of limited number of features is not suitable for representing complex objects; thus, a dense 3D reconstruction method is required.

To solve this problem, 3D occupancy mapping theory can be considered as a way for building dense 3D volumetric representation. Teixeira et al. accomplished 3D reconstruction of shiphull consisting of volumetric submaps using acoustic images. In this research, alignment of submaps is optimized by pose-graph [4]. However, they mounted a concentrator lens on acoustic camera which can obtain 3D information directly within a narrow range by narrowing the field of view of the elevation angle. Therefore, this approach cannot sense a wide field range, which is a major benefit of acoustic camera. Detailed description of the elevation angle will be given in the next session.

In our previous study, we treated the acoustic camera as a range sensor, instead of an image sensor and designed a full probabilistic mapping framework which can reconstruct 3D environment using acoustic images from multiple viewpoints [5]. However, it assumed that camera poses corresponding to each viewpoint are already given which is not effective in realistic problem.

In this paper, in order to solve the problem in our previous study [5], 3D local maps are generated from an effective movement which is rotating around the acoustic axis (i.e., roll rotation of acoustic camera) and graph optimization

*This work was in part funded by ImPACT Program of Council for Science, Technology and Innovation (Cabinet Office, Government of Japan), JSPS KAKENHI Grant Number 18H06483, and a part of this study is the result of "HRD for Fukushima Daiichi Decommissioning based on Robotics and Nuclide Analysis" carried out under the Center of World and Intelligence Project for Nuclear S&T and Human Resources Development by the Ministry of Education, Culture, Sports, Science and Technology of Japan.

[1]Y. Wang, H. Woo, Y. Tamura, A. Yamashita and H. Asama are with the Department of Precision Engineering, Graduate School of Engineering, The University of Tokyo, Japan. {wang,woo,tamura,yamashita,asama}@robot.t.u-tokyo.ac.jp

[2]Y. Ji is with the Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University, Japan. ji@mech.chuo-u.ac.jp

is implemented to realize accurate pose estimation of the camera and generation of a 3D global map simultaneously. As a result, it is possible to reconstruct a dense underwater 3D environment robustly and efficiently.

The remainder of this paper is organized as follows. Section II explains principles of the acoustic camera briefly. Section III describes 3D occupancy mapping framework with the acoustic camera via roll rotation. Section IV introduces the graph optimization algorithm in order to build the 3D global map from the 3D local maps. The effectiveness of the proposed method is evaluated with the experimental results in Section V. Finally, Section VI gives conclusions and future works of this paper.

II. Principles of Acoustic Camera

An acoustic camera is an active sonar which insonifies a wide range of 3D fan-shape ultrasonic wave in the forward direction and receives the reflection signal after hitting an object. Each point within the wave can be represented in a polar coordinate system by range r, azimuth angle θ and elevation angle ϕ within the scope of r_{cam}, θ_{cam} and ϕ_{cam} respectively as shown in Fig. 2(a).

As a multiple beam sonar, acoustic wave can be considered as an integration of 2D beam slices in the azimuth angle direction as shown in Fig. 2(b). For each 2D sonar wave, only range data is available which means the information of elevation angle is lost. The 2D acoustic images are generated by projecting intensity of reflection signal to the imaging plane as shown in Fig. 2(a). Range r and azimuth angle θ information is acquirable from the image. More details on the principles of the acoustic camera can be found in [6].

III. 3D Mapping

A. 3D Occupancy Mapping

3D occupancy mapping is used to solve uncertainty problem of the elevation angle in our previous study [5][7]. An acoustic camera is considered as a range sensor; thus, areas that beams have passed can be classified as occupied, free and unknown voxels. Note that the entire 3D space is divided into small voxels in this study. In order to perform the 3D mapping based on the classified voxel information in the 3D space, we take several principles into consideration as follows [5]:

- Each area with the same range and azimuth angle has the same probability.
- There is less uncertainty in free space than occupied space.
- Unknown area exists because of possible occlusion.

B. Camera motion and local maps

Camera motion may affect the result of 3D reconstruction [8][9]. For a deterministic method, corresponding points in different images can be found by intersection of two arcs for elevation angles, which can be considered as triangulation [10]. On the other hand, in the case of the probabilistic method we proposed, since each voxel belonging to one arc for the elevation angle has the same probability in one

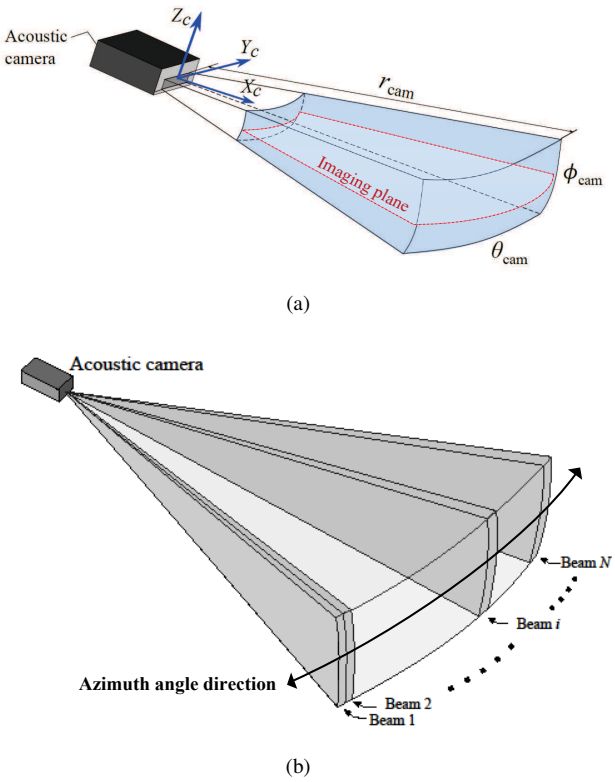


Fig. 2. Acoustic projection model:(a) geometrical model and imaging plane of the acoustic image and (b) beam slices in acoustic camera.

observation, we have to narrow the occupied points using intersection. The arcs for the elevation angles from different observations are better to intersect rather than coincide or parallel. In other words, translation of z-axis, rotations of roll and pitch (i.e., rotation on the x and y axes) are relatively effective.

In a realistic problem, it is difficult to build a whole map of an entire space directly with one measurement. Therefore, a common way is to generate several local maps and fuse them into the final global map. In order to generate a local map, camera motion should be as simple as possible. For one single motion, roll rotation can be one of the most effective ways for leading to intersection of arcs with uncertainty and easy to be implemented by using pan & tilt module mounted on the acoustic camera in general. As a result, roll rotation is chosen to generate local maps in this study.

IV. Graph Optimization

A. SLAM Framework

As we mentioned above, in order to realize 3D reconstruction of underwater scene, measurements should be taken from multiple viewpoints in different camera position. However, it is difficult to acquire accurate camera pose at every measurement. In order to build a 3D global map robustly and precisely, pose estimation method of the camera is necessary. Hence, we designed a novel simultaneous localization and mapping (SLAM) framework which is divided into the two stages: front-end and back-end stages as shown in (Fig. 3).

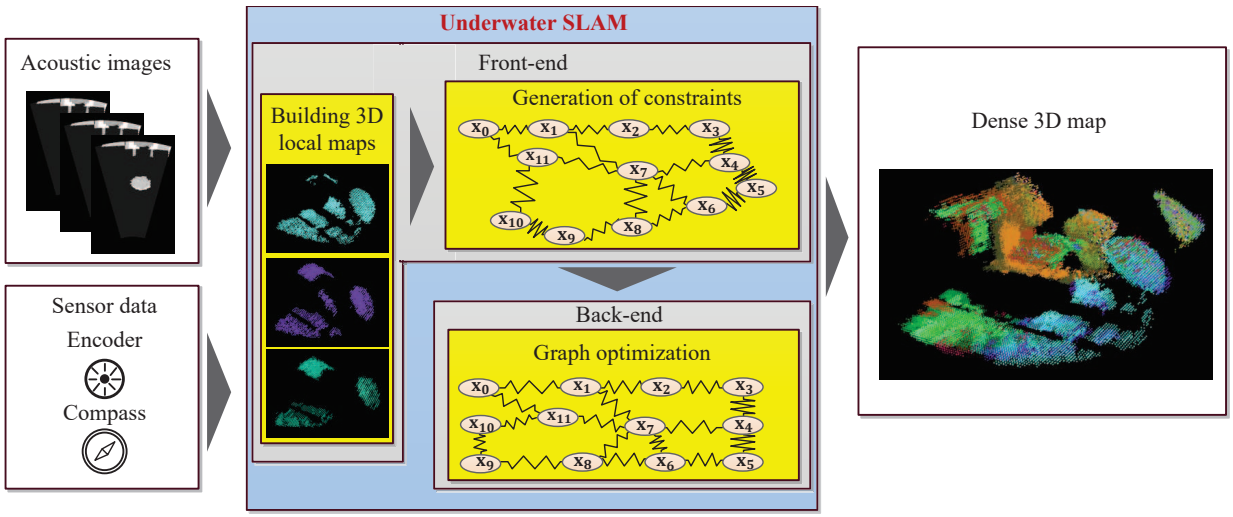


Fig. 3. Overview of SLAM framework using graph optimization. Local maps are generated from acoustic images of i-th rotation and odometry information. In the front-end stage, ICP constraints are generated by 6 DoF matching of current local map and previous local maps. Next, constraints are passed to back-end stage. Finally, global map and corrected camera pose are generated by graph optimization simultaneously.

In the front-end stage, local maps are built and matched using 3D registration through iterative closest point (ICP) [11]. Results of ICP transform and odometry data of camera movement are used as constraints which are passed to the back-end stage. In the back-end stage, we implement optimization of pose graph to find the best configuration of camera poses and realign the local maps to generate a correct global map.

B. Front-end Stage

In the front-end stage of the framework, first, we generate local maps using odometry data and acoustic images from one rotation. Due to noise and other problems like slipping, odometry data may include error which means local maps may be aligned in a wrong position. Here, ICP is used to match different local maps which is a 6 degrees of freedom (DoF) matching method for 3D registration. A coarse to fine principle is followed that max correspondence distance is large in the first ICP matching step and becomes small for the second ICP matching step. The final ICP transform matrix is reserved and transferred to the back-end stage of the framework. Denote that $\mathbf{x}_i$ is the i-th pose. In order to generate local maps, robot remains stable and does only roll rotation of acoustic camera which is considered as one scan. Roll angle is assumed to be accurate here and camera pose $\mathbf{x}_i$ is the pose at the beginning of the i-th scan. Figure 4 shows the local maps built from robot pose $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$. It is worth mention that ICP method does not work so well when two consecutive scans are too far to each other which means local maps from camera pose $\mathbf{x}_i$ and pose $\mathbf{x}_{i+1}$ should have enough overlap as shown in Fig. 4.

C. Back-end stage

Local maps are realigned using graph optimization, also known as graph-based SLAM [12].

The trajectory of robot can be described by the sequence of robot poses $\mathbf{x}_{1:T} = \mathbf{x}_1, \ldots, \mathbf{x}_T$. Each pose is considered as a node in one graph. While moving, it acquires a sequence of odometry measurements $\mathbf{u}_{1:T} = \mathbf{u}_1, \ldots, \mathbf{u}_T$ and the measurements of the environment $\mathbf{z}_{1:T} = \mathbf{z}_1, \ldots, \mathbf{z}_T$. Graph SLAM is a full SLAM method consists of estimating the posterior probability of robot trajectory $\mathbf{x}_{1:T}$ and the map $\mathbf{m}$ of the environment given all the measurements plus an initial position $\mathbf{x}_0$. The posterior probability can be written as:

$$p(\mathbf{x}_{1:T}, \mathbf{m} | \mathbf{z}_{1:T}, \mathbf{u}_{1:T}, \mathbf{x}_0). \tag{1}$$

In this study, initial camera pose $\mathbf{x}_0$ is taken as origin point of global coordinate. Let $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)^{\mathrm{T}}$, $\mathbf{z}_{ij}$ and Ω_{ij} be the mean and the information matrix of a virtual measurement between the node (pose) i and the node j. This virtual measurement can be considered as the position of $\mathbf{x}_j$ seen from $\mathbf{x}_i$ based on the observation. $\hat{\mathbf{z}}_{ij}$ denotes the prediction of a virtual measurement given a configuration of the nodes $\mathbf{x}_i$ and $\mathbf{x}_j$, which is calculated by ICP algorithm in this study. The information matrix here is calculated by the root mean square error (RMSE) of the corresponding point sets between the two nodes, with a line process weight for loop closure detection[13]. The log-likelihood l_{ij} of a measurement $\mathbf{z}_{ij}$ is

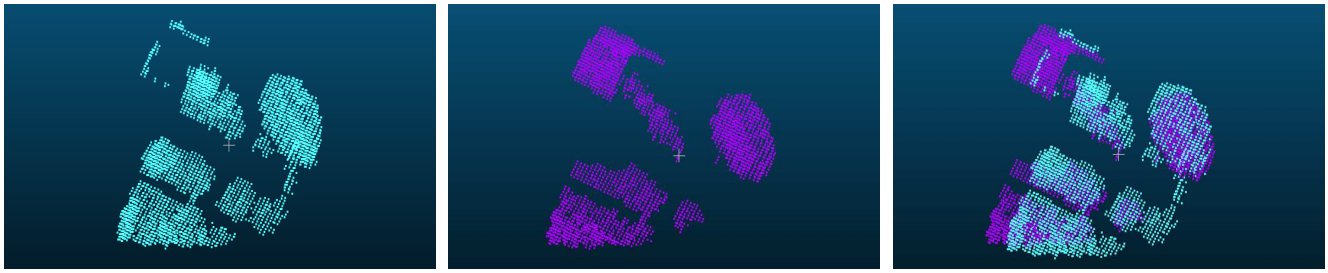
$$l_{ij} \propto \left[\mathbf{z}_{ij} - \hat{\mathbf{z}}_{ij}(\mathbf{x}_i, \mathbf{x}_j)\right]^{\mathrm{T}} \Omega_{ij} \left[\mathbf{z}_{ij} - \hat{\mathbf{z}}_{ij}(\mathbf{x}_i, \mathbf{x}_j)\right]. \tag{2}$$

Here, the error between two nodes is defined as:

$$\mathbf{e}_{ij}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{z}_{ij} - \hat{\mathbf{z}}_{ij}(\mathbf{x}_i, \mathbf{x}_j). \tag{3}$$

In order to find the optimal configuration of the nodes $\mathbf{x}^*$, maximum likelihood approach is used to minimize the sum of the error as follows:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \, \Sigma_{ij} \mathbf{e}_{ij}^{\mathrm{T}} \Omega_{ij} \mathbf{e}_{ij}. \tag{4}$$



(a) (b) (c)

Fig. 4. Examples of local maps generated from simulation data: (a) is the local map generated from a certain pose $\mathbf{x}_i$ and (b) is the local map generated from $\mathbf{x}_{i+1}$. Enough overlap is necessary between two consecutive local maps as shown in (c). Global map is generated from the integration of local maps.

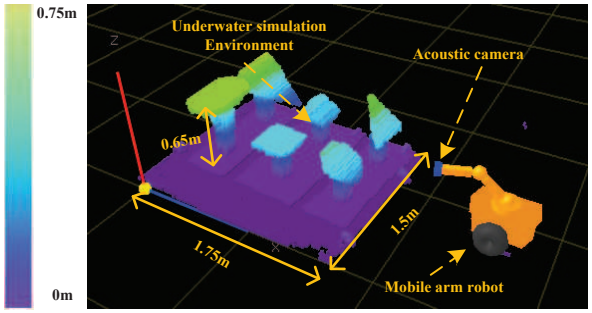


Fig. 5. Underwater robot equipped with acoustic camera on end effector of 3 DoF manipulator in underwater simulation environment. Color here is to show the height of simulation environment.

In this study, Levenberg-Marquardt algorithm is used to find the best configuration of camera poses $\mathbf{x}^*$. Figure 3 shows an example of graph optimization. In front-end stage, constraints between each node are generated and the optimal configuration of nodes is calculated in the back-end stage in order to get a minimum sum of error.

V. Experiment

To verify our method, simulation experiments were conducted with virtual acoustic images generated from acoustic camera imaging simulator [6].

A. Simulation setting

An underwater simulation environment that consists of six artificial objects fixed on ground was generated as shown in Fig. 5. The resolution of the simulation environment was 15mm.

An underwater robot equipped with an acoustic camera was used as construction machine model as shown in Fig. 5. We assumed that the acoustic camera was mounted on end effector of 3 DoF manipulator with a pan & tilt module. To generate local maps, the motion of the robot is as follow:

- Robot alternatively moves its wheels and stops to implement roll rotation of acoustic camera.
- Robot stops in a certain position, alternatively moves its arm and stops to implement roll rotation of acoustic camera.

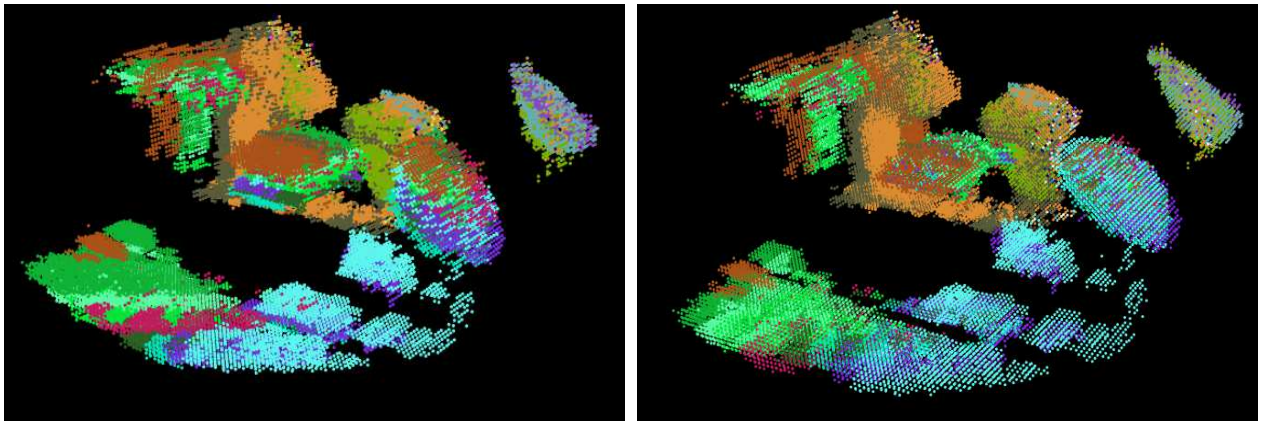
This is to say during roll rotation of acoustic camera, robot does not move in order to sense the area and acquire local maps. In other cases, robot moves its wheels or arm in order to sense a larger area. In a realistic problem, the manipulator is much heavier than the camera, so that roll rotation is considered stable without vibration. The initial robot pose was set as a global coordinate frame.

In this simulation, odometry data could be obtained under assumptions that the underwater robot equipped with sensors including encoders on the wheels and each joint of the manipulator, and the acoustic camera had a built-in compass. Here, the pitch and roll angles from compass were considered to be accurate with little rounding error, but the yaw angle is not reliable which is consistent with real data. In order to simulate noise in odometry, we added Gaussian noise with a standard deviation about maximum 50 percent of control input u_i to robot pose when robot moved. Robot moved in the first eight poses to get closer to the environment and stopped, continued to move its arm to sense the environment.

B. Simulation result

After local maps are generated, graph optimization proposed in Section IV was performed. we implemented it by utilizing Open3D library [14]. Figure 6 shows the realignment result between the local maps generated by 3D occupancy mapping described in Section III. Here, each local map is represented by a different color. To reconstruct the entire 3D environment, 14 local maps. Each local map is generated from acoustic images every 10 degree during a 360 degree roll rotation. It can be seen from the result that there are aberrations of local maps in Fig. 6(a). In other word, several planes exist which are supposed to be the same plane. On the other hand, these staggered layers overlapped after graph optimization, as shown in Fig. 6(b). To evaluate the 3D reconstruction results quantitatively, distances between 3D point cloud of each result and the ground truth (Fig. 5) were calculated, using nearest neighbor distance based on Euclidean distance between the two points. The average distances were 0.049 m in case of the 3D point cloud before optimization and 0.028 m in case of the 3D point cloud after optimization.

Figure 7 shows the camera trajectory of position from odometry, proposed method and ground truth. The movement can be separated into two steps. In the first 8 poses robot moves towards the environment, during each node robot stops to rotate the camera in order to sense the area. When



(a) (b)

Fig. 6. Registration result:(a) shows local maps before optimization (b) shows local maps after optimization. Different colors here represent different local maps. A 3D global map consists of 14 local maps.

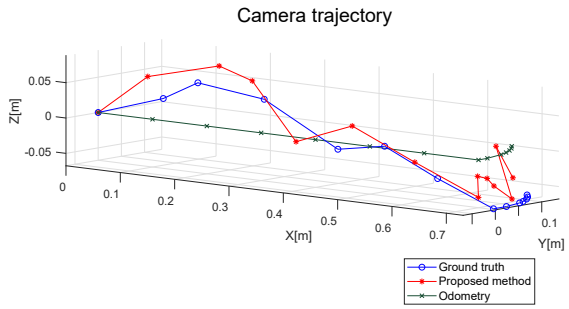


Fig. 7. Trajectory of camera position. Result of our proposed method provides promising correction of camera position.

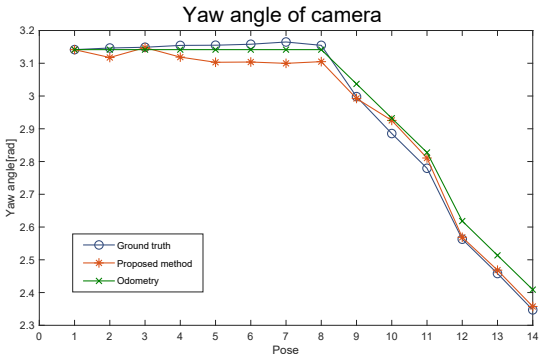


Fig. 8. Yaw angle of acoustic camera. At first, robot keeps moving without changing the yaw angle of camera. From the 9-th pose, robot starts to move its arm in order to sense the environment.

robot gets close enough to the environment, robot stops moving and starts to move its arm to sense the area. The red trajectory is estimated from our proposed method which better fits the blue trajectory of ground truth. To evaluate the rotation of camera, since roll angle and pith angle are considered accurate in this research, we mainly focus on yaw angle of camera. Figure 8 shows the yaw angle of camera from odometry, proposed method and ground truth. At the beginning, robots moves towards the environment without changing its arm angle so that odometry value is closer to the ground truth. However, from the 9-th pose, when robot starts to change its arm angle, much noise was added to the input of control that the error of odometry became larger. Our proposed method gets a better result which is much closer to the ground truth. Due to noise and distortion in local maps, 3D registration using ICP is not perfect. In some cases, like the yaw angle of camera from the 1-st pose to the 8-th pose, if noise in odometry is small, 3D registration may introduce noise to the system which leads to larger errors in estimated poses.

The above mentioned results show that our framework can successfully correct the odometry error to estimate the camera trajectory and build an accurate 3D global map of a surrounding environment.

VI. Conclusions and Future Work

In this paper, a graph optimization-based 3D dense mapping framework using acoustic images is proposed. Local maps are generated from roll rotation of the acoustic camera. A graph optimization scheme is implemented to realize pose estimation and build a global map of a surrounding environment. Experimental results demonstrate that our SLAM framework for the acoustic camera can reconstruct dense 3D model of underwater targets successfully.

Future work may include real experiment, considering the noise during roll rotation and changing the algorithm of ICP for a more robust 3D registration.

Acknowledgement

The authors would like to thank S. Fuchiyama, A. Ueyama, K. Okada and their colleagues at KYOKUTO Inc. and Y. Yamamura, F. Maeda, S. Imanaga at TOYO Corp for their kind help and useful advice on acoustic camera research.

REFERENCES

[1] E. Belcher, W. Hanot, and J. Burch, "Dual-frequency identification sonar (didson)," *Proceedings of the 2002 IEEE International Symposium on Underwater Technology (UT2002)*, pp. 187–192, Apr. 2002.

[2] N. T. Mai, H. Woo, Y. Ji, Y. Tamura, A. Yamashita, and H. Asama, "3-d reconstruction of line features using multi-view acoustic images in underwater environment," *Proceedings of the 2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligence Systems (MFI2017)*, pp. 312–317, Nov. 2017.

[3] N. T. Mai, H. Woo, Y. Ji, Y. Tamura, A. Yamashita, and H. Asama, "3-d reconstruction of underwater object based on extended kalman filter by using acoustic camera images," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 1043–1049, 2017.

[4] P. V. Teixeira, M. Kaess, F. S. Hover, and J. J. Leonard, "Underwater inspection using sonar-based volumetric submaps," *Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2016)*, pp. 4288–4295, Oct. 2016.

[5] Y. Wang, Y. Ji, H. Woo, Y. Tamura, A. Yamashita, and H. Asama, "3d occupancy mapping framework based on acoustic camera in underwater environment," *Proceedings of the 12th IFAC Symposium on Robot Control (SYROCO2018)*, Aug. 2018.

[6] S. Kwak, Y. Ji, A. Yamashita, and H. Asama, "Development of acoustic camera-imaging simulator based on novel model," *Proceedings of the 2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC2015)*, pp. 1719–1724, Jun. 2015.

[7] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: An efficient probabilistic 3D mapping framework based on octrees," *Autonomous Robots*, vol. 34, no. 3, pp. 189–206, Apr. 2013.

[8] T. A. Huang and M. Kaess, "Towards acoustic structure from motion for imaging sonar," *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2015)*, pp. 758–765, Sep. 2015.

[9] Y. Yang and G. Huang, "Acoustic-inertial underwater navigation," *Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA2017)*, pp. 4927–4933, May 2017.

[10] Y. Ji, S. Kwak, A. Yamashita, and H. Asama, "Acoustic camera-based 3d measurement of underwater objects through automated extraction and association of feature points," *Proceedings of 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI2016)*, pp. 224–230, Sep. 2016.

[11] P. J. 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, Feb. 1992.

[12] G. Grisetti, R. Kummerle, C. Stachniss, and W. Burgard, "A tutorial on graph-based slam," *IEEE Intelligent Transportation Systems Magazine*, vol. 2, no. 4, pp. 31–43, 2010.

[13] S. Choi, Q.-Y. Zhou, and V. Koltun, "Robust reconstruction of indoor scenes," *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR2015)*, pp. 5556–5565, Jun. 2015.

[14] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," *arXiv:1801.09847*, pp. 1–6, Jan. 2018.